



HAL
open science

A multivariate analysis approach for the Imaging Atmospheric Cherenkov Telescopes System H.E.S.S

F. Dubois, G. Lamanna, A. Jacholkowska

► **To cite this version:**

F. Dubois, G. Lamanna, A. Jacholkowska. A multivariate analysis approach for the Imaging Atmospheric Cherenkov Telescopes System H.E.S.S. *Astroparticle Physics*, 2010, 32 (2), pp.73. 10.1016/j.astropartphys.2009.06.003 . hal-00616888

HAL Id: hal-00616888

<https://hal.science/hal-00616888>

Submitted on 25 Aug 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Accepted Manuscript

A multivariate analysis approach for the Imaging Atmospheric Cherenkov Telescopes System H.E.S.S

F. Dubois, G. Lamanna, A. Jacholkowska

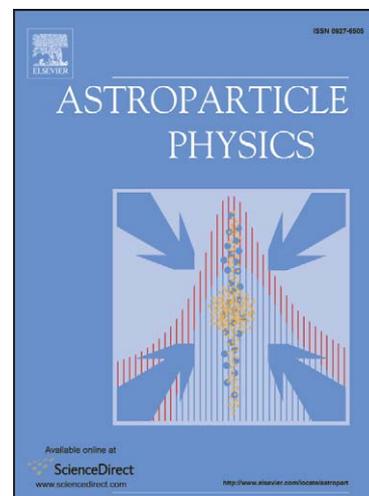
PII: S0927-6505(09)00096-6
DOI: [10.1016/j.astropartphys.2009.06.003](https://doi.org/10.1016/j.astropartphys.2009.06.003)
Reference: ASTPHY 1416

To appear in: *Astroparticle Physics*

Received Date: 23 March 2009
Accepted Date: 19 June 2009

Please cite this article as: F. Dubois, G. Lamanna, A. Jacholkowska, A multivariate analysis approach for the Imaging Atmospheric Cherenkov Telescopes System H.E.S.S, *Astroparticle Physics* (2009), doi: [10.1016/j.astropartphys.2009.06.003](https://doi.org/10.1016/j.astropartphys.2009.06.003)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



A multivariate analysis approach for the Imaging Atmospheric Cherenkov Telescopes System H.E.S.S.

F. Dubois^{*,a}, G. Lamanna^{*,a}, A. Jacholkowska^b

^a*LAPP, Laboratoire de Annecy-le-Vieux de Physique des Particules, UMR/IN2P3-CNRS, Université de Savoie, 9 Chemin de Bellevue - 74941, Annecy-le-Vieux, France*

^b*LPNHE, Laboratoire de Physique Nucléaire et de Hautes Energies, UMR/IN2P3-CNRS, Universités Paris VI & VII, 4 Place Jussieu - 75252 Paris, France*

Abstract

A multivariate particle classification approach is proposed and applied to the analysis of the data from H.E.S.S. (High Energy Stereoscopic System). The combination of results from the three shower reconstruction methods: Hillas, Model and 3D-Model, leads to a substantial gain in the discrimination power between photons and hadrons. The construction and use of a combined effective estimator improves by several factors the signal-to-background ratio which is extremely important in case of studies of the faint and extended sources. The results of this approach are presented for a typical set of sources. The consequent gain in the sensitivity is shown through a comparison to the H.E.S.S. published results.

Key words: Multivariate analysis, H.E.S.S., Gamma-rays, Cherenkov, IACT

1. Introduction

In the past decade, the High Energy Stereoscopic System (H.E.S.S.) successful observations [1] resulted in a multitude of revolutionary results in the VHE gamma domain for various types of galactic and extra-galactic sources. The H.E.S.S. experiment, an array of four of new generation Imaging Air Shower Telescopes (IACT) located in Namibia, is presently the most sensitive detector of the astrophysical VHE gamma-ray sources (> 100 GeV). The four telescope array system provides a sensitivity to a 1% of the Crab Nebula flux in ~ 25 hours observation. A fifth telescope (twice in size) is presently under construction and will improve the sensitivity of the H.E.S.S. array below 100 GeV. Till now, the H.E.S.S. collaboration has published the detection of more than 50 sources of different types: SNRs, PWNs, unidentified galactic sources and AGNs.

One limiting factor in the standard H.E.S.S. analysis is the identification of photons against the

hadron cosmic-rays which constitute a major part of the measured candidates. At present, the hadron discrimination procedures do not allow the morphological studies of the faint and extended sources or the detection of the very tiny fluxes expected from the exotic gamma-ray signals as from the galactic dark matter. To improve the hadron rejection, a promising approach is based on a use of the maximum information that is provided by the shower reconstruction and the Monte Carlo simulations in case of photons and hadrons. At present, three reconstruction methods are qualified in H.E.S.S.: Hillas parameter reconstruction [2], Model reconstruction [4] and 3D-Model reconstruction [5], which yield complementary discriminating variables for the hadron background suppression. Although most of the published analysis have been performed by applying the standard method of *Hillas parameterization*, with the advent of the H.E.S.S. telescopes with large mirrors and fine-pixel cameras together with the stereoscopy, new and more elaborate techniques have been introduced and also applied in a good fraction of the H.E.S.S. analysis:

- a *Model analysis* technique based on the comparison of the shower images with a more realistic semi-

*Corresponding author

E-mail addresses : duboisf@lapp.in2p3.fr (F. Dubois), giovanni.lamanna@lapp.in2p3.fr (G. Lamanna)

Preprint submitted to Elsevier

analytical pre-calculated model of the image [3] [4]. - a *3D – Model analysis* method based on a simple 3-dimensional elliptical reconstruction of the gamma-ray induced air shower [5].

In this paper, a combination of the discriminating variables from the three reconstructions is proposed, resulting in an estimator called later as X_{eff} , which was successfully used for the particle identification in LEP¹ analysis. Moreover, a construction of an estimator varying between 0 and 1, characterizing particle type, may lead in future to analysis with weighted event techniques. In the following, after a brief description of the three different reconstruction methods in section 2, a multivariate classification approach based on the combination of the various discriminating variables is defined and detailed in section 3. In section 4 the combined energy and direction estimators are discussed. In section 5, in order to validate the proposed procedure, a sample of astrophysical observation data published by H.E.S.S. are re-analyzed with X_{eff} estimator and the results are compared with those from the standard analysis.

2. The H.E.S.S. data analysis methods

2.1. Hillas analysis

The classical analysis method for IACT data was first introduced by M. Hillas in 1985 [2], where so-called Hillas parameters of the shower images in the camera focal plane are defined: the length L and the width W of the elliptical image; the total image amplitude (image size); the *angular distance* d between the center of the camera and the image center of gravity; the azimuth angle ϕ of the image main axis and the orientation angle α , as shown in (Fig. 1). The use of the Hillas parameters leads to a robust and efficient reconstruction and analysis method, as detailed elsewhere (see [6] for a review). In general, in single telescope observations, the image parameters *length* and *width* provide a way to estimate the shower direction, while the shower energy is usually estimated from the image *size* and nominal distance (d). The stereoscopic imaging technique [3] provides a simple geometric reconstruction of the shower: the source direction is given by the intersection of the shower image main axes in the camera. The energy is estimated from a

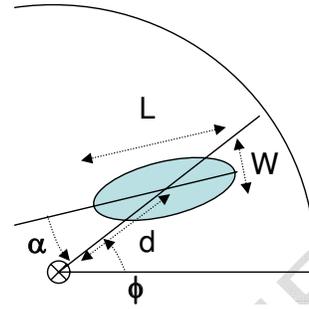


Figure 1: Definition of the geometrical parameters used for the Hillas shower reconstruction.

weighted average of each single telescope energy reconstruction. The discrimination of gamma events against the dominant background of the charged cosmic-ray atmospheric showers may be faced in different ways [3] [13] [14], taking the advantage of existing correlations between Hillas parameters and particle type. Among them the *Scaled Cuts* technique is usually applied and commonly used in the H.E.S.S. standard analysis. The geometric parameters of the shower image, width w and length l , are compared to the averaged expected values ($\langle w \rangle$, $\langle l \rangle$) and their variances (σ_w , σ_l) as obtained from Monte Carlo simulated gamma-ray data as a function of the image total charge amplitude q and the reconstructed *impact distance* ρ of the shower. According to this technique the main Hillas image parameters renormalized to the expected values are the *Scaled Width* (SW) and *Scaled Length* (SL):

$$SW = \frac{w(q, \rho) - \langle w(q, \rho) \rangle}{\sigma_w(q, \rho)}, \quad (1)$$

$$SL = \frac{l(q, \rho) - \langle l(q, \rho) \rangle}{\sigma_l(q, \rho)} \quad (2)$$

The scaled parameters are then easily combined in stereoscopic observations in *Mean Scaled Width*:

$$MSW = \frac{\sum_{tels} SW}{\sqrt{ntels}} \quad (3)$$

and *Mean Scaled Length*:

$$MSL = \frac{\sum_{tels} SL}{\sqrt{ntels}} \quad (4)$$

$ntels$ being the number of telescope of the IACT array (e.g. 4 in case of H.E.S.S.). These two vari-

¹The CERN particle physics experiment at LEP - the Large Electron Positron collider

ables are almost uncorrelated for gamma-ray showers (as it will be shown further in the text with Monte Carlo simulations) therefore are often combined in a single discriminating variable *Mean Scaled Sum (MSS)*:

$$MSS = \frac{MSW + MSL}{\sqrt{2}} \quad (5)$$

2.2. Model analysis

A more elaborate analysis technique pioneered by the work of the CAT (Cerenkov Air Telescope) collaboration and now extensively applied in the H.E.S.S. data analysis is the Model analysis whose details can be found in [4] and references therein. Here in short recall, the so-called Model method is based on the comparison between the registered shower images and a predicted pixel-by-pixel modeling of the expected image within a realistic detector simulation. The event reconstruction is based on a maximum likelihood fit of the shower direction and energy, with use of all available pixels in the camera, without requirement on the image cleaning. The probability density function (p.d.f.) of observing a signal S in a given pixel, under the assumption of an expected amplitude μ , a fluctuation of the pedestal σ_p (due to the night sky background and electronics) and a fluctuation of the single photo-electron (*p.e.*) signal $\sigma_s \approx 0.4$ (which matches the photomultiplier resolution) is given by $P(S|\mu, \sigma_p, \sigma_s)$, defined as:

$$P(S|\mu, \sigma_p, \sigma_s) = \sum_{n=0}^{\infty} \frac{e^{-\mu} \mu^n}{n! \sqrt{2\pi(\sigma_p^2 + n\sigma_s^2)}} \exp\left(-\frac{(S-n)^2}{2(\sigma_p^2 + n\sigma_s^2)}\right) \quad (6)$$

The reconstruction of the events *e.g.* the primary energy, the direction and the impact parameter of the detected photon (identically for the case of a single telescope or a stereoscopic array), is obtained through the maximization of a log-likelihood function:

$$L = 2 \sum_{i=1}^{Pixels} \log[P_i(S_i|\mu, \sigma_p, \sigma_s)] \quad (7)$$

In the Model analysis the separation between γ and hadron candidates relies on a *goodness - of -*

fit (G) variable. The average value of the log-likelihood can be calculated analytically following to:

$$\begin{aligned} \langle \ln L \rangle &= \\ &= \sum_{i=1}^{Pixels} \int_{S_i} P(S_i|\mu_i, \sigma_{p_i}, \sigma_{s_i}) \times \ln P(S_i|\mu_i, \sigma_{p_i}, \sigma_{s_i}) dS_i \\ &= \sum_{i=1}^{Pixels} [1 + \ln(2\pi) + \ln(\sigma_{p_i}^2 + \mu_i \times (1 + \sigma_{s_i}^2))] \quad (8) \end{aligned}$$

As the variance of $\ln L$ being close to 2, the *goodness - of - fit (G)* can be defined as a normal variable:

$$G = \frac{\langle \ln L \rangle - \ln L}{\sqrt{2 \times N_{dof}}} \quad (9)$$

where N_{dof} is the number of degree of freedom.

2.3. 3D Model analysis

A 3D Model-reconstruction method of the gamma-ray induced air-shower has been recently introduced and applied in the H.E.S.S. data analysis. The atmospheric shower is modeled as a Gaussian photosphere with anisotropic light angular distribution and then used to predict the collected light in each pixel of a camera. The main shower parameters, *e.g.* incident direction, shower core position on the ground, slant depth of shower maximum, average lateral spread of the Cerenkov photosphere and primary energy, are derived from the fit of the pixel content in the images. The *3D - Model Analysis* selects electromagnetic showers on the basis of their rotational symmetry with respect to the incident direction. Details on the *3D - Model Analysis* can be found in [5]. Monte Carlo simulation studies have demonstrated the dependency of the photosphere 3D-width (in units of radiation length) on the slant depth of the shower maximum. This property together with the assumption of a rotational symmetry of the electromagnetic shower allows to reject most of the background hadrons through a dimensionless and almost zenith angle independent quantity ω , the reduced 3D-width:

$$\omega = \frac{\sigma_T \rho}{D_S} \quad (10)$$

where σ_T is the transverse standrad deviation of teh 3D shower, ρ is the density of air and D_S is the column density at shower maximum (see [5] for furtehr details).

The ω discriminating variable, the reduced 3D-width, which shows excellent abilities for the photon/hadron separation, together with Hillas parameters, will contribute to the over-all combination of the discriminating variables as described in section 3. The three methods have been applied in the H.E.S.S. data analysis mainly for cross-check purposes while the Hillas analysis has been always considered as the reference standard method. In some cases a linear combination of the Hillas and Model approaches have produced competitive background rejection and the corresponding results have been published by the H.E.S.S. collaboration. At present, the need of improving the sensitivity of the experiment for more detailed studies, requires the combination of the complementary performances of these three methods. The terms of their comparison can be found elsewhere [4]. Here it is important to summarize the main differences among them:

- **Sensitivities:** The three methods show very similar sensitivities. If the Model and 3D Model analysis have a $\sim 20\%$ higher efficiency for gamma rays than the Hillas analysis they also keep more background events and are therefore less competitive in terms of signal-to-background ratio. The gamma efficiency as a function of the OFF-axis angle (distance of the shower axis to the center of the camera) is more flat for the Model and 3D Model analysis since they both do not rely on the images size as in the case of Hillas analysis.
- **Angular Resolution:** The Model analysis at low energy performs significantly better ($\sigma_{68} \simeq 0.07^\circ$ at 0.4 TeV) than other analysis ($\sigma_{68} \simeq 0.1^\circ$ for 3D Model and $\sigma_{68} \simeq 0.12^\circ$ for Hillas). The Hillas and 3D Model methods at higher energy takes over ($\sigma_{68} \simeq 0.05^\circ$ for 3D Model, $\sigma_{68} \simeq 0.07^\circ$ for Hillas and $\sigma_{68} \simeq 0.1^\circ$ for Model at 20 TeV). The angular distances to shower true direction (θ parameter) reconstructed by the three methods show low level of correlation due to different pattern reconstruction on ground.
- **Energy Resolution:** The Model analysis performs significantly better ($\Delta E/E \simeq 12\text{-}13\%$) for energy up to 20 TeV than other analyses ($\Delta E/E \simeq 30$ to 16% as a function of energy for 3D Model and $\Delta E/E \simeq 16$ to 20% for Hillas).

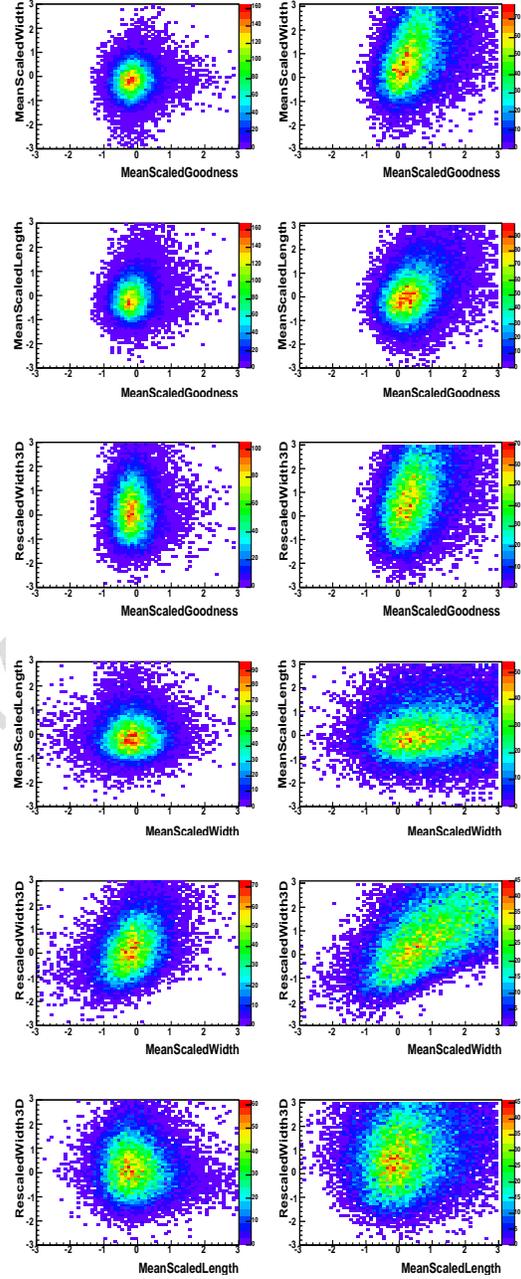


Figure 2: The reciprocal combinations of pair of discriminating variables among MSL, MSW, MSG and 3RW do not show any correlation for the *Gamma* sample (left panel) while a residual correlation exists for the background-*Hadron* sample (right panel).

3. The X_{eff} combined method

The classification and regression method approach (e.g. neural-network, random-forest, fisher),

usually applied in data analysis, combines parameters from a given reconstruction procedure (the shower image parameters in the case of the IACT analysis) which characterize differently signal and background events. In the presented approach (see [7] for more details), these image parameters, which are already powerful by themselves and applied to the H.E.S.S. analysis providing already excellent background discrimination for bright sources, will be combined in a single estimator. This procedure presents a two-fold aspect: a more powerful hadron background suppression as based on the maximum information obtained with three complementary reconstructions, and a single analysis parameter leading to a simplified study of applied cuts for a global optimization of the photon and hadron efficiencies. Therefore, in the following, the aim of this work will be to combine the main *gamma*-tagging discriminating variables of each of the three reconstruction methods with the purpose of exploiting their complementary discriminating power improving the *gamma*/hadron separation, as compared to the standard H.E.S.S. analysis results. The discussed multivariate method consists of a unique resulting discriminating variable X_{eff} , whose value, associated to each event, has the power of an event-by-event gamma-mistag probability (*hadroness*) estimator. The definition of the X_{eff} mis-tagging probability function, introduced in [8] [9] and [10], follows the relation:

$$X_{eff}(\{x_i\}) = \frac{\eta H(\{x_i\})}{(1-\eta)G(\{x_i\}) + \eta H(\{x_i\})} \quad (11)$$

where η is the mistag fraction of the gamma class of events (e.g. the relative background fraction):

$$\eta = \frac{N_{Hadron}}{N_{Hadron} + N_{Gamma}}, \quad (12)$$

where N_{Hadron} is the background number of events while N_{Gamma} the signal events in any data sample. $G(\{x_i\})$ and $H(\{x_i\})$ are the *global* multidimensional *probability density functions* (*p.d.f.*) for events in which the tag identifies Gamma (the right tag) and Hadron (the wrong tag).

In the following, in order to describe the application of the method and the resulting analysis performance in terms of gamma/hadron separation, two distinct data sets are used: *Gamma*, corresponding to Monte Carlo simulated data at different zenith and offset conditions and *Hadron*, firstly treated

with H.E.S.S. proton Monte Carlo, secondly analyzed with real off-source observation data sets, obtained in stereoscopic mode with at least three telescopes.

The discriminating variables x_i from the three reconstructions entering the X_{eff} combination are respectively: 1) the *Mean Scaled Width* (*MSW*), the *Mean Scaled Length* (*MSL*) for the Hillas reconstruction ([2]); 2) the *Mean Scaled Goodness* (*MSG*) for the Model reconstruction ([4]); 3) the *Rescaled Width* (*3RW*), for the 3D-Model reconstruction ([5]). The use of multidimensional global *p.d.f.s* describes properly the possible correlations between the discriminating variables, while they are quite well approximated by the product of the corresponding one-dimensional probability density functions, when the variables are not correlated:

$$X_{eff}(\{x_i\}) = \frac{\eta \prod_i h_i(x_i)}{(1-\eta) \prod_i g_i(x_i) + \eta \prod_i h_i(x_i)} \quad (13)$$

where $g_i(x_i)$ and $h_i(x_i)$ are the *gamma* and *hadron* respective probability density functions *p.d.f.s* of any discriminating variable x_i . The simplified assumption of the factorization of the *p.d.f.s* in the X_{eff} equation is fully justified for the case of *Gamma* sample due to the absence of any correlation (see figure 2) between discriminating variables, while a limited correlation is manifested by the variables in the case of the *Hadron* sample. However in the case under exam, these residual correlations

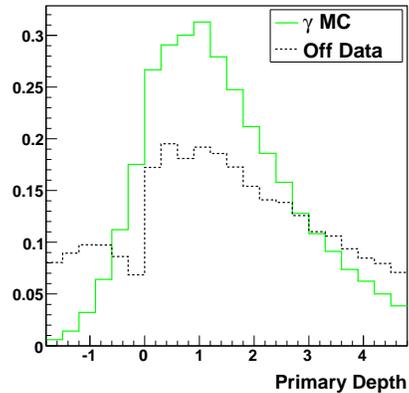


Figure 3: Reconstructed primary depth in radiation lengths for photon and proton Monte Carlo.

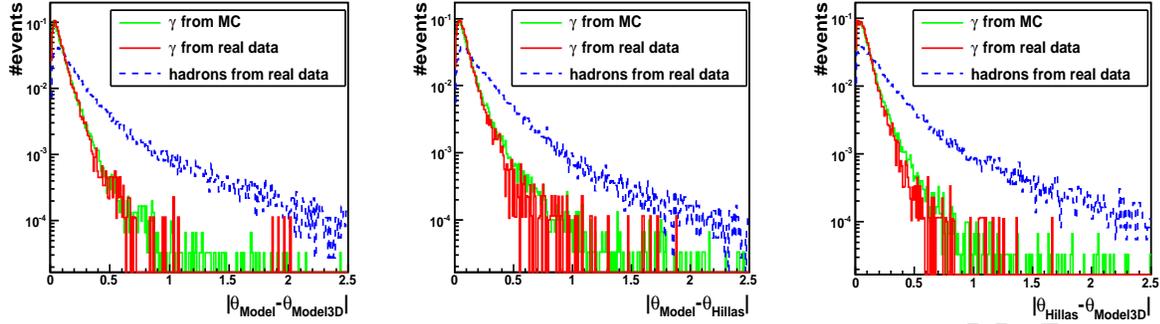


Figure 4: Dispersion between reconstructed directions for any pair of analysis methods: $|\theta_{Model} - \theta_{Hillas}|$, $|\theta_{Model} - \theta_{3D}|$ and $|\theta_{Hillas} - \theta_{3D}|$. The distributions for simulated *Gamma* (MC) events together with preselected *Gamma*-candidates (ON-OFF) and *Hadron*-candidates (OFF) samples are compared.

in the variable distributions for hadrons, have no measurable effects. During various tests performed in the qualification procedure of the X_{eff} method, the possible systematic effect induced by the use of the factorization of the p.d.f. functions has been checked and found negligible. This aspect is discussed later in the text (see section 4.3) when possible systematic effects induced by the X_{eff} are detailed. Therefore, all presented results are derived for simplicity with single discriminating variables p.d.f.s.

3.1. Event pre-selection

The first step in the analysis is a series of pre-selection criteria:

- Image cleaning requiring pixels in the image to be above a lower threshold of 5 p.e. and to have a neighbour above 10 p.e., and vice versa. This cut is the first step in the moment analysis and makes γ -ray showers approximating better a narrow elliptical shape, before application of the Hillas reconstruction method.

- Accepting events within a maximum offset angle from the camera centre of 2° allows to reduce any zenith angle dependent effects or variations in the NSB (Night Sky Background) level across the field of view as well as to limit the radial variation of the telescope system acceptance (see [15] for more details).

- The first cosmic-ray interaction depth generating showers in the atmosphere is one free parameter of the model analysis and the corresponding reconstructed value in radiation lengths (X_0) ranges from -1 and 4 for γ -ray showers. These two limits are

applied to the reconstructed primary depth distribution (Fig. 3) to exclude some fraction of hadrons.

- One more pre-selection cut is mainly devoted to guarantee a certain level of consistency among the different reconstruction methods. It concerns the distribution of the shower directions reconstructed by the three analysis methods. In figure 4 the distributions of the relative dispersions of the direction reconstruction for any pair of analysis methods: $|\theta_{Model} - \theta_{Hillas}|$, $|\theta_{Model} - \theta_{3D}|$ and $|\theta_{Hillas} - \theta_{3D}|$ are shown for the preselected samples of *Gamma*-MC events, *Gamma* and *Hadron* data-set events. The good agreement between the *Gamma*-MC and *Gamma* data-set distributions, together with different shape of the distribution of the *Hadron* data-set compared to *Gamma* events, allows to apply a pre-selection aimed to retain only those events for which the above relative dispersions are contained within a limit corresponding to $\sim 90\%$ containment for γ -ray events (namely a common upper limit of $\sim 0.5^\circ$).

- A final event-preselection criteria, able to decrease the background by more than 20% (removing the tails of the *Hadron* distributions), is based on the application of a lower and an upper cuts (e.g. -3 and +3) on the distribution of the discriminating variables (before computing from them the corresponding p.d.f.s): only those events for which all four variables have values contained between these limits are retained (see figure 5).

The distributions of the discriminating variables were studied as a function of the dynamical range of the H.E.S.S. system. In previous works, a good agreement between the distributions of the discrim-

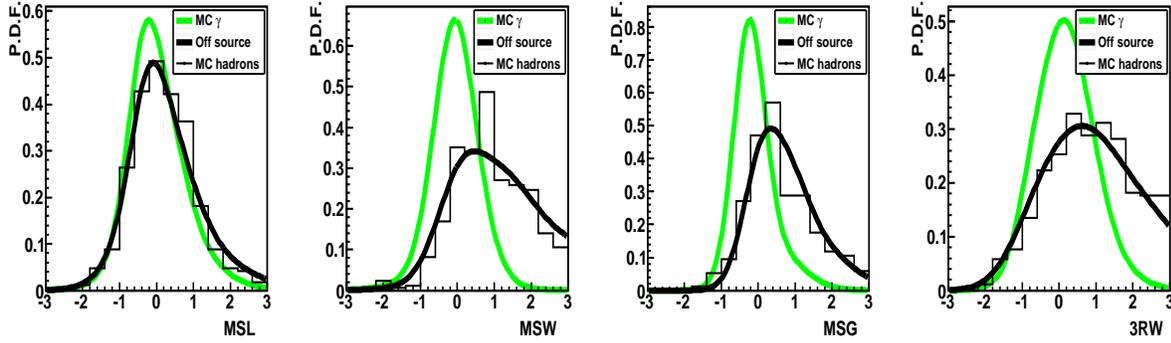


Figure 5: Probability density functions derived from the best fit resulting functions of the distributions of the discriminating variables (e.g. MSL, MSW, MSG and 3RW from left to right) for respectively Monte Carlo simulated *Gamma* and *OFF*-data sets. The cross represents the *p.d.f.* deduced from Monte Carlo simulated protons.

inating variables for simulated photons (*Gamma*) and ON-OFF real data were well established and discussed by [4] and [5]. Training samples of simulated photon and OFF-data-sets were examined as a function of the zenith angle of the observations (from 0° to 60°), energy range (from ~ 100 GeV to ~ 100 TeV), photon spectral index (from 1.8 to 3.4), minimum image size cut (40, 60, 80, 200 p.e.) and finally different offsets, the point source distance from the camera centre, (from 0° to 2°). The relative variations of the distribution (in terms of spread, normalization and mean value) are contained within a few percent, without any appreciable effect on the profile of the resulting X_{eff} estimator, neither in the final results of the analysis of a set of astrophysical sources considered for validation purposes (see section 4.3 and 5). A common choice was then assumed for the distribution of the four discriminating variables and a unique parameterization of the resulting *p.d.f.s* (fig. 5) was adopted and applied for all analysis.

Thus the resulting X_{eff} (*hadroness*) estimator applied in the following is defined as:

$$X_{eff} = \frac{\eta \prod_{1 \leq i \leq 4} h_i(x_i)}{(1 - \eta) \prod_{1 \leq i \leq 4} g_i(x_i) + \eta \prod_{1 \leq i \leq 4} h_i(x_i)} \quad (14)$$

where h_i and g_i are the *p.d.f.s* for *Hadron* and *Gamma* respectively as a function of the four discriminating variables: *MSL*, *MSW*, *MSG* and *3RW* (as shown in fig. 5). The comparison of the *p.d.f.* for *Gamma* and *Hadron* shows a good discriminating power of each variable. The Monte

Carlo simulations for protons provide *p.d.f.* distributions in good agreement with the OFF-data. In the following the OFF-data *p.d.f.* will be used. In figure 6 an example of the resulting events classification according to the value of the X_{eff} discriminating estimator is shown for *ON*-source, *OFF*-source and *ON-OFF* data samples. The X_{eff} classification allows to quantify the resulting event-by-event mistag probability providing a *hadroness* test statistic for the γ/h separation.

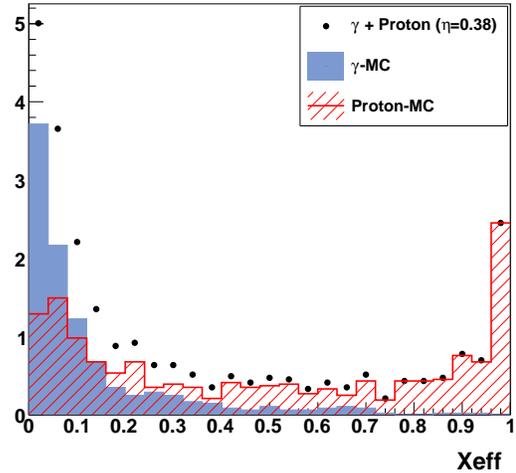


Figure 6: Example of the X_{eff} -*hadroness* (gamma-mistag) estimator distributions for two samples of simulated *Gamma* and *Hadron* events. A specific X_{eff} value is estimated for each event. The two samples are merged in a unique sample corresponding to $\eta = 0.38$ and analysed to validate the analysis method.

In the equation (14) the mistag fraction η can be defined a priori or estimated with a likelihood optimization. Since it corresponds to the background fraction related to the effective signal-to-background ratio, which is therefore a function of the specific data set under study (*e.g.* the fluency of the specific astrophysical source). In order to adopt the most appropriate η value and to check the compatibility between the X_{eff} (*hadroness*) estimator efficiency in the data and the Monte Carlo a likelihood function composed of the *p.d.f.s* of the discriminating variables was used here:

$$\mathcal{L} = \eta \prod_{1 \leq i \leq 4} h_i(x_i) + (1 - \eta) \prod_{1 \leq i \leq 4} g_i(x_i). \quad (15)$$

The optimal η value is the one which minimizes the sum of $-\mathcal{L}(\eta)$ over all events of the data sample:

$$\min \left[-2 \sum_{events} \ln(\mathcal{L}(\eta)) \right] \quad (16)$$

Successful tests of this method applied on samples with known amounts of *Gamma* and *Hadron* events (for which therefore η is precisely known) leads to a precise (less than 3% uncertainty) determination of the resulting optimal η . This result demonstrates as well the compatibility and the good agreement between data and Monte Carlo for photon sample (see figure 7). However as reported in [10] the effective power of the X_{eff} discriminating method is not given by the value of η . Even for a choice of an a-priori value of η , *e.g.* 50%, the

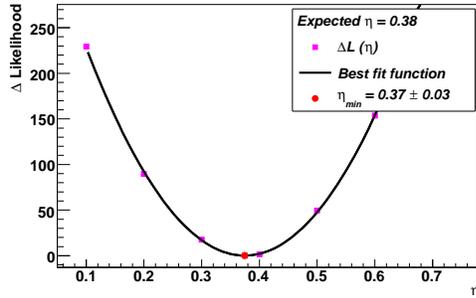


Figure 7: Two samples of simulated *Gamma* and *Hadron* events are merged and analysed to validate the analysis method. The maximization of the likelihood function for the optimal η value estimation provides a value (0.37 ± 0.03) consistent within 3% with known a priori relative background fraction (being 0.38).

use of the discriminating variables and the resulting X_{eff} function ensures that the γ/h separation is successfully achieved as well as a compatibility with the Monte Carlo results.

In Fig. 8 the distributions of the Signal-to-Background ratio (S/B) and the significance σ as a function of the X_{eff} value for a *Crab* data-set are shown. Five different distributions are compared for five respective different values chosen for the η factor. Whatever is the choice of the η value, the corresponding values of the maxima of these three analysis quality parameters are not affected. The optimisation of the events selection through a cut on the X_{eff} value corresponding to the maximum of these three quality parameters for a specific analysis, will be a function of η .

4. The H.E.S.S.- X_{eff} data analysis method

The X_{eff} data analysis approach aims to take advantage of the complementary properties of the three described reconstruction methods and to improve the background rejection.

4.1. Event selection and background rejection

The X_{eff} method, as described above, is then applied to the preselected data-sets. An unique value

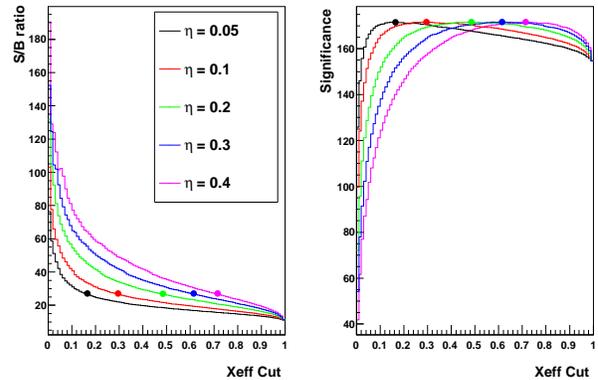


Figure 8: The signal-to-background ratios (S/B) and the significances N_σ as a function of the X_{eff} -Cut value for the *Crab* data sample. Five different distributions are compared for five corresponding different values assumed for η . While the choice of the optimal η affects the X_{eff} distribution and the consequent choice of the corresponding X_{eff} -Cut, the correlation between η and the optimal X_{eff} -Cut allows for invariant values of the corresponding quality parameters (see bullet markers).

Table 1: The X_{eff} cut selections as well as cuts on the distance (θ) from the reconstructed shower position to the source, the image amplitude and the distance from the centre of the field of view, optimized for various cases, as described in the text, are listed .

Configuration Name (e.g.)	η	$X_{eff}Cut$ Max.	θ_{cut}^2 (deg. ²) Max.	Amplitude (p.e.) Min.	Distance ($^\circ$) Max.
Flare (AGN)	0.1	0.3	0.02	40	2.0
Bright (Crab, MSH-1552)	0.1	0.3	0.0125	80	2.0
Faint (AGN, PWN)	0.5	0.3	0.0125	80	2.0
Search (J1849)	0.5	0.3	0.0125	200	2.0

of the X_{eff} , e.g. an $X_{eff}Cut$, is chosen to separate the γ -like events from the bulk of *hadron*-like candidates. The optimum selection cut yields the maximum values of the corresponding quality parameters for the data analysis of a given astrophysical source: namely the significance N_σ level achieved, as defined through the likelihood ratio method described in [11], which also depends on the square root of the observation time N_σ/\sqrt{t} , the Signal-to-Background ratio (S/B), the number of photons N_γ and the quality factor $Q_{f=\epsilon_\gamma/\sqrt{\epsilon_h}}$. The cut selection criteria generally depend on the energy spectrum of the sources and therefore their optimisation is achievable with the Monte Carlo simulations, requiring a possible optimization of the $X_{eff}Cut$ selection separately for much harder or much softer energy spectra. However, after having deeply investigated the systematic effects due to the spectral dependent choice, as a rule an universal $X_{eff}Cut = 0.3$ is applied for every analysis and for any kind of source, in order to preserve the *a priori* nature of the analysis. Such an approach should also be applied in source searches. Furthermore, if for the background estimation, the OFF-source data samples are used instead the Monte Carlo proton simulations, the application of the general and common background *pdfs* avoids a subjective approach to the source discovery procedures and the introduction of the additional biases in statistical optimization of the signal-to-background ratio. Table 2 summarize the combination of cut selections together with the choice of the nominal cuts on the distance (θ) from the reconstructed shower position to the source, the image amplitude and the distance from the centre of the field of view. Two different values of η are adopted: 0.1 for sources whose flux is intense as much as at least 10% the Crab flux, while $\eta = 0.5$ is the choice for all other fainter sources (as

well as for source search and morphological study of extended sources). The stability of the analysis results as a function of selection criteria optimisation is addressed later in the text (see section 4.3).

Figure 9 shows the resulting energy dependency of the photon effective area for the full H.E.S.S. array as a function of true Monte Carlo simulated photon energy and for observations at zenith angles of 20° , 45° and 60° and when the X_{eff} analysis method is applied. In order to establish and summa-

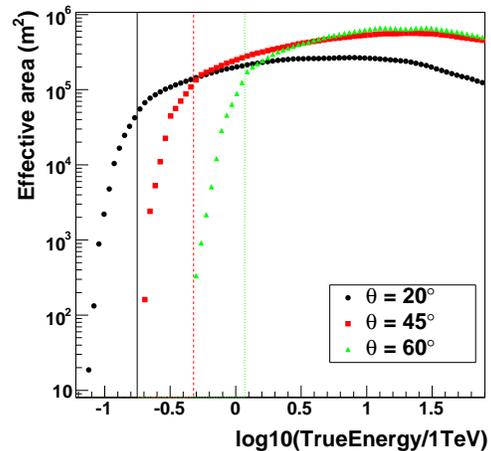


Figure 9: The effective collecting area of the full H.E.S.S. arrays for the X_{eff} analysis as a function of true Monte Carlo simulated photon energy and for observations at zenith angles of 20° , 45° and 60° . The vertical lines denote the safe energy thresholds for each zenith angle, increasing in zenith angle from left to right.

rise the competitive results achieved through the application of the X_{eff} analysis, the receiver operator characteristic (*ROC*) diagram, showing the fraction of accepted gamma (e.g. photon efficiency ϵ_γ) as a function of the corresponding fraction of

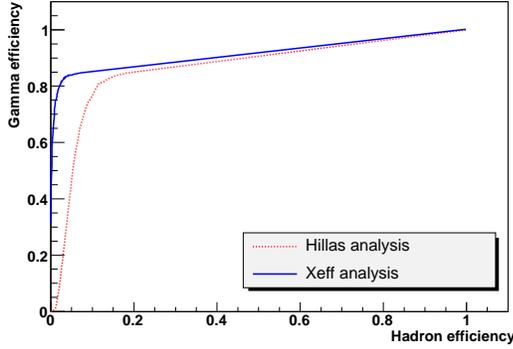


Figure 10: Receiver operator characteristic diagram applied to the γ/h separation for the case of the *standard* H.E.S.S. *Hillas* analysis and the X_{eff} analysis.

accepted hadron (e.g. proton efficiency ϵ_h), is presented (see figure 10). Same diagram is produced for both the *standard* H.E.S.S. *Hillas* analysis and the X_{eff} analysis. The *ROC* diagram for Model and 3D-Model analysis are comparable to the Hillas one within a few percents. In fact even if they have higher efficiency for γ -rays compared to Hillas ($\sim 20\%$), they also keep more background events, which in turns leads to very similar sensitivities for all three methods. The results here presented are based on Monte Carlo simulation analysis and validated with a H.E.S.S. data sample of the Crab observation. This comparison shows a general superiority achieved in the γ/h separation by combining the three different reconstruction methods (through the X_{eff} approach) against the standard selection based exclusively on optimised cuts on the Hillas scaled parameters. It is also useful to compare the different approaches in terms of the corresponding quality factors $Q_f = \epsilon_\gamma / \sqrt{\epsilon_h}$. In Fig. 11 the energy and zenith angle dependencies of the ratio between the X_{eff} and the Hillas quality factors ($Q_f(X_{eff})/Q_f(Hillas)$) are shown. For the computation of the $Q_f(X_{eff})$ and $Q_f(Hillas)$ the γ and *hadron* efficiencies are estimated starting from an identical original sample of simulated events pre-selected only according to the image cleaning requirements (e.g. pixels in the image to be above a lower threshold of 5 p.e. and to have a neighbour above 10 p.e.).

4.2. Combined direction and energy estimations

The X_{eff} approach is here proposed to optimize the γ -event selection, however since it is based on

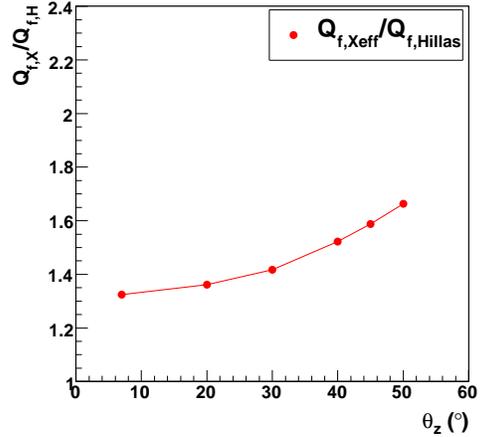
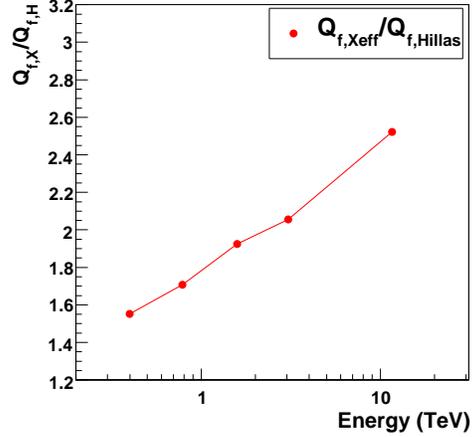


Figure 11: Ratios between the X_{eff} quality factor $Q_f(X_{eff})$ and the Hillas one $Q_f(Hillas)$. The $Q_f(X_{eff})/Q_f(Hillas)$ ratio is shown as a function of zenith angle (bottom) and reconstructed energy (top).

the combination of the three independent reconstruction methods, a combination of the correlated estimates of single physical quantities, namely the primary γ energy and direction, has also to be studied. After having selected γ candidate events through the X_{eff} analysis it is clearly desirable to quote a single quantity to characterize the energy and direction of the photon which is critical for the following spectral and morphological analysis of any astrophysical sources. As already shown, for the event pre-selection, a certain agreement was already required among the three reconstruction methods in terms of primary photon directions. This guarantees that the distributions of the reconstructed

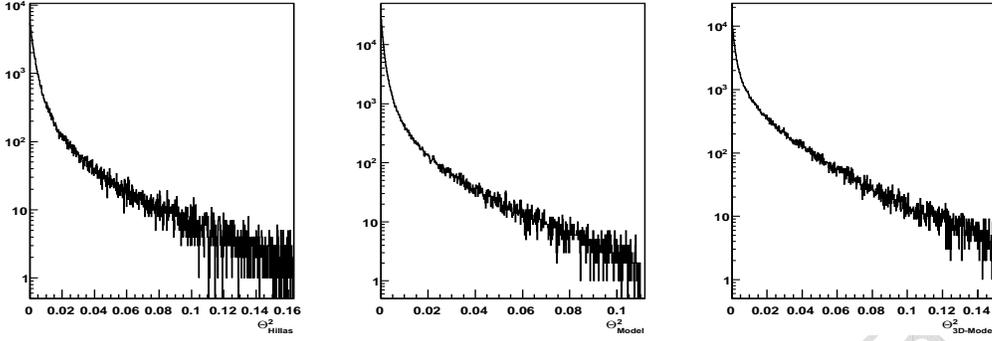


Figure 12: The distribution of the relative error in the reconstructed square angular distances to the shower true direction (θ^2) for Monte Carlo simulated 0.8 TeV energy γ sample and for the three analysis methods. The resulting σ_{68} containments, commonly used for the angular resolution estimation, are 0.08° (Hillas), 0.05° (Model) and 0.1° (3D-Model).

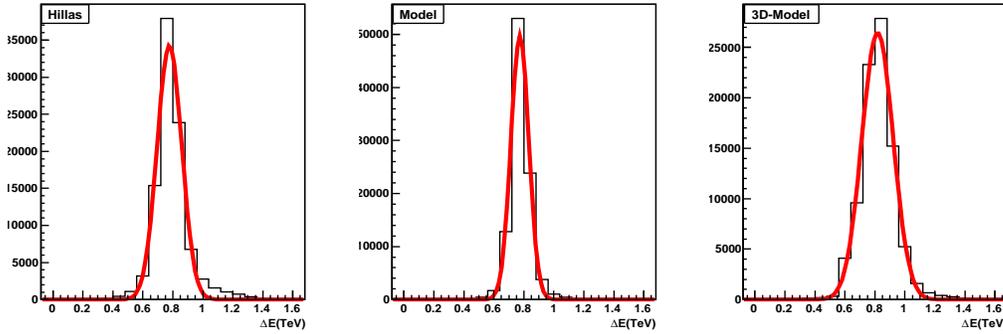


Figure 13: The distribution of the relative error in the reconstructed energy per event for Monte Carlo simulated 0.8 TeV energy γ for the three analysis methods. The root mean square widths of the energy distributions are 13% (Hillas), 9% (Model) and 19% (3D-Model), while the gaussian σ from the fit are respectively 9%, 7% and 15%.

observables have almost a Gaussian behaviour for the three methods (see fig. 12). This also happens for the primary photon energy. It is shown in figure 13 with an example of the distribution of the relative error in the reconstructed energy per event for Monte Carlo simulated 0.8 TeV energy γ for the three analysis methods. Choosing the physical quantity reconstructed by one method out of three would be one of the solutions. The alternative one is to combine different estimates by weighing each result x_i by a factor which takes into account the covariance matrices between estimates (see [12] for more details). Supposing the physical quantity x is measured using M ($M=3$ in our case) different methods, and x_i the unbiased estimate through the i th method, an improved \hat{x} estimate of the studied quantity is obtained by requiring a linear combina-

tion of the individual estimates:

$$\hat{x} = \sum_{i=1}^M w_i x_i, \quad (17)$$

where the weighting factors w_i for various estimates x_i are constant and are related to unbiased estimates, which means

$$\sum_{i=1}^M w_i = 1 \quad (18)$$

The variance of \hat{x} is thus defined as:

$$\sigma^2(\hat{x}) = \sum_{i=1}^M \sum_{j=1}^M E_{i,j} w_i w_j, \quad (19)$$

where $E_{i,j}$ are elements of the covariance matrix between estimates of the i th and j th methods, the diagonal elements giving the variance of the individual estimates (σ_H , σ_M and σ_{3D} , for Hillas, Model and 3D-Model methods respectively), while the off-diagonal elements describe the correlations between pairs of estimates:

$$\mathbf{E} = \begin{bmatrix} \sigma_H^2 & E_{H,M} & E_{H,3D} \\ E_{H,M} & \sigma_M^2 & E_{M,3D} \\ E_{H,3D} & E_{M,3D} & \sigma_{3D}^2 \end{bmatrix} \quad (20)$$

The *Best Linear Unbiased Estimates* (BLUE) method is then applied which consists in looking for the M values of w_i which provide the best unbiased estimates of the observable \hat{x} and minimize the corresponding variance $\sigma^2(\hat{x})$.

Such a method is specifically applied to the determination of the best estimate of the primary photon direction and energy (which correspond to the observable \hat{x}) of the γ -candidate-events as reconstructed by the $M=3$ reconstruction methods. The error matrix \mathbf{E} is estimated by Monte Carlo simulation, by applying to a γ sample simulated at conditions similar to the real data sample (e.g. for the Crab data analysis, MC photon spectrum generated with $\Gamma = 2.6$ and 50° zenith angle was considered) the same selection criteria as for real data. Therefore covariances between estimates is computed as a function of different data-set conditions: e.g. energy and spectral index, camera offsets and zenith angles. The \mathbf{E} error matrix determines the weights w_i of the three different estimates by satisfying the unbiased condition (in the equation 18) and minimizing the equation (19). Finally the obtained weights w_i are applied to combine together respectively the observables for the three different reconstructions. Figure 14 shows the angular resolution (68% containment) as a function of *Gamma* original energy for simulated showers. The events candidates are selected according to the X_{eff} analysis, while the chosen reconstruction is respectively the Hillas, Model, 3D-Model to be compared with the estimated combined directional measurement referred as X_{eff} . One can deduce the expected improvement when the X_{eff} results are compared with expected angular resolutions from the other three methods separately: at lower energy (0.1 to 1 TeV) the combined 68% containment angular resolution follows the most competitive Model reconstruction, while at higher energy (greater than few TeV) the Hillas and 3D-Model

methods perform better. For completeness in figure 14 the angular bias is also shown in order to provide the order of magnitude of relative systematic uncertainties in the direction reconstruction for different reconstructions. Here it is also important to remind that the reconstructed angular distances to the shower true direction (the θ parameter) from the three methods show low level of reciprocal correlation due to different pattern reconstruction on ground and which makes envisagable the combination of their estimates. The Hillas analysis reconstructs better the events that are well within the array, whereas the Model analysis performs better with events that are not too close to one telescope. The 3D Model does its best with high telescope-multiplicity events, which concentrate at the center of the array (see [4] and references therein for more details).

The same combinatorial approach is applied to obtain a better estimate of the reconstructed energy of the gamma candidates. Figure 15 shows the energy resolution and the energy bias as a function of *Gamma* original energy for simulated showers for the combined result (referred as X_{eff}) and the three independent reconstructions.

4.3. Systematic uncertainties

The systematic effects to the spectral measurements of a source, related to the detector inefficiencies and reconstruction have been studied with Crab-Nebula data [15]. The following table summarizes the most important factors contributing to the systematic error estimation as extracted from the H.E.S.S. publication [15]:

Table 2: Summary table showing the various estimated contributions to the systematic flux error [15].

Uncertainty	Flux	Index
MC Shower interactions	1%	
MC Atmospheric sim.	10%	
Broken pixels	5%	
Live time	1%	
Selection cuts	8%	0.08
Background est.	1%	0.01
Run-by-run variability	15% -	
Data set variability		- 0.05
Total	20%	0.09

At first order, these uncertainties affect in similar

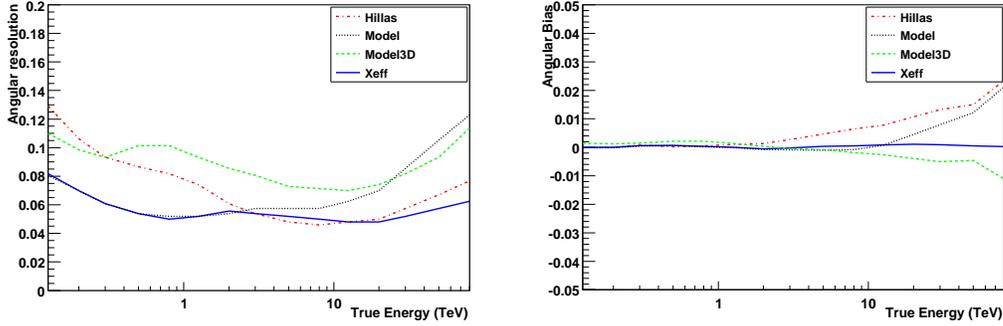


Figure 14: The angular resolution (left panel) and the angular bias (right panel) as a function of Γ energy, resulting from the X_{eff} analysis and estimated through the correlated combination of the directional measurements provided by the three reconstruction methods. The X_{eff} results are compared with expected angular resolution and bias from the three methods separately.

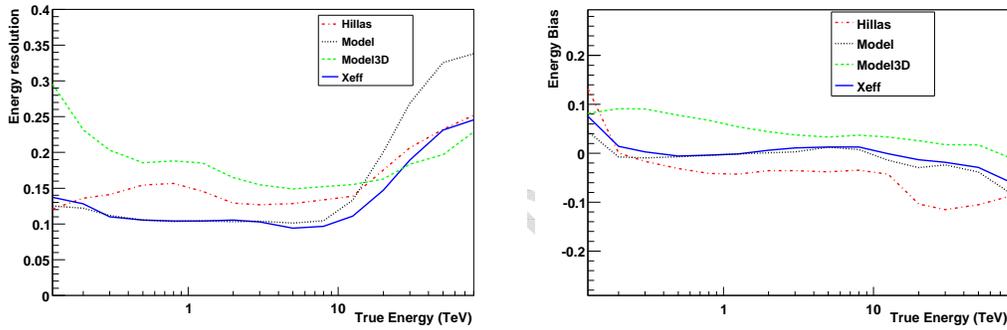


Figure 15: The energy resolution (left panel) and the energy bias (right panel) as a function of Γ true energy, resulting from the X_{eff} analysis and estimated through the correlated combination of the energy measurements provided by the three reconstruction methods. The X_{eff} results are compared with expected energy resolution and bias from the three methods separately.

way the results obtained with the three methods of shower reconstruction. In the following, the systematic effects specific to the use of the X_{eff} composed estimator are discussed.

The first point that should be underlined is related to the domain of the validity of the equation (11) with parameter η value comprised between 0.05 and 0.95 which implies the necessity of performing a strong pre-selection of events in the initial sample. This guarantees a correct behaviour of the formula and the lowest possible dependency on the background fluctuations - a pre-selected signal/background ratio of the order of 1 would be an optimal choice. The inherent systematic effects of each reconstruction method were checked with MC simulations and extensive studies as a func-

tion of the zenith angle and the offset for different types of sources have been performed in the past analyses as described in [2], [15], [4] and [5]. It is assumed in the presented analyses, that a proper simulation of the discriminating variables leads to a correct description of the X_{eff} construction in the MC simulations. This shows the importance of the description of the H.E.S.S. detector in the simulations and of the comparisons of the X_{eff} mis-tagging performances between MC and Data samples. Related to this point, it was necessary to compare the correlation factors concerning the discriminating variables produced by Hillas, Model and 3D-Model reconstructions which show, as well for MC as for Data, no correlation for photons and some residual values for the hadron background.

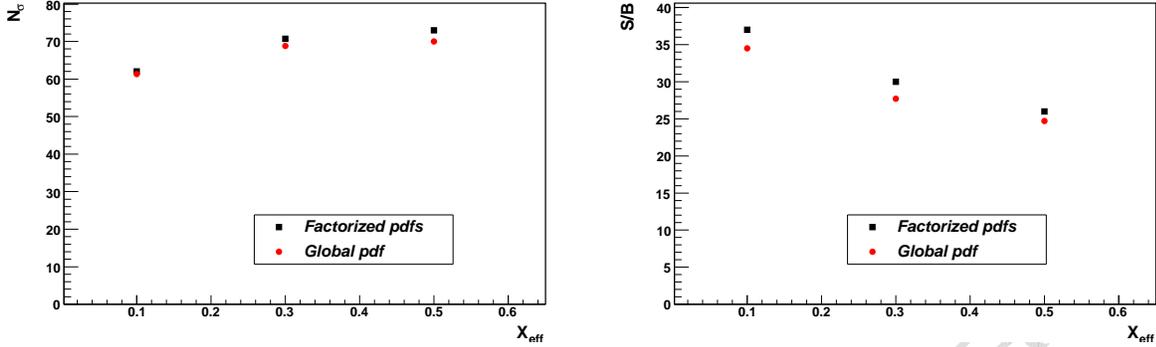


Figure 16: The significance levels N_σ and the S/B -ratios for three chosen X_{eff} -cut (0.1, 0.3 and 0.5) in the case the X_{eff} function is built through the factorisation of single $p.d.f.s$ and same estimators obtained with an X_{eff} using instead the single global correlated $p.d.f.$.

There are at least two possibilities to overcome the correlation problem: reconstruct X_{eff} with global multi-variable $p.d.f.s$ for photons and hadrons, or perform an iteration on X_{eff} variable by fitting the $p.d.f.$ from the X_{eff} distribution. The first solution has been studied and the differences in results on signal-to-background separation found with global, multivariable $p.d.f.s$ were below few %, to be considered here as negligible. In order to provide a quantitative example of such results, a test on a Crab data set is here considered. The differences between the S/B -ratios and significance levels N_σ for an X_{eff} function built through the factorisation of single $p.d.f.s$ and same estimators obtained with an X_{eff} using instead a single multi-variable correlated $p.d.f.$ are shown in figure 16. One can deduce the relative differences contained within few percents, second that the approach X_{eff} with the factorisation of single $p.d.f.s$ performs better as expected. The stability of the results was checked as a function of the X_{eff} -cut for three values: 0.1, 0.3 and 0.5, as also shown in figure 16.

Finally, a good test when determining possible systematic effects on the results of data analysis, is to study the implication of variation of cut selection (the X_{eff} cut) within a reasonable domain. The stable dependency of the S/B ratio and the significance level N_σ on the X_{eff} cut can be deduced by figure 8. In this example, always referred to a Crab data set, when considering X_{eff} cut variation at fixed η value or by varying η as well, one can deduce the low level systematic contribution due to the employed method corresponding to $\leq 10\%$ variation on the resulting S/B ratio and $\leq 3\%$ in

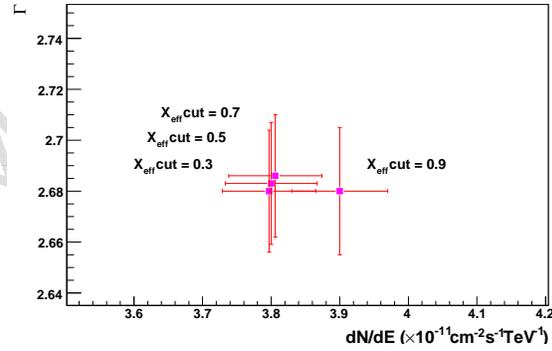


Figure 17: Stability plot of the results of the spectral analysis for a Crab data-set, by varying the value of the X_{eff} cut selection = 0.1, 0.5, 0.7 and 0.9.

the obtained N_σ . The stability plot of the spectral analysis results is shown in figure 17. Here the power-law best fit results for the same Crab data set are compared through the correlation between spectral index Γ vs differential flux dN/dE ($E > 1$ TeV) for different values of X_{eff} -cut = 0.1, 0.5, 0.7 and 0.9. The spectral analysis is stable within the statistical uncertainties, showing the total negligible systematic effect of the choice of X_{eff} cut.

In conclusion the systematic effects introduced by the multivariate X_{eff} approach is totally negligible compared to the main sources of uncertainties listed in Table 2.

5. The H.E.S.S.- X_{eff} data analysis validation

The improvement achieved with the X_{eff} data analysis in terms of sensitivity, reconstruction and background rejection, is validated by the re-analysis of some astrophysical sources already observed with the H.E.S.S. system. The X_{eff} data analysis results are here shown and compared to the corresponding results obtained with a standard Hillas analysis already published by the H.E.S.S. collaboration.

5.1. Point-like sources

A first cross-check required to validate the method is to study some point-like sources at different background condition, namely Galactic (e.g. the Crab nebula, SNR G0.9+0.1) and Extragalactic (e.g. PKS2155, ES0347 and H2356).

5.1.1. Crab-Nebula

The Crab nebula, a reference source in VHE gamma-ray astronomy, has a size comparable with the H.E.S.S. point spread function (PDF), therefore it is usually treated as a point-like source even if it belongs to the family of extended pulsar wind nebulae (PWN). First observed at high energies in 1989 the Crab was then re-observed and confirmed by a number of other experiments (see [15] and reference therein).

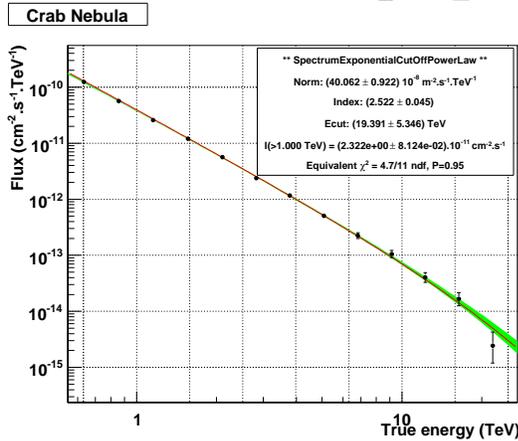


Figure 18: Energy spectrum resulting from the X_{eff} analysis for data set III, fit residuals to the power law and exponential cut-off fit are also shown.

The H.E.S.S. collaboration has published the analysis of the Crab-data based on different observations from 2004 to 2005 and for a total time

of 22.9 hrs after data quality selection [15]. A sub-sample of those data (quoted as the III data-set), with four telescopes, limited average zenith angle observation (47.9°) and longer live-time (10.6 hours), has been re-analysed by applying the X_{eff} method and with the choice of so called *Bright-cuts*. The results of this analysis are summarised in the Table 3. As expected the X_{eff} analysis applied to this data-set provides better results in terms of background rejection, with S/B ratio more than a factor 3 better than in Hillas-standard analysis and an improved significance of 22%.

The differential energy spectrum of the Crab signal as observed by the H.E.S.S. instrument, has been determined with a parameterization described by a power law with slope $\Gamma = 2.39 \pm 0.03_{stat} \pm 0.09_{sys}$ and an exponential cutoff at $(14.3 \pm 2.1_{stat} \pm 2.8_{sys})$ TeV. The integral flux above 1 TeV is $2.26 \pm 0.08_{stat} \pm 0.45_{sys} \times 10^{-11} \text{cm}^{-2} \text{s}^{-1} \text{TeV}^{-1}$. The energy spectrum of the Crab for the data-set III has been measured with the X_{eff} analysis. The results are consistent with the one published in [15] (see Table 4). The reconstructed X_{eff} integral flux is compatible with the Hillas-standard result within the statistical uncertainties, which indicates that the reconstructed flux is not strongly dependent on the details of the analysis method. The differential spectra are also in agreement in case of power law fit function plus exponential energy cutoff as well as in the case of a power law best fit function (PL), compared to the results of the published analysis on the full Crab data sample (ALL). Even if compatible within the uncertainties, the X_{eff} Crab spectrum (fig. 18) shows less steepening at an energy cut-off (E_c) of $(19.4 \pm 5.3_{stat} \pm 3.9_{sys})$ TeV, slightly higher than in the Hillas spectrum.

5.1.2. SNR G0.9+0.1

The H.E.S.S. instrument has detected for the first time a VHE (> 100 GeV) gamma-ray emission from the composite supernova remnant G0.9+0.1. The flux and the location of this point-like source make it an important case for the X_{eff} analysis validation purpose: it is one of the weakest source ever detected at TeV energies (with a flux representing only 2% of the flux from the Crab above 200 GeV); it is located in the region around the Galactic Center where the cosmic-ray and diffuse gamma-ray background make the source signal more difficult to detect. The H.E.S.S. collaboration has published results on the G0.9+0.1 emission

Table 3: Comparative results for the analysis of the Crab nebula, listed by data set along with excesses, significance and signal-to-background ratios. The X_{eff} analysis is competitive in terms of background rejection power when compared to standard Hillas analysis published results as well as the Model and 3D-Model analysis standalone.

Source	Data Set	Method	ON	OFF	α	excess	significance N_σ	S/B
Crab	III	<i>Hillas-stand.</i> (publ.)	4759	2417	0.20	4283	94.2	8.8
	III	<i>Model</i>	10079	37635	0.07	7293	99	2.7
	III	<i>3D-Model</i>	7460	15732	0.1	5958	99	3.8
	III	X_{eff} -Bright	5013	902	0.20	4833	115.1	26.8

Table 4: Flux and spectral measurements of the Crab, divided up by data set. The results of power-law plus an exponential cut-off best fits with the X_{eff} -Bright-cuts analysis for a data-set (III) is compared to published results (*Std.*) with *Hillas-Standard-cuts* analysis. The results of the power-law (PL) best fits are also compared. The quoted errors do not take into account the systematic effects.

Data Set	Method	I_0 (1 TeV) $\times 10^{-11}$ ($\text{cm}^{-2}\text{s}^{-1}\text{TeV}^{-1}$)	Γ	E_c (TeV)	$F_{>1\text{TeV}}$ $\times 10^{-11}$ ($\text{cm}^{-2}\text{s}^{-1}$)
ALL	<i>Std.</i>	3.76 ± 0.07	2.39 ± 0.03	14.3 ± 2.1	2.26 ± 0.08
III	<i>Std.</i>	3.84 ± 0.09	2.41 ± 0.04	15.1 ± 2.8	2.31 ± 0.10
III	X_{eff}	4.01 ± 0.09	2.52 ± 0.04	19.4 ± 5.3	2.32 ± 0.08
ALL	<i>Std.</i> (PL)	3.45 ± 0.05	2.63 ± 0.01		2.11 ± 0.03
III	X_{eff} (PL)	3.81 ± 0.07	2.67 ± 0.02		2.27 ± 0.06

based on a 2004 data set corresponding to a total live time of about 50 hours [16]. The Hillas standard analysis resulted in a highly significant detection with 13 σ . The dataset is fitted by a power law in energy with photon index $\Gamma = 2.40 \pm 0.11$. The X_{eff} analysis was conducted on exactly the same dataset and with the choice of *Faint-source cuts*. For cross-check a classic Hillas analysis through the application of *Standard cuts* and requiring a 200 p.e. threshold was also conducted. The comparison among the results of the two analysis methods are detailed in table 5. The gain in significance with the X_{eff} analysis is $\sim 30\%$, and the background rejection is improved of a factor ~ 3 , for a reduced number of excess events.

The H.E.S.S. collaboration has published the detection (and the discovery in some cases) of VHE γ -ray emission from distant extragalactic sources: AGNs, Blazars and BL Lac objects. For our validation purpose a sample of such sources is considered since they exhibits various level of spectral hardness and corresponding integral flux for a similar background level. In the follow the results of the

X_{eff} analysis are summarized.

5.1.3. The PKS2155-304 big-flare

The 28 July 2006 flare of the PKS 2155-304 blazar observed by H.E.S.S., for its peculiar fluency, energy spectrum, time variability and for the point-like size of the source, is an excellent prototype for the validation of the multivariate analysis. The H.E.S.S. published results are based on a standard analysis [17] of an equivalent exposure of 1.32 h live time at a mean zenith angle of 13° in the early hours of MJD 53944. As reported in the corresponding publication [17] the analysis was performed with Hillas method and *loose cuts* [15] (40 p.e. minimal amplitude; looser cuts on Hillas discriminating variables with consequently higher gamma statistics but lower background rejection power), yielding to an energy threshold of about 170 GeV. This approach could be sensitive to systematic bias due to important inalienable background on short time scale even if, at first sight, it could not be an issue since the average flux observed during this outburst correspondes to about 7 times the Crab flux. Furthermore, it would simplify the analysis avoiding

Table 5: Comparative results for the analysis of the SNR G0.9+01, listed by data set along with excesses, significance and signal-to-background ratios. The X_{eff} analysis is competitive in terms of background rejection power when compared to standard Hillas analysis.

Source	Method	ON	OFF	α	excess	significance N_σ	S/B
G0.9+01	<i>Hillas</i> (200 p.e.)	1069	10801	0.065	368	13 (publ. in [16])	0.5
	X_{eff} -Bright (200 p.e.)	348	1715	0.071	225	16	1.8

low-statistics issues on short time scale flux estimation. The standard analysis is based on ON-source data taken from a circular region of radius $\theta = 0.2^\circ$ centered on PKS 2155-304, which corresponds to a 99% containment of ON-OFF candidate events and in agreement with the angular resolution (68% containment) provided by the Hillas reconstruction method at low energies (see fig. 14).

The X_{eff} -analysis was conducted on exactly the same data-set and with the choice of *Flare cuts* applied to ON-source data taken from a circular region of radius $\theta = 0.14^\circ$ centered on PKS 2155-304, corresponding to 95% events containment according to the angular resolution provided by the combined method at low energies (see fig. 14). The comparison between the two approaches is shown in Table 6. In Fig. 19 the distributions of ON and OFF source events in θ^2 resulting from the Hillas analysis and X_{eff} analysis are compared. The vertical arrows denote the θ^2 selection cuts corresponding to respectively 90% and 95% containment of excess events after background subtraction. Both distributions, which are in perfect agreement with their expected point-spread functions, show the different level of background rejection achieved with the two methods as well as the improvement in terms of angular reconstruction when the X_{eff} method is applied: 90% of the excess γ -events are reconstructed within $\theta^2 = 0.02^\circ$, while they are within $\theta^2 = 0.075^\circ$ in the *Hillas* analysis.

The X_{eff} -analysis is applied to a selected sample of 8420 events, which is about 30% less candidates ON source than the sample selected for the Hillas analysis. The excess events selected by the two analysis differ each other for the same extent but with a comparable resulting level of significance for both methods. The X_{eff} approach shows also in this analysis a competitive background rejection power, resulting in about one order of magnitude higher value of the signal-to-background ratio. All

these considerations support the hypothesis that the measured gamma-ray flux from the PKS 2155-

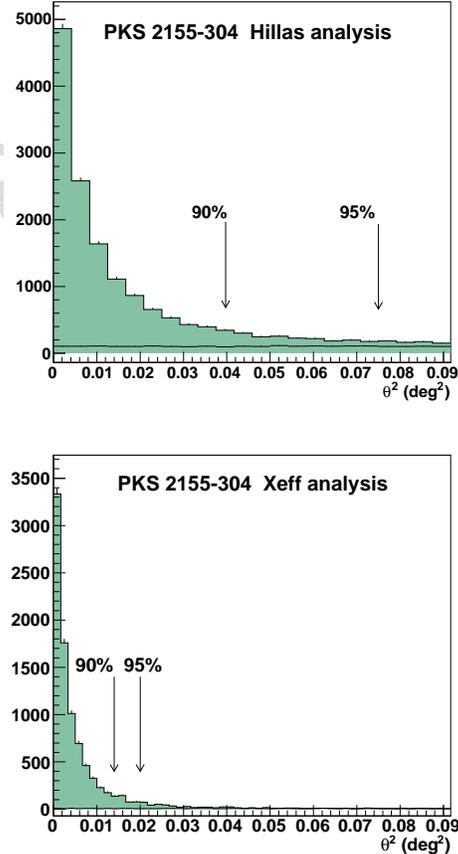


Figure 19: Distributions of ON and OFF source events in θ^2 for the PKS2155-304 MJD 53944 flare data sets resulting from the Hillas analysis (top) and X_{eff} analysis (bottom). The vertical arrows denote the θ^2 selection cuts corresponding to respectively 90% and 95% containment of excess events after background subtraction. Both distributions are in perfect agreement with their respective Monte Carlo derived point-spread functions.

Table 6: Comparative results for the analysis of the MJD 53944 PKS2155-304 flare, listed by methods (published *Hillas-loose* cuts analysis and *X_{eff}-Flare* cuts analysis) along with radius of integrated region around the source, excesses, significance and signal-to-background ratios.

Source	Method	θ_{cut}^2 ($deg.^2$)	<i>ON</i>	<i>OFF</i>	α	excess	N_σ	S/B
PKS2155-304	<i>Hillas-looses</i> [17]	0.04	12480	3296	0.215	11771	168	17
	<i>X_{eff}</i>	0.02	8420	399	0.20	8340	165	105

304 2006 flare as studied with the *X_{eff}*-analysis is based on a statistically significant sample of photon events with a high level of purity, e.g. lowest possible level of background contamination.

The spectral analysis has been performed with the *X_{eff}* method. The fit to the data with a simple power law function resulted in a spectral index $\Gamma = 3.35 \pm 0.02_{stat} \pm 0.1_{sys}$ compatible with the published one ($\Gamma = 3.19 \pm 0.02_{stat} \pm 0.1_{sys}$) and an average integral flux above 200 GeV equal to $\Phi_{>200GeV} = 1.66 \pm 0.05_{stat} \pm 0.33_{sys} \times 10^{-9} \text{cm}^{-2}\text{s}^{-1}$ consistent with the published result of $\Phi_{>200GeV} = 1.72 \pm 0.05_{stat} \pm 0.34_{sys} \times 10^{-9} \text{cm}^{-2}\text{s}^{-1}$.

The almost pure sample of photons from PKS2155-304 Big-flare obtained with *X_{eff}* selection may be further used in the energy-time analysis and calibration studies between data and Monte Carlo simulations.

5.1.4. 1ES 0347-121

The BL Lac object 1ES 0347-121 was observed for a total live time of 25.4 hours and at zenith angles ranging from 12° to 40° [18]. The H.E.S.S. observation results in an excess of 327 events, corresponding to a statistical significance of 10.1 standard deviations for an energy spectrum ranging from 250 GeV to 3 TeV and described by a power law with photon index $\Gamma = 3.10 \pm 0.23$. The same data set have been analysed through the *X_{eff}* approach resulting in 290 excess events corresponding to about 50% higher statistical significance and a factor of three improved S/B ratio (see Table 7).

5.1.5. H 2356-309

The BL Lac object H 2356-309 was observed with a total exposure of about 40 hours live-time. The H.E.S.S. collaboration has published the results of a standard Hillas analysis and a 3D-model standalone analysis, both allowing for corresponding 9.7σ and 11.6σ significance levels [19]. The

differential energy spectrum of this source is well-described by a power law with a photon index $\Gamma = 3.09 \pm 0.24$. The competitive results of the *X_{eff}* analysis of the same data set are shown in table 7: the high level of background rejection power allows to get a factor of two significance level for about the same number of excess events.

For the above studied sources the agreement between the results of spectral analysis with the standard Hillas method and the *X_{eff}*-approach, already shown in previous examples, is confirmed. This consolidates the applicability of the multivariate approach for the H.E.S.S. analysis.

5.2. Extended sources

The *X_{eff}* method is conceived with the purpose of improving the morphological studies of extended sources and the sensitivity to faint γ -emission thanks to its background rejection power together with the best estimation of the direction and energy of the photon candidates. To validate such achievement a test has been performed on some Galactic sources (e.g. Vela-X, MSH 15-52 and J1849)

5.2.1. Vela-X

The Vela-X cocoon is an extended region to the south of the Vela pulsar and part of the Vela supernova remnant (SNR) complex region which is source of non-thermal radiation. H.E.S.S. has observed a strong signal originated from this region. Vela X is an important prototype for the validation of the analysis method introduced here: it is one of the most extended source ever detected by a IACT observatory and it has a peculiar energy spectrum showing the first clear evidence of a peak from VHE gamma-ray sources. The published H.E.S.S. Hillas-analysis of Vela-X [20] resulted in an excess of 2152 events for a total live time of 16.4 hours, within an integration region of 0.8° around the best-fit centre

Table 7: Comparative results for the analysis of extragalactic sources (1ES 1101-232, 1ES 0347-121 and H 2356-309), listed by methods (published *Hillas* analysis and X_{eff} analysis) along with excesses, significance and signal-to-background ratios.

Source	Method	<i>ON</i>	<i>OFF</i>	excess	N_σ	S/B
1ES 0347-121	Hillas [18]	1167	9241	327	10.1	0.4
	X_{eff}	553	3067	290	14.7	1.1
H 2356-309	Hillas [19]	3776	35280	591	9.7	0.4
	3D-Model [19]	1706	13784	453	11.6	0.4
	X_{eff}	1213	6593	618	22.2	1.0

of gravity of the emission region. The background level in the Vela-X analysis (both the published and the X_{eff} ones) was estimated using the *ON*–*OFF* method, based in a sample of runs taken at similar zenith angles which contain no excess gamma-ray signal. This method allows the background level to be estimated in cases where the integration region from the source is comparable to the size of the field of view. For the published analysis the choice of *Hillas-hard* cuts (200 p.e. image amplitude threshold) was motivated by the need of improving the angular resolution and reducing systematic effects due to uncertainties in the background estimation. On the contrary the X_{eff} analysis with the choice of *Faint*-source cuts (implying only 80 p.e. minimal threshold) is supposed to be at least performant as the *Hillas-hard* analysis due to a competitive background rejection and an improved angular resolution without being obliged to increase the minimal image amplitude to 200 p.e.. In Table 8 the published and the X_{eff} results are compared: the X_{eff} *Faint* cuts (with a choice of integration region of 0.8° around the best-fit centre of gravity of the emission region) resulted in slightly reduced statistics of selected γ -events and $\sim 32\%$ improvement in the resulting significance and about almost factor of 3 increase of the signal-to-background ratio. Furthermore for completeness the X_{eff} analysis with 80 p.e. amplitude threshold was applied and the results are comparable to those obtained with *Hillas* analysis but at 200 p.e. amplitude threshold. In conclusion the X_{eff} analysis leads to global improvement in γ -events precision identification critical for morphological studies and in lower energy sensitivity for further spectral investigation.

These results validate and confirm the X_{eff} multivariate approach as a successful way to analyse faint extended sources.

The published energy spectrum of Vela-X as ob-

served by H.E.S.S., has been measured with a differential spectrum described by a power law with slope $\Gamma = 1.45 \pm 0.09_{stat} \pm 0.2_{sys}$ and an exponential cutoff at energy of $(13.8 \pm 2.3_{stat} \pm 4.1_{sys})$ TeV. The integral flux above 1 TeV is $1.28 \pm 0.17_{stat} \pm 0.38_{sys} \times 10^{-11} \text{cm}^{-2} \text{s}^{-1} \text{TeV}^{-1}$.

The energy spectrum as measured with the X_{eff} analysis is consistent with the published ones (see Table 4) but it is extended at lower energy (250 GeV against the 550 GeV) with competitive results obtained for 80 p.e. threshold. This last achievement is what one has to look forward when exploiting the introduced multivariate analysis potential. This result combined with a competitive angular resolution, which is better than 0.07 degrees even at the energy threshold of the measured flux (see fig. 14), allows to investigate more in details the morphology of extended sources and as a function of the energy regime of the non-thermal emission. Fig. 20 shows the gaussian smoothed sky map of the Vela-X cocoon region with smoothing width of 0.09° (as performed in the published map [20]) and obtained with the X_{eff} analysis with 200 p.e. threshold. The circle represents the H.E.S.S. integration region for the spectral measurement (0.8°).

The comparison with the published map [20] shows a good agreement as far as we apply a smoothing width compatible with the *Hillas* reconstruction point spread function (PSF). When a smaller smoothing width, more consistent with the improved X_{eff} PSF, and a lower energy threshold are applied, some new morphological features are revealed. Further model investigation on the Vela-X region morphology is therefore envisaged [7]. It goes far behind the aim of this paper and will be detailed in a dedicated analysis of the H.E.S.S. collaboration [21].

Table 8: Comparative results for the analysis of the 2004-2005 Vela X nebula cocoon observations, listed by methods (the published standard Hillas-*hard cuts* analysis and the X_{eff} -*Faint-cuts* analysis, both applied to a $\theta^2=0.64^\circ$ integration angle) along with excesses, significance and signal-to-background ratios.

Source	Method	ON	OFF	α	excess	N_σ	S/B
Vela X	Hillas- <i>hard</i> (200 p.e.) [20]	9610	7879	0.94	2152	16.7	0.29
	X_{eff} -Faint (200 p.e.)	3505	2230	0.9	1515	22	0.76
	X_{eff} -Faint (80 p.e.)	8789	8006	0.9	1945	13.5	0.27

Table 9: Flux and spectral measurements of Vela X. The results of the power-law plus an exponential cut-off best fits with the X_{eff} -*Faint* analysis (with 80 p.e. threshold) is compared to published results with *Hillas – hard* analysis (with 200 p.e. threshold). The uncertainties associated to each measurement are in order the statistical and systematic ones.

Method	E_{MIN} GeV	E_{MAX} TeV	Γ	E_c TeV	$F_{>1TeV}$ ($\times 10^{-11} \text{cm}^{-2} \text{s}^{-1}$)
Hillas	550	65	$1.45 \pm 0.09 \pm 0.2$	$13.8 \pm 2.3 \pm 4.1$	$1.28 \pm 0.17 \pm 0.38$
X_{eff}	250	65	$1.46 \pm 0.07 \pm 0.2$	$13.6 \pm 2.9 \pm 3.9$	$1.71 \pm 0.14 \pm 0.41$

Table 10: Comparative results for the analysis of the Supernova Remnant MSH 15-52 2004 observations, listed by methods (the published standard Hillas analysis and the X_{eff} -*Bright-cuts* analysis, along with excesses, significance and signal-to-background ratios.

Source	Method	ON	OFF	excess	N_σ	S/B
MSH 15-52	Hillas- <i>standard</i> (80 p.e.) [22]	3706	10154	1469	25	1.0
	X_{eff} - <i>Bright</i> (80 p.e.)	2069	6249	1371	39	2.0

5.2.2. MSH 15-52

The Supernova Remnant MSH 15-52 is an other extended prototype source chosen for validation purpose. It has been observed by H.E.S.S. and a γ -ray signal was detected at 25 sigma level during an exposure of 22.1 hours live time [22]. The published results by the H.E.S.S. collaboration concern a selected sample of data taken at a mean zenith angle of 37° and analysed using the standard Hillas method. The data are consistent with a power law energy spectrum from 280 GeV to 40 TeV and a photon index $\Gamma = 2.27 \pm 0.03 \pm 0.2$. The energy spectrum as measured with the X_{eff} analysis is consistent with the published one. In Table 10 the published and the X_{eff} results are compared: the X_{eff} *Bright*-source cuts resulted in slightly reduced statistics of selected γ -events and 50% improvement in the resulting significance and about factor of 2 increase of the signal-to-background ratio. The Hillas analysis

of the morphological study of MSH 15-52 was conducted with a higher energy threshold (~ 900 GeV) by applying a 400 p.e. amplitude threshold on the image size of the γ -candidates. This cut is meant to reduce the Hillas angular resolution to $\leq 0.07^\circ$ and improve the morphological investigation: the resulting extended emission reveals an elliptically shaped region around the pulsar PSR B1509-58, with semi-major axis $\sim 6'$ in the NW-SE direction and semi-minor axis $\sim 2'$. This morphology coincides with diffuse pulsar wind nebula as observed at X-ray energies by ROSAT [22]. As in the case of Vela-X also for MSH 15-52 the X_{eff} analysis leads to a global improvement in γ -events precision identification and to a lower energy improved sensitivity, critical for further morphological investigation: the improved X_{eff} PSF would allow for more detailed studies and at a lower energy regime since the 400 p.e. amplitude threshold is not required to achieve a comparable

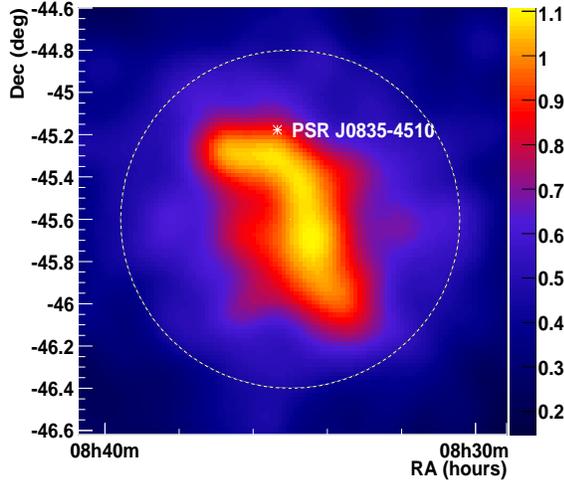


Figure 20: Gaussian smoothed sky map of the Vela-X cocoon region obtained with the X_{eff} analysis with 200 p.e. threshold and 0.09° smoothing width.

angular resolution (0.07°). Further model investigation on the MSH 15-52 PWN morphology is therefore envisaged [7].

Finally the X_{eff} analysis method has demonstrated competitive performance in searching for signal from a series of faint sources. As an example the IGR J18490-0000 was observed by the H.E.S.S. system during scans of the Galactic plane [23]: the preliminary results of Hillas (hard cuts and 200 p.e. image amplitude threshold) analysis have been published in [24]. The analysed dataset includes 49 hours of livetime with an average zenith angle of 30° . An excess was detected with a peak significance of $6.4-6.7\sigma$. The X_{eff} analysis (with *Searchcuts*) provides an excess detection for $>6\sigma$ average significance and a $\sim 8\sigma$ peak significance level.

6. Conclusions

A multi-variable-classification approach based on a composed estimator for the photon and hadron selection in H.E.S.S. experiment has been studied.

The proposed method aims to use the results of the three reconstructions of the photon candidate shower parameters by combining the identification estimators as well as the kinematical variables: angle and energy. This leads to optimizations of the signal-to-background ratios in a multivariable space

(here 4 dimensions) by use of the maximal information derived from the individual reconstruction methods. The followed approach ensures, by its principle, the best optimization of the quality factor and allows to reach the lowest energy threshold in the X_{eff} analyses. The gain in the quality factor is due also to a more stringent pre-selection possibility offered by comparing the reconstructed kinematical parameters, and to be applied before the X_{eff} combination. It has also been checked, that the error weighted combination of the kinematical variables provides improved resolution functions in a large energy range.

The results from X_{eff} analysis of a series of benchmark sources are very competitive as compared to the standard H.E.S.S. analyses based only on one reconstruction procedure. For these sources, the main improvement obtained with X_{eff} estimator concerns the signal-to-background ratio and the quality factor in general. For the point-like or the extended sources, keeping the photon efficiency on about the same level, the improvement factor varies between 3 and 10 depending on the analysis type, while the resulting significance increases of about 30%.

The future developments with combined estimators will follow different directions: inclusion of other possible discriminating variables, preferentially uncorrelated with those presented in this paper, and comparison and testing of the other commonly used procedures [25].

Finally, it should be underlined, that the X_{eff} estimator is well suited for the weighted event-by-event analyses and a Likelihood parameterization approach for the investigation of the theoretical modeling of the morphological properties of the extended sources.

Acknowledgments

The authors would like to thank the H.E.S.S. collaboration for the technical support and the fruitful discussions about the three different methods in use for the data analysis. The authors thank Prof. W. Hofmann, spokesperson of the H.E.S.S. Collaboration for allowing us to use H.E.S.S. data in this publication.

References

- [1] H.E.S.S. collaboration, <http://www.mpi-hd.mpg.de/hfm/HESS/pages/publications/>.

- [2] A. Hillas, Proc. 19th I.C.R.C.(La Jolla), Vol.3, 445 (1985).
- [3] A. Daum *et al.*, *Astropart. Phys.* **8**, 1 (1997).
- [4] M. de Naurois, Proc.Cherenkov 2005, astro-ph/0607247v1 (2005).
- [5] M. Lemoine-Goumard, B. Degrange and M. Tluczykont, *Astropart. Phys.* **25**, 195 (2006).
- [6] W. Hoffmann *et al.*, *Astropart. Phys.* **12**, 135 (1999).
- [7] F. Dubois, *Ph.D. Thesis*, LAPP, 2009.
- [8] D. Buskulic *et al.*, (ALEPH Collab.), *Phys. Lett. B* **377** (1996) 205.
- [9] D. Buskulic *et al.*, (ALEPH Collab.), *Phys. Lett. B* **384** (1996) 449.
- [10] D.E. Jaffe, F. Le Diberder and M.-H. Schune, LAL 94-67 and FSU-SCRI 94-101.
- [11] T.-P. Li & Y.-Q. Ma, *Astroph. J.* **272**, 317 (1983)
- [12] L. Lyons, D. Gibaut and P. Clifford, *Nucl. Instrum. Methods* **A270**, 110-117 (1998).
- [13] G. Mohanty *et al.*, *Astropart. Phys.* **9**, 15 (1998).
- [14] P.T. Reynolds *et al.*, *Astroph. J.* **404**, 206 (1993).
- [15] H.E.S.S. collaboration, F.Aharonian *et al.*, *Astron. & Astrophys.* **457**, 899 (2006).
- [16] H.E.S.S. collaboration, F.Aharonian *et al.*, *Astron. & Astrophys.* **432**, L25-L29 (2005).
- [17] H.E.S.S. collaboration, F.Aharonian *et al.*, *Astroph. J. Lett.* **664**, L71 (2007).
- [18] H.E.S.S. collaboration, F.Aharonian *et al.*, *Astron. & Astrophys.* **473**, L25-L2 (2007).
- [19] H.E.S.S. collaboration, F.Aharonian *et al.*, *Astron. & Astrophys.* **455**, 461 (2006).
- [20] H.E.S.S. collaboration, F.Aharonian *et al.*, *Astron. & Astrophys.* **448**, L43 (2006).
- [21] H.E.S.S. collaboration, F.Aharonian *et al.*, *Vela X follow-up 2009* to be published.
- [22] H.E.S.S. collaboration, F.Aharonian *et al.*, *Astron. & Astrophys.* **435**, L17-L20 (2005).
- [23] H.E.S.S. collaboration, F.Aharonian *et al.*, *Astroph. J.* **636**, 777 (2006).
- [24] R. Terrier *et al.* for the H.E.S.S. collaboration, Proc. Gamma-2008 Heidelberg (312) (2008).
- [25] Toolkit for Multivariate Data Analysis (*TMVA*), <http://tmva.sourceforge.net/>.