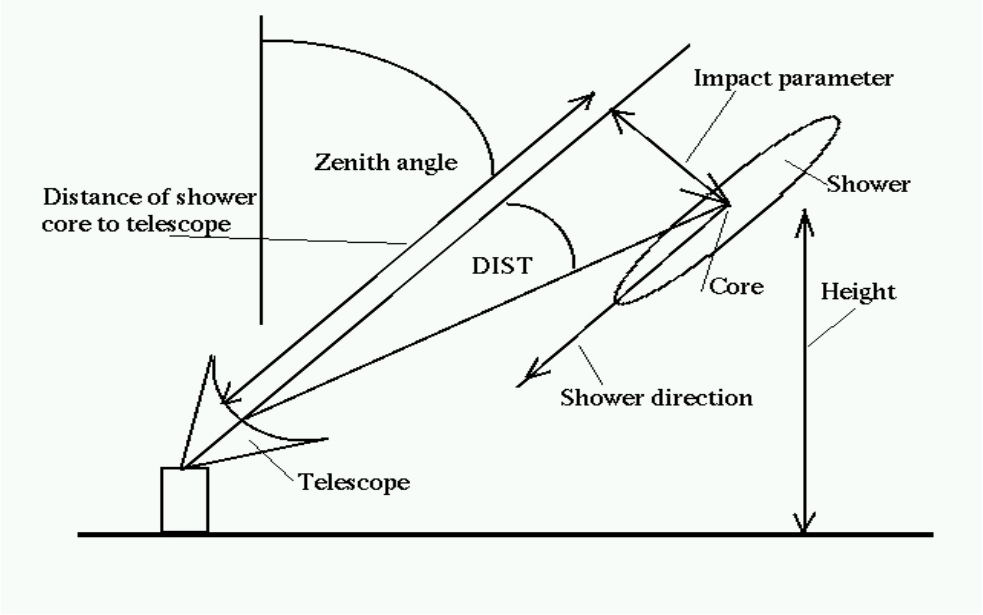


Figure A.7: This figure illustrates the imaging of a shower onto the telescope camera. It clarifies how a changing shower core position (changing energy and zenith angle) and a changing impact parameter effects the shape of the image.
It should be noted that Cherenkov light is not isotropy emitted but rather is only in the forward direction in light cones whose opening angle changes with the density of the air and the velocity of the particle.

The ***height*** of the shower maximum from the ground depends on the ***energy*** of the primary particle. The larger the energy, the deeper the shower develops in the atmosphere (see chapter 1, concerning theory and detectors). This changes the view of the shower and the image seems ***bigger*** and with a larger DIST, as illustrated in Fig. A.7.

A variation in the ***zenith angle*** has a similar effect. If this angle changes, the ***distance*** of the shower maximum to the telescope also changes. Tor ***larger*** zenith angles the maximum moves ***away*** from the telescope and the image seems generally smaller (smaller SIZE, WIDTH and LENGTH).

Similarly, a different ***impact parameter*** for the shower changes the ***ellipticity*** of the image.

Fig. A.6 shows the dependence of WIDTH and LENGTH on SIZE and DIST. It can be seen that for ***higher*** energies, gammas and protons are ***better separated,*** due to more available pixel information. In the figure a selection cut of SIZE > 100 PhE and a zenith angle < 30° has been applied.

To take ***into account the dependence*** just mentioned, the ***static cut intervals*** of Equ. A.34 need to be ***replaced*** by variable ones. This means that the ***cut intervals*** become ***functions*** of SIZE, DISTANCE and the zenith angle. This has been examined by Daniel Kranich in his PhD thesis [Kra01]. The dependence of WIDTH and LENGTH to the NSB can be corrected, as has shown Martin Kestel [Kes01].

Now let us assume that the cuts (static or dynamic) have been optimized and that we wish to apply them on a real dataset in order to extract the signal that we are interested in. It is clear that a pure signal dataset can ***never*** be achieved. The background can be suppressed ***to some degree*** but there will always be some background remaining together with the signal events. Apart from the fact that many hadronic showers look very similar to γ-showers just for statistical reasons, there exists also a certain fraction of hadronic showers that produce already at the first interaction a large fraction of energetic π^0 such

that these showers develop almost as pure electromagnetic showers. These showers are indistinguishable from the signal.

Optimization of cuts using real recorded data instead of Monte Carlo samples

It is possible to optimize cut values by only using ***recorded*** data and without touching MC data. The essential point is to ***know*** the background in the dataset, either by OFF-data or by some other kind of background estimation.

The procedure goes as follows:One can make a ***good guess*** for the initial values of the cut intervals, apply it to a set of real ON-data and estimate its background content after cut. Now the cut intervals can be optimized by ***maximizing the significance***.

Then the optimized cuts can be used on orthogonal datasets for analysis. This technique has been used by Daniel Kranich [Kra01].

A.3.4 Background determination by using the ALPHA-plot

Earlier it was mentioned that it is never possible to obtain a pure γ-events sample from a recorded data set because it will always be ***contaminated*** with hadronic background. For the calculation of the signal amplitude and its significance the background needs to be subtracted from the contaminated dataset. Therfore, a precise information about the background is needed. There are basically ***two possibilities***:

1. ***Two separate*** telescope ***runs*** are taken, one with the telescope pointing to a source (***ON-data run***) and the other one with the telescope pointing a few degrees away from the source (***OFF-data run***). After the selection cut are applied to both runs, the background can be determined precisely from the OFF data run and can be subtracted from the ON data run. Using Equ. A.39 the significance of the detection can be calculated.

2. For point source studies, the ***ALPHA plot*** can be used for this purpose since (see Fig. A.8) the γ events accumulate in the region up to about 12°, owing that the images point towards the center of the camera. By ***interpolating the shape of the background*** (which should be as flat as possible) the background in the region up to 12° can be estimated.

The latter method works quite well and was originally worked by G. Hermann in Heidelberg and has been successfully applied by Daniel Kranich [Kra01] in his thesis. The background estimation is achieved by fitting a ***simple polynomial*** to the distribution. The required conditions for the polynomial is that it has a ***horizontal tangent*** at 0° (because of the symmetry of the signed ALPHA distribution to the y-axis). While Daniel Kranich [Kra01] used a polynomial of second order, in this analysis a polynomial of order 2.5 is applied because background shape studies have shown that this is the function that provides the best fit to the background distribution.

$$f(ALPHA) = a + b\,|ALPHA|^{2.5};\ ALPHA \geq 0 \tag{A.41}$$

Assuming that the signal distribution of ALPHA is to first order approximated by a ***Gaussian distribution***, it can be used to estimate the signal events [Kra01]:

$$g(ALPHA) = \frac{N_{Signal}}{\sqrt{2\pi}\sigma} e^{-\frac{ALPHA^2}{2\sigma^2}} \tag{A.42}$$

The ***combination*** of both $h(\alpha) = f(\alpha) + g(\alpha)$ has 4 free parameters and gives a very good measure of the excess events by estimating the background from the ALPHA distribution in the region [20°, 70°] (see Fig.A.8).

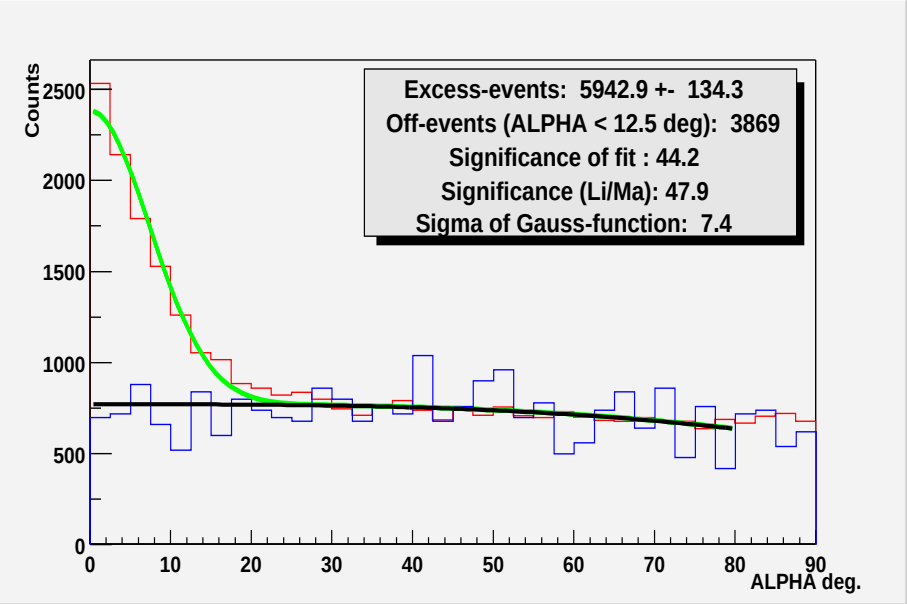


Figure A.8: This plot shows the data of a part of the dataset of Mkn 421 during its flare period from February to May 2001 (red line). The simple static cuts of Equ. A.34 have been applied. Background and excess events are estimated by fitting a combination of a Gaussian distribution (green line) and a polynomial (black line). For comparison, the blue line shows an off data dataset which has much lower statistics. We will see that the linear discriminant method and new image parameters improve the separation significantly. The new method finds almost a factor 3 more excess-events and it also doubles the significance. The expression 'significance of the fit' denotes the ratio of excess events and its error, both of them obtained by the fit. The expression 'significance (Li/Ma)' is the significance that has been calculated by using Equ. A.39 (with $\alpha = 1$).

This procedure works well if the ALPHA background distribution is ***quite flat***, which unfortunately is not necessarily the case since the limited diameter of the camera ***truncates*** events that are large or have sizable impact parameters. Events that are truncated by the camera border are reconstructed ***incorrectly*** with an ALPHA close to 90°. Consequently, for this method to work properly, the distribution must be ***forced*** to become as flat as possible. This can be accomplished , for example, by excluding events with a large DIST.

A.3.5 The linear discriminant analysis as a dynamical cut and as a tool to quantify the discrimination power of image parameters

The ***linear discriminant analysis*** (***LDA***) is used to find optimized and unbiased linear combinations of variables in order to separate classes of events [Fab97]. The method was invented by Fisher [Fis37]. In this case, the classes are signal and background events. It deals naturally with dependencies between variables and is not useful if they are NOT dependent on each other. In this case, normal cuts give better results. In this section I show how a combination of static cuts and the LDA achieves a good separation.

As practical side information, an estimator, called ***discriminating power***, of how well two samples can be separated, is obtained. The linear discriminant method is nothing more than a simple ***linear*** neural net with only ***one neuron***, a so-called perceptron with ***linear output***, trained by an analytical method (linear least squares method), rather than doing stepwise optimization.

The main ***advantage*** of the LDA, compared to a neural net and to other cut-optimization methods as well, is that it is ***analytically optimized*** while all other methods need an ***iterative*** maximization procedure whose result depends on ***initial values*** and which are often ***not reproducible***.

It is exactly this reproducibility that is needed when seeking an ***estimator*** that quantifies how good a given parameter set is able to separate two classes of events.

Description of the method

The method finds ***discriminant factors*** which are ***linear combinations*** of the input parameters. For the case of ***two*** classes, there is only one discriminant factor which is called ***discriminant variable*** and which is the optimum linear combination for the separation between the two classes. The method is first described for the case of q classes and then simplified to two classes.

Each class has N_i events. The total number of events is:

$$N_{tot} = \sum_i N_i \tag{A.43}$$

Let $\mathbf{x}^i(n) = \left(x_1^i(n), \ldots, x_p^i(n)\right)$ be a vector of p input parameters which represents the event n in class i. The ***expectation value*** of parameter x_j in class i is

$$E\left[x_j^i\right] = E_j^i = \frac{1}{N_i}\sum_{n=1}^{N_i} x_j^i(n) \tag{A.44}$$

and for all events of ***all classes combined*** we have:

$$\begin{aligned} E\left[x_j\right] = E_j &= \frac{1}{N_{tot}}\sum_{i=1}^{q}\sum_{n=1}^{N_i} x_j^i(n) \\ &= \frac{1}{N_{tot}}\sum_{i=1}^{q} N_i E_j^i \end{aligned} \tag{A.45}$$

The overall ***covariance matrix*** element between the variables x_k and x_l for all classes combined is given by

$$T_{kl} \equiv covar\left[x_k, x_l\right] = E\left[\left(x_k - E\left[x_k\right]\right) \times \left(x_l - E\left[x_l\right]\right)\right] \tag{A.46}$$

The covariance matrix can be ***split into two parts***, the '***within' matrix*** $\mathbf{W}$ and the '***between' matrix*** $\mathbf{B}$

$$T_{kl} = W_{kl} + B_{kl} \tag{A.47}$$

whereas

$$W_{kl} = \frac{1}{N}\sum_{i=1}^{q}\sum_{n=1}^{N_i}\left(x_k^i(n) - E\left[x_k^i\right]\right) \times \left(x_l^i(n) - E\left[x_l^i\right]\right) \tag{A.48}$$

$$B_{kl} = \sum_{i=1}^{q}\frac{N_i}{N_{tot}}\left(E_k^i - E_k\right) \times \left(E_l^i - E_l\right) \tag{A.49}$$

$\mathbf{W}$ describes the variances of one parameter within the group whereas B describes the differences between groups. Let us now introduce the linear combination of the input parameters

$$u(n) = \sum_{j=1}^{q} y^j x_j(n) = \mathbf{y}^T\mathbf{x}(n) \tag{A.50}$$

where $\mathbf{y}^T$ is the transposed matrix of $\mathbf{y}$ and $\mathbf{y}$ is the equivalent of ***weight vectors*** in the case of neural nets. The variance of $u(n)$ is given by

$$var(u) = \frac{1}{N_{tot}}\sum_{i=1}^{q}\sum_{n=1}^{N_i}\left(u^i(n) - E\left[u^i\right]\right) \tag{A.51}$$

$$= \mathbf{y}^T\mathbf{T}\mathbf{y} \tag{A.52}$$

$$= \mathbf{y}^T\mathbf{W}\mathbf{y} + \mathbf{y}^T\mathbf{B}\mathbf{y} \tag{A.53}$$

The first term in the last line is equivalent to the variance ***within*** the classes while the second term describes the differences ***between*** the classes. To optimize the discrimination, the first term must be ***minimized*** while the second one must be ***maximized***. The discriminating power is defined as

$$f(u) \equiv \frac{\mathbf{y}^T\mathbf{B}\mathbf{y}}{\mathbf{y}^T\mathbf{T}\mathbf{y}} \tag{A.54}$$

and the ***optimization*** is done by deriving $f(u)$ and setting it equal to zero:

$$\frac{\partial f(u)}{\partial(y_j)} = 0$$

This leads to the expression

$$\mathbf{T}^{-1}\mathbf{B}\mathbf{y} = f(u)\mathbf{y} \tag{A.55}$$

which means that $f(u)$ is maximal if it is an ***eigenvalue*** of $\mathbf{T}^{-1}\mathbf{B}$ and that $\mathbf{y}$ (the weight vector) is the corresponding eigenvector.

Now in our case of ***two classes***: We have gamma events and hadronic events. They will be denoted with indices G and H, respectively. The ***expectation values*** of the two classes for each parameter can be expressed as a vector:

$$\mathbf{E} = \left(E_1^i, ..., E_p\right)^T \tag{A.56}$$

The matrix becomes

$$\mathbf{B} = \mathbf{v}^T\mathbf{v} \tag{A.57}$$

	ALPHA	WIDTH	LENGTH	CONC	DIST	log SIZE	$\cos\theta$	RMS
ALPHA	100	-0.5	-0.1	1.0	0.4	-0.5	-0.3	-0.1
WIDTH		100	42.6	-59.6	-5.9	36.5	15.4	-7.6
LENGTH			100	-70.0	-21.1	35.7	10.3	-3.8
CONC				100	22.0	-47.7	-14.7	5.8
DIST					100	-7.5	0.3	0.5
log SIZE						100	5.6	-1.6
$\cos\theta$							100	-20.0
RMS								100

Table A.1: The table shows the correlation in percent for the most important image parameters. WIDTH, CONC and LENGTH are strongly correlated. SIZE and CONC also show a dependence. Interesting is the correlation between zenith angle and NSB (pedestal RMS). ALPHA is not correlated with any of the image parameters.

with

$$\mathbf{v} = \frac{\sqrt{N_G N_H}}{N_{tot}} \left(\mathbf{E}^G - \mathbf{E}^H\right) \tag{A.58}$$

Now the solution to the maximization problem (Equ. A.55) for our ***weights*** $\mathbf{y}$ gets the simple form

$$\mathbf{y} = \mathbf{T}^{-1}\mathbf{v} \tag{A.59}$$

The ***discrimination power*** is a very useful quantity as we will see later. It can be used to quantify how well the datasets have been discriminated. It shows how well the parameter-set is able to discriminate between the two classes. Its value ranges from 'zero' to 'one'. 'One' means perfect separation and 'zero' means no separation.

$$D = f(u) = \mathbf{v}^T\mathbf{y}\,;\; D \in [0,1] \tag{A.60}$$

Equ. A.50 can be used to calculate the center (offset of origin) between the two distributions $u^i(n)$ for gammas and hadrons (i=G,H) by taking the ***average of the two expectation vectors***:

$$u_0 = \frac{1}{2}\mathbf{y}^T\left(\mathbf{E}^G + \mathbf{E}^H\right) \tag{A.61}$$

Finally, the ***output of the LDA*** (similarly to the output of a neural net) takes the form

$$LDA_{out}(n) = u(n) - u_0 = \mathbf{y}^T\mathbf{x}(n) \tag{A.62}$$

Before applying the method, the ***image parameters*** that enter as inputs to the method must be ***chosen.***

Application of the LDA to Monte Carlo Data samples

As a first step it must be stated that the separation is ***best*** if the number of events in both samples are ***equa***l. If this is not the case, the datasets have to be ***normalized*** to the same number of events.

As explained above, the LDA is expected to achieve an ***improvement in separation*** only in the case of ***correlated*** variables. It uncorrelates dependencies. All parameters that are ***not*** correlated should be treated by a ***static cut***. Tab. A.1 shows the ***correlation matrix*** for a real data sample (mainly hadronic events).

The correlation matrix for MC gamma events looks similar but is not exactly the same. The table shows that ***ALPHA*** has absolutely ***no correlation*** to the other parameters at all. It is used separately to ***estimate*** the background. It is interesting to see that there is ***significant correlation*** between the night sky background (***pedestal RMS***) and the ***zenith***

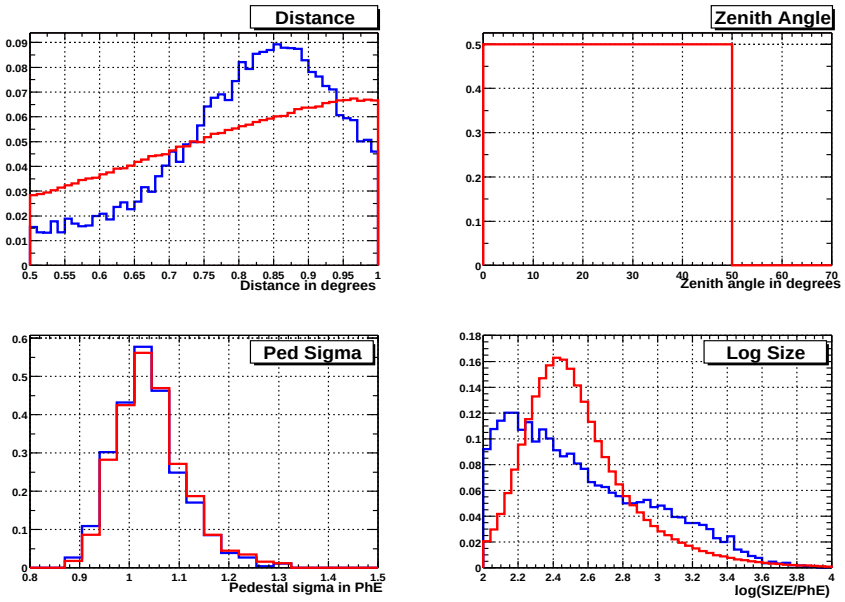


Figure A.9: Distributions of the four parameters SIZE, DIST, Pedestal RMS and the zenith angle, for gamma events (blue) and hadronic events (red) after normalization. The SIZE distribution shape for hadrons and gammas is quite different because of a) a different spectral index of the γ-MC and the OFF-data and because of b) different effective areas for hadrons and γ's. The pedestal RMS (NSB) distribution is almost the same because the NSB has been simulated to the γ-MC in a way that it is equal to the OFF-dataset.

angle. The correlation between ***SIZE*** and the other three parameters: ***WIDTH, LENGTH*** and ***CONC*** is logical. The parameter ***CONC*** has such a high degree of correlation that it is essentially represented by ***WIDTH*** and ***LENGTH*** .

The image parameters can be divided into two groups:

1. ***cut parameters*** for background suppression and
2. ***non-cut parameters*** e.g. estimators.

The distributions of the ***non-cut parameters*** as ***SIZE***, ***DIST***, ***pedestal RMS*** and the ***zenith angle*** have to be ***equal*** for both, gamma events and hadronic events, because we ***do not*** wish to cut on them. Still they are included in the LDA input parameters to ***correct for dependencies*** of the cut parameters on them (similar to dynamic cuts).

To make the distributions of the non-cut parameters equal

1. the night sky background is simulated in the MC gamma sample in a manner such that it has the ***same distribution*** with respect to ***zenith angle*** as the OFF-data set (which is used as hadronic events).
2. The z***enith angle distribution*** was normalized to be ***flat*** in both data sets since it is important to have equal cut efficiencies for all zenith angles.

In this analysis the ***SIZE*** distributions and the ***DIST*** distributions are ***not normalized*** to each other. The experience showed that the separation does not improve when they are normalized. Fig. A.9 shows the distribution of these four parameters after normalization.

If the parameters are included ***linearly***, the LDA represents a ***cut plane*** in a ***multi-dimensional space***, separating the two datasets (represented by vectors) maximally with

respect to discriminating power. Figure A.10 shows the output of equ. A.62 after optimization (equ. A.59and equ. A.60). The discriminating power is 0.62. The parameters included are:

$$\begin{aligned} Input[0] &= WIDTH \\ Input[1] &= LENGTH \\ Input[2] &= CONC \\ Input[3] &= SIZE \\ Input[4] &= \cos\theta \end{aligned} \tag{A.63}$$

Again, ***ALPHA*** is not included in this list, since it is a ***independent*** parameter. It is used to ***estimate*** the background.

A cut in the ***one dimensional output*** of the LDA can be chosen. Depending on the position of the cut, a higher or lower percentage of background is removed while retaining the signal events respectively. Fig. A.11 shows the cut efficiencies ε_G and ε_H against each other. Fig. A.12 demonstrates the performance of the LDA for a cut which achieves maximal significance. The cut efficiency increases for increasing energy and remains ***flat*** for increasing zenith angle, as it should be.

If ***higher orders and combinations*** of image parameters are included in the LDA then it no longer corresponds to a linear cut plane, but rather to a ***multidimensional surface*** that follows the multi-dimensional ***shape*** of the parameter distributions in a more ***adaptive*** manner (in a similar way as dynamic cuts). Consequently, the separation improves. The following terms have been appended to the previous parameter list:

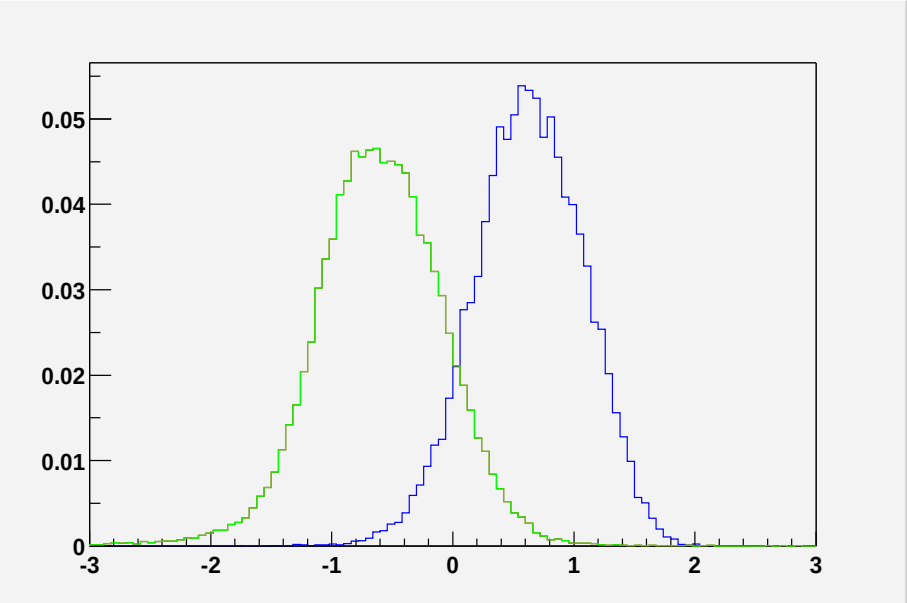
$$\begin{aligned} Input[5] &= WIDTH^2 \\ Input[6] &= LENGTH^2 \\ Input[7] &= CONC^2 \\ Input[8] &= \frac{WIDTH * LENGTH}{SIZE} \end{aligned} \tag{A.64}$$

Fig. A.13, Fig. A.11 and Fig. A.10 show the improvement in separation using higher orders and a combination of these parameters. The excess events found with linear input gives 11400 +- 190. This increases to 12600 +- 200 for the LDA with higher order inputs.

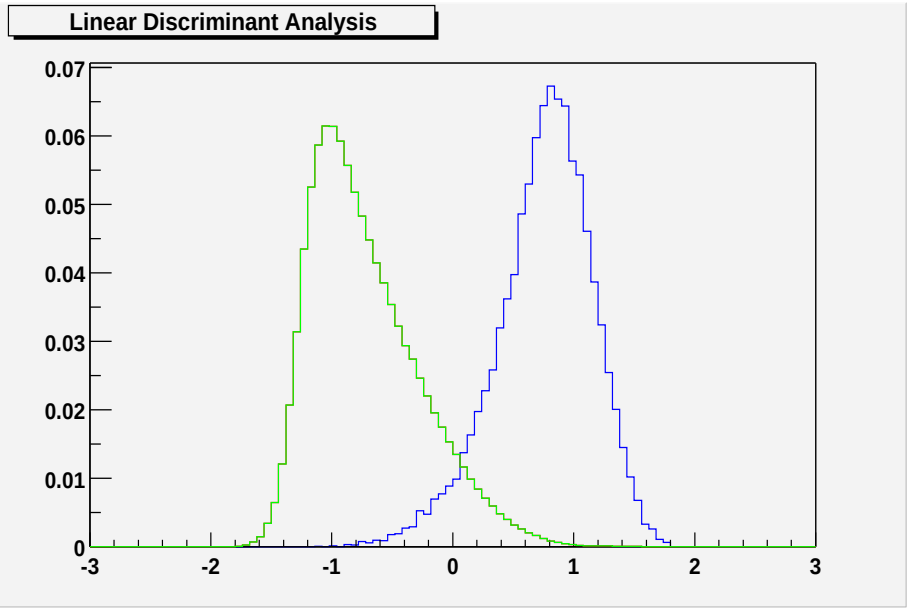
The Fig. A.14 shows the excess events found as a function of the cut position of the LDA output. A ***plateau*** appears on the left side of the plot which can be interpreted as the ***total signal events*** in this data set. A fit on the plateau gives $N_{Ex} = 15030 \pm 71$. When compared with the number of excess events found at maximum significance (12600 +- 200), this number results in a ***total cut efficiency of 83 %***. It must be noted that before application of the selection cut, a ***'precut'*** on the minimum SIZE < 60 PhE, on the zenith angle < 50° and on the distance 0.4 < DIST < 1.05 was performed, thus previously reducing the total number of signal events in the dataset. It is clear that this plot here does ***not*** tell us anything about the ***trigger efficiency*** nor about the efficiency of the 'precut'.

A.3.6 Comparison of the discrimination power of the LDA and the dynamical cuts, tested with real Mkn 421 data

For the comparison of ***dynamical cuts*** of the analysis program 'jacuzzi' at MPI Munich [Kra01] with the ***LDA*** of this thesis, the ***same*** dataset was used ***without*** pointing corrections (pointing corrections increase the significance of the signal). For case of the dynamical cuts a correction of the parameter with respect to the NSB (the so-called 'zonking' [Kes01]) and also a cut on the ***asymmetry*** was applied (see the following section). Both

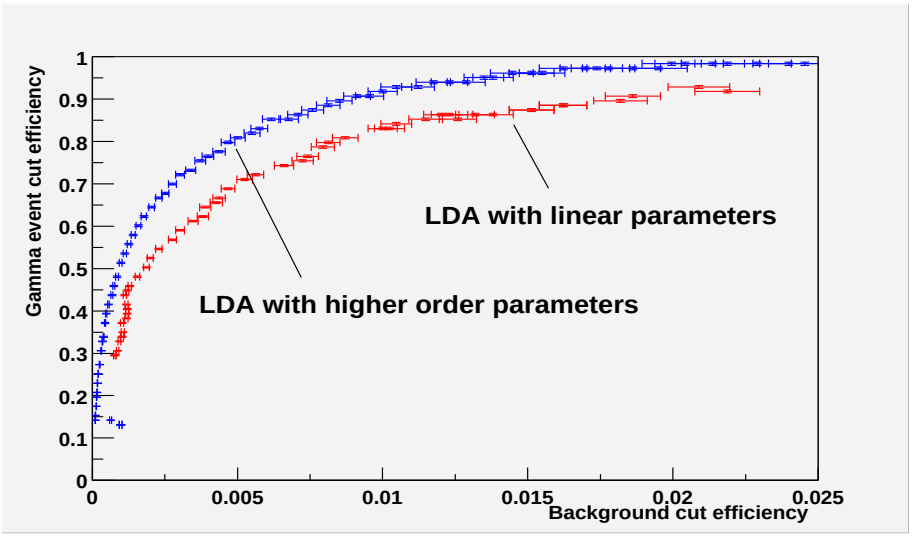


(a) Linear inputs

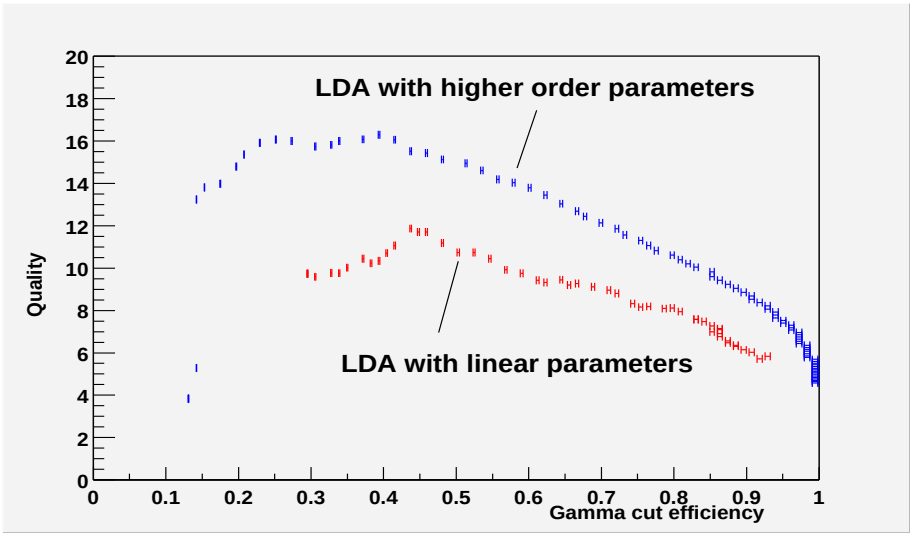


(b) Higher order inputs

Figure A.10: Output of the LDA A.62 for gamma events (blue) and hadronic events (green). The areas are normalized to one. In (a) the parameters are included linearly (D=0.62) and in (b) higher polynomial orders are included as well (D=0.748). It is clearly shown that the discrimination increases by including higher orders of the parameters.



(a)



(b)

Figure A.11: (a) Gamma cut efficiency vs background cut efficiency and (b) Quality factor for the LDA vs LDA-output. Both plots show the performance of LDA with only linear inputs and with higher order inputs. The inclusion of higher orders in the LDA improve the quality factor and its separation capability significantly.

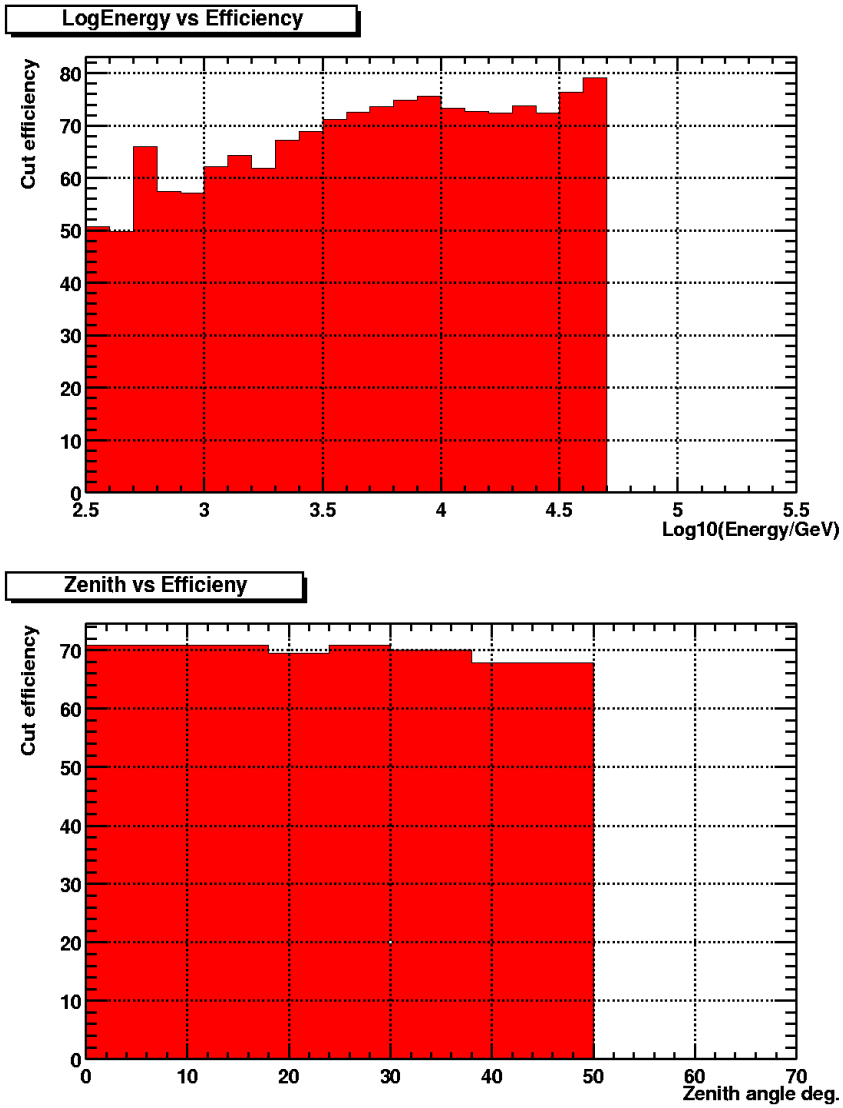
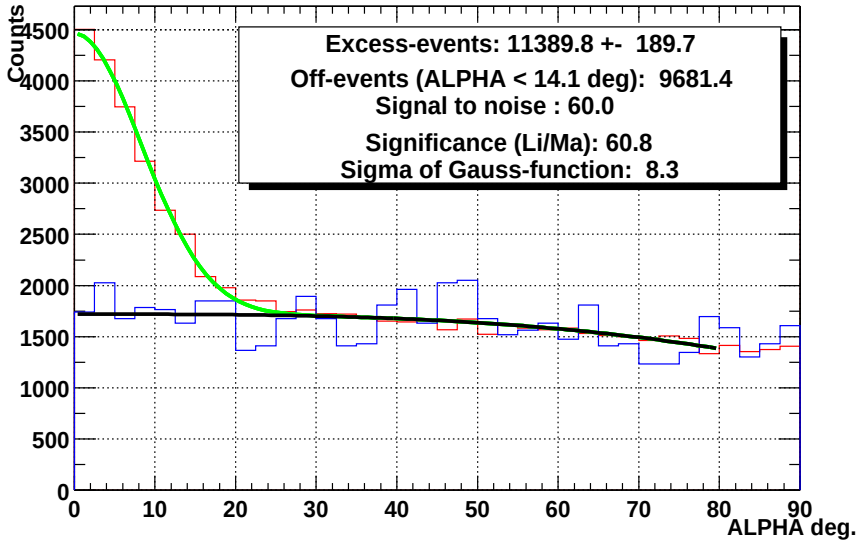
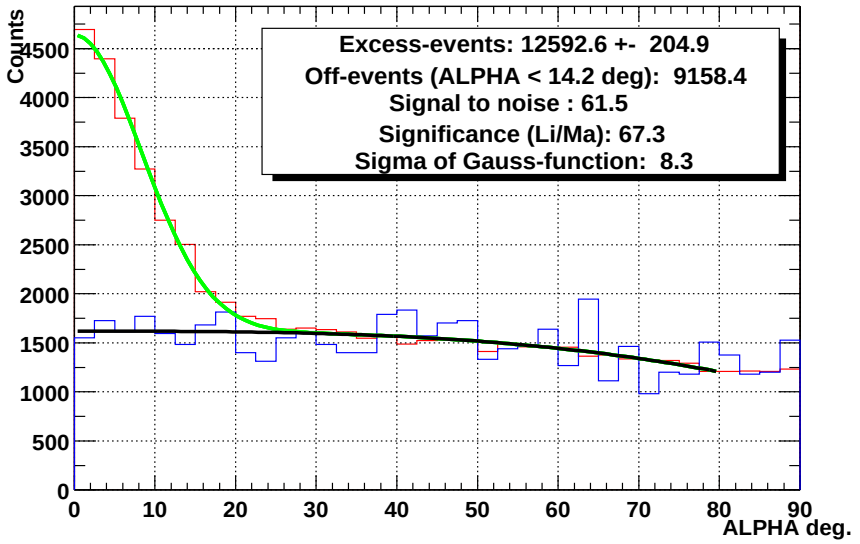


Figure A.12: These two plots show the γ cut efficiency for the LDA with parameter input of only linear order and for a cut which achieves maximal significance. In the upper the γ cut efficiency is plotted against the logarithm of the energy in log(GeV) and in the lower plot against the zenith angle. The cut-efficiency increases with energy and is flat with zenith angle as it should be.



(a)



(b)

Figure A.13: The linear discriminant analysis is applied to a sample Mkn 421 dataset of 167 h observation time. a) The upper plots shows the result for the LDA with linear inputs. It yields in a significance (Li/Ma) of 60.8. b) The lower plots shows the result for the LDA with higher orders of image parameters for the input. The background is more suppressed while more signal is found (increase from 11400 events to 12600 events). The significance improves by 10 %. Here a previous filter cut on SIZE > 60 PhE, the 0.4 DIST > 1.05 and on the zenith angle <50° was applied.

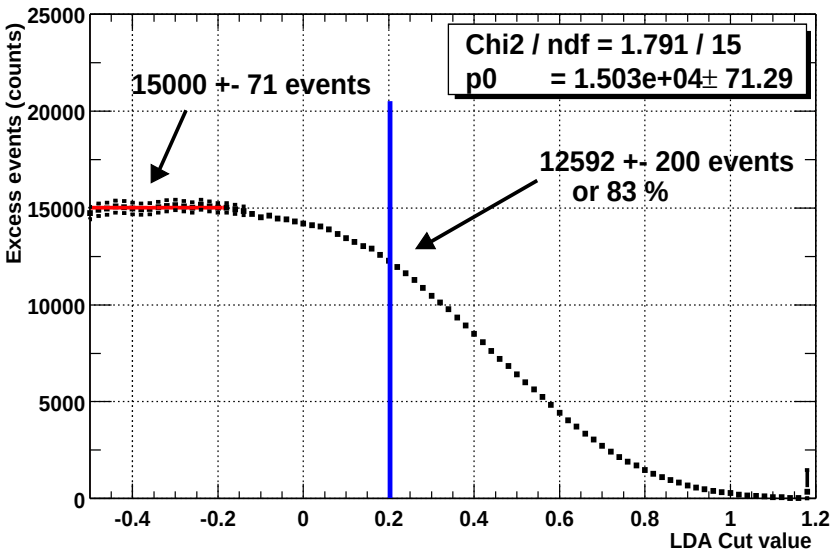


Figure A.14: When moving the LDA cut position a plateau appears on the left side which indicates the maximum excess events in this dataset. This maximum represents the total signal events in this dataset. Here a previous cut on SIZE > 60 PhE, the 0.4 DIST > 1.05 and on the zenith angle <50° was applied. After the LDA selection (higher order inputs) about 83 % of the total events have been selected.

improvements ***have not been*** included in the ***LDA*** procedure up to now. The result of the dynamical supercuts can be seen in Fig. A.15.

As a result of not applying pointing corrections, the ALPHA distribution is wider and unfortunately, 'jacuzzi' was fitting the ALPHA distribution with a fixed variance for the Gaussian distribution, expecting a narrower distribution (it was not made for that). Consequently, the fit has large chisquare. Still, the fit result can be ***roughly*** compared with the result achieved by the LDA. The excess events found by 'jacuzzi' are 7700 +- 130 and the LDA (in this simple version) found 12600 +- 200. The significance of the LDA is around 20% higher. It has to be kept in mind that jacuzzi also finds higher significances when a pointing correction has been applied beforehand.

As a ***conclusion*** it can be stated that the LDA is at least as efficient than dynamical cuts.

A.3.7 Conclusion about the efficiency of the LDA separation method and an outlook

The LDA proves to be a very efficient separation method that can well compete with classic methods like static cuts and the dynamic cuts and has ***equal*** or better ***performance***. It shows to be especially efficient for very low energy events (< 1 TeV) and very high energy events (> 10 TeV).

The ***main advantage*** of the ***LDA*** is clearly the cut optimization algorithm that is ***analytically solvable***. This does not only speed up the optimization process but it does also ***not depend*** on any initial value as any other stepwise optimizing methods (e.g. 'MINUIT'-package). Its results are 100 % reproducible unlike any other kind of separation methods as dynamic cuts or neural networks.

For this reason it is especially suited to ***quantify the separation power*** of different image parameters sets in order to find the most efficient one. The LDA will be used in the

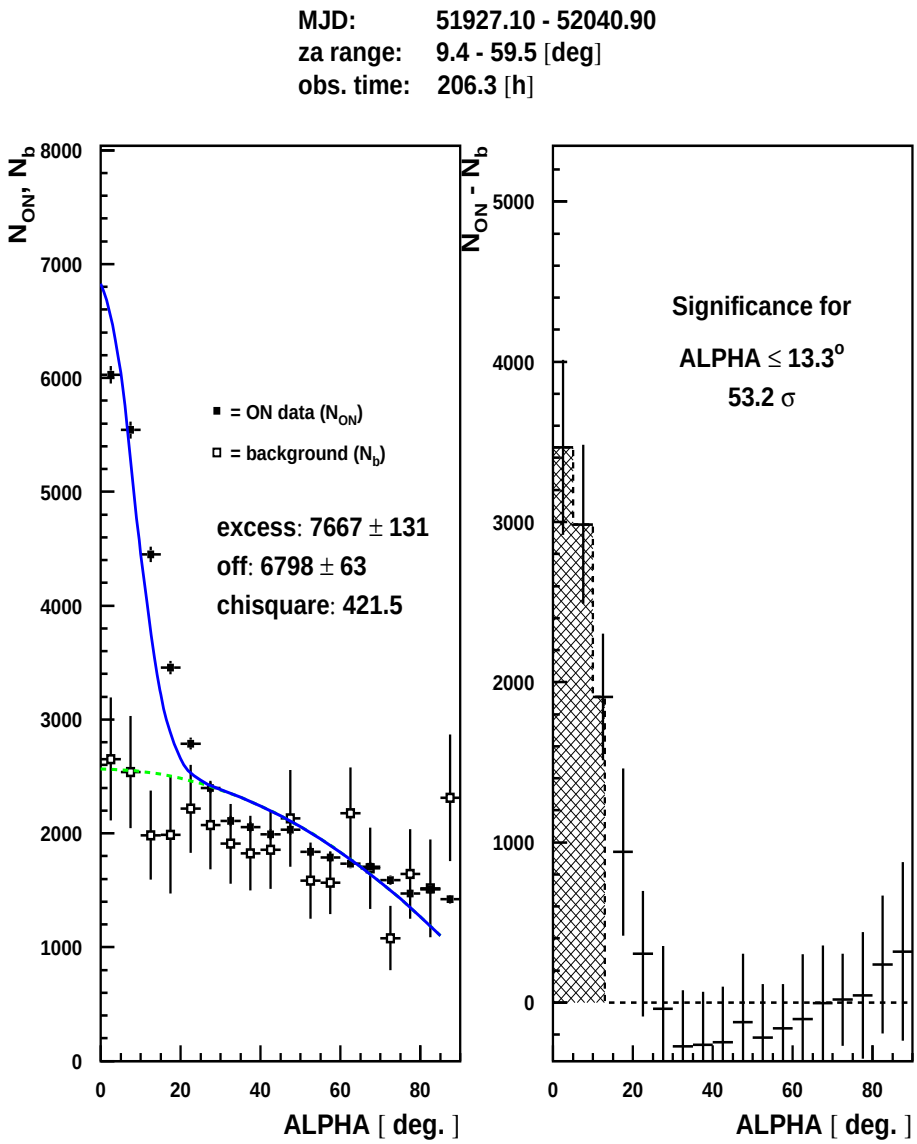


Figure A.15: Significance determined by the 'jacuzzi' program (dynamic selection cuts) as applied on the whole Mkn 421 dataset from February 2001 to April 2001 without pointing correction.

following section for this purpose.

For ***future improvements*** of the performance the author suggests to interested people to ***introduce a binning in energy*** and optimize the LDA in ***each*** energy bin separately. This will solve the difficulty of having ***different*** energy spectra for the ***background*** and the ***signal*** and will further ***increase*** the separation power.

A.4 New image parameters and cleaning algorithms

The LDA has been defined and can be used to check the separation ability of ***new*** image parameters. Each new parameter can be included as an ***additional parameter*** in the LDA.

There are ***two*** estimators that tell us about the ***separation power*** of the parameters. These are the ***discriminant power*** (Equ. A.60) and the ***final significance*** can be compared with different parameter sets, upon application to the data. With both we can ***judge*** whether an improvement has been achieved. Three things were tested here.

1. ***New*** image parameters were developed which describe more features in the image structure,
2. ***Weights*** were introduced in the calculation of Hillas parameters and
3. The ***cleaning level*** of the 'dynamical' image cleaning has been varied systematically in order to find the one which results in the highest separation power for the image parameters.

A.4.1 Introduction of weights in the calculation of Hillas parameters

Up to now, in the definition of the Hillas parameters presented above, Equ. A.25 to Equ. A.33the weight has always been assumed to be equal to the charge in one pixel ($w_i = q_i$). Here weights of the following form have been tested:

$$w_i = q_i^n \quad (NONE) \tag{A.65}$$

$$w_i = \frac{q_i^{n+0.5}}{\sqrt{q_i + \sigma_i^2}} \quad (REL) \tag{A.66}$$

In Equ. A.65 (these weights are denoted in the following as '***NONE***'), q_i is potentiate with an exponent n. This has the effect that pixels with ***more signal*** have ***more weight*** in the calculation of the HILLAS variances. This makes them less sensitive to noise fluctuations (coming from NSB).

Equ. A.66 (these weights are denoted as '***REL***') a weight is introduced that represents the ***statistical error*** of the signal in the pixel ($\sigma_{tot,i} = \sqrt{q_i + \sigma_i^2}$). A value 0.5 has been added to the exponent to keep the ***effective exponent*** in both cases ***equal***. The exponent become comparable in this manner. For a small NSB the weights 'REL' converge to the weights 'NONE'.

A program has been written which ***systematically compares*** the separation power of a MC Gamma dataset and a hadronic background dataset that was recorded by CT1. The ***NSB*** has been simulated in the MC gamma dataset to have a zenith angle distribution ***equivalent*** to that of the hadronic dataset (See section A.2.1).

The program tries systematically

1. Different ***image cleaning levels***,
2. Different ***cleaning types*** and
3. Four different ***exponents*** (n=0.5, 1.0, 1.5, 2.0).

This is done for ***both*** types of weights , 'NONE' and 'REL' (Equ. A.65and Equ. A.66).

To be able to judge the separation, ***two*** parameters were introduced. I called them:

1. ***Separation***

$$S = \frac{\langle p_G \rangle - \langle p_H \rangle}{\sqrt{\sigma_G^2 + \sigma_H^2}} \tag{A.67}$$

   which is the difference in the mean value of the parameter divided by its combined variance. It should be as ***large*** as possible for good separation. This parameter does not take into account the ***shape*** of the distributions.

2. The ***overlapping integral*** is ***sensitive*** to the shape of the distributions

$$\begin{aligned} O &= \frac{\int p_G(x)\, p_H(x)\, dx}{\int p_G(x)\, dx \int p_H(x)\, dx} \\ &= \frac{\sum_i \left(p_G^i\, p_H^i\right)}{\left(\sum_i p_G^i\right)\, \left(\sum_i p_H^i\right)} \end{aligned} \tag{A.68}$$

   It measures the ***overlap*** of the two distributions and should be as ***small*** as possible for good separation.

With the combination of these two parameters we have a ***tool*** to judge possible improvements in ***separation*** for different configurations.

The results for the parameter WIDTH, LENGTH, CONC are shown in Fig. A.17, Fig. A.16, Fig. A.18, Fig. A.19, and A.20. The exponent is expressed in different colors. Blue is $n = 0.5$, green is $n = 1.0$, pink is $n = 1.5$ and red is $n = 2.0$. Different cleaning algorithms are expressed in different marker styles and line styles. ***Circles*** and a ***continuous line*** represent ***'classic'*** cleaning, ***triangles*** and a ***dashed line*** are '***island'*** cleaning and ***stars*** with ***dotted line*** are ***'mountain'*** cleaning.

Separation of WIDTH, LENGTH and CONC

Tab. A.2 illustrates the ***cleaning levels*** for the ***best*** separations and overlaps for the ***two*** different cleaning algorithms, ***'classical'*** and ***'island'*** cleaning. The weights REL and NONE show almost ***identical*** results (see Fig. A.17, Fig. A.16, Fig. A.18 and Fig. A.19). The first two rows show the old classic version of HILLAS parameters. The cleaning level consists of two numbers, the 'image core limit' and the 'image border limit'. The best cleaning algorithms for WIDTH, LENGTH and CONC are printed in bold. A cut of $0.4 < DIST < 1.05$, $SIZE > 100$ and $12\,deg \leq Zenith\,angle \leq 22\,deg$ was applied beforehand ('precut').

As a conclusion can be stated the following:

1. There is ***no visible difference*** between the weights 'NONE' and 'REL'

2. ***Slightly reducing*** the cleaning level to 2.7/2.0 ***increases*** the separation a little for ***'classic'*** cleaning.

3. The ***best separation*** and ***overlap*** (for all three parameters) is achieved for the ***'island' cleaning method*** with a ***very low*** cleaning level of 1.0/0.3.

4. In case of the parameter ***WIDTH*** the separation ***increases*** by increasing the ***exponent*** n from $n = 1$ to $n = 1.5$.

5. In case of ***LENGTH*** $n = 1$ remains the best choice.