A multivariate analysis approach for the Imaging atmospheric Cerenkov Telescopes System H.E.S.S.
F. Dubois, G. Lamanna, A. Jacholkowska

To cite this version:

HAL Id: in2p3-00380934
http://hal.in2p3.fr/in2p3-00380934
Submitted on 5 May 2009

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

F. Dubois*, G. Lamanna* and A. Jacholkowska†

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

Abstract. We present a multivariate classification approach applied to the analysis of data from the H.E.S.S. Very High Energy (VHE) γ-ray IACT stereoscopic system. This approach combines three complementary analysis methods already successfully applied in the H.E.S.S. data analysis. The proposed approach, with the combined effective estimator \( X_{eff} \), is conceived to improve the signal-to-background ratio and therefore particularly relevant to the morphological studies of faint extended sources.

Keywords: Multivariate analysis, Hadron rejection

PACS: 95.75.pq

H.E.S.S. DATA ANALYSIS METHODS

Hillas analysis

The classical analysis method for IACT analysis, was first introduced by M. Hillas in a famous paper of 1985 [1], and it is based on the Hillas parameters of the shower images in the camera focal plane: the length \( L \) and width \( W \) of the elliptical image; the total image amplitude (image size). The discrimination of \( \gamma \) events against the dominant background of cosmic-ray atmospheric showers is provided by comparing the geometric parameters of the shower image, width \( w \) and length \( l \) to the averaged expected values \( \langle w \rangle, \langle l \rangle \) and their variances \( \sigma_w, \sigma_l \) as obtained from Monte Carlo simulated \( \gamma \)-ray data. According to this technique the main Hillas image parameters renormalized to the expected values are the Scaled Width \( SW \) and Scaled Length \( SL \), and then combined, for the case of stereoscopic observations, in the Mean Scaled Width:

\[
MSW = \frac{\sum_{\text{els}} SW}{\sqrt{ntels}}
\]  

and the Mean Scaled Length:

\[
MSL = \frac{\sum_{\text{els}} SL}{\sqrt{ntels}}
\]

Model analysis

A more elaborate analysis techniques pioneered by the work of the CAT (Cherenkov Air Telescope) collaboration and now extensively applied also in the H.E.S.S. data analysis is the Model analysis whose details can be found in [2] and references therein. The event reconstruction is based on a maximum likelihood method which uses all available pixels in the camera. In the Model analysis the separation between \( \gamma \) candidates and hadrons is done by a goodness – of – fit \( G \) variable, computed from the maximum likelihood value:

\[
G = \frac{\langle \ln L \rangle - \ln L}{\sqrt{2 \times N_{dof}}}
\]

and the easily combined in stereoscopic observations in Mean Scaled Goodness:

\[
MSG = \frac{\sum_{\text{els}} G}{\sqrt{ntels}}
\]

3D Model analysis

One more original image analysis techniques has been recently introduced and applied to the H.E.S.S. data analysis. It consists on a 3D-reconstruction method of the \( \gamma \)-ray induced air-shower. 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 the cameras. A scaled variable computed from the air-shower width is used for discrimination between \( \gamma \) candidates and the hadrons: \( \text{RescaledWidth3D} \). The \( 3D - \text{Model Analysis} \) selects electromagnetic showers on the basis of their rotational symmetry with respect to the incident direction. More details can be found in [3].
The proposed 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 estimator. More detailed description of the approach can be found in [4]. The definition of $X_{eff}$, introduced in [5] (and references therein), is given below, where $G$ and $H$ are the multi-dimensional probability density for $x_i$ value of each discriminating variable, respectively associated to Gamma and Hadrons, for any discriminating variable $x_i$.

$$X_{eff}(\{x_i\}) = \frac{\eta H(\{x_i\})}{(1-\eta)G(\{x_i\}) + \eta H(\{x_i\})}$$ (5)

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

$$\eta = \frac{N_{Hadrons}}{N_{Hadrons} + N_{Gamma}},$$ (6)

where $N_{Hadrons}$ is background number events while $N_{Gamma}$ the signal events in any data sample. $G(\{x_i\})$ and $H(\{x_i\})$ are the multi-dimensional probability density distributions (pdfs) for events in which the tag identifies Gamma (the right tag) and Hadrons (the wrong tag).

The multidimensional pdfs include properly the possible correlations between the discriminating variables, while, when the variables are not correlated, the multidimensional distributions are approximated by the product of the one-dimensional probability density distributions. The discriminating variables which we deal with when the three analysis methods are combined together are respectively: 1) The Mean Scaled Width (MSW), the Mean Scaled Length (MSL) for the Hillas analysis; 2) The Mean Scaled Goodness (MSG) for the Model analysis; 3) The Rescaled Width (3RW) for the 3D analysis. 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 will be used: Gamma corresponding to Monte Carlo simulated data at different zenith and offset conditions and Hadrons corresponding to real observation data set off-source obtained in stereoscopic mode with at least three telescopes. In Fig. 1 the distributions of the pdfs for the four discriminating variables are shown for Gamma and Hadrons data sets. A preselection of the discriminating variables is applied by requiring the corresponding distributions to be contained between a lower and an upper limits (e.g. -3 and +3). This allows to remove the tails of the Hadron distributions. The good agreement between simulated photons (Gamma) and ON-OFF real data are well established and are discussed in [2] and [3].

![FIGURE 1. Probability density functions derived from the best fit functions of the discriminating variable distributions for respectively simulated photons (Gamma) and OFF data (Hadrons) samples.](image)

**FIGURE 1.** Probability density functions derived from the best fit functions of the discriminating variable distributions for respectively simulated photons (Gamma) and OFF data (Hadrons) samples.

**FIGURE 2.** Example of the $X_{eff}$-hadroness ($\gamma$-mistag) estimator distributions for samples of ON, OFF, and ON-OFF data. A specific $X_{eff}$ value is estimated for each event.

In Fig. 2 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 $\gamma$/h separation.
The H.E.S.S. data analysis through the application of the $X_{\text{eff}}$ approach aims to take advantage of the complementary properties of the three described analysis methods in order to improve the background rejection and the resulting significance. Therefore the consistency among the shower directions reconstructed by each of the three analysis methods is also required. In Fig. 3 the distributions of the absolute dispersions of the direction reconstruction for any pair of analysis methods: $|\theta_{\text{Model}} - \theta_{\text{Hillas}}|$, $|\theta_{\text{Model}} - \theta_{3D}|$ and $|\theta_{\text{Hillas}} - \theta_{3D}|$. The preselected (within the chosen lower and upper limits of the corresponding discriminant variables) Gamma (Monte Carlo) and Hadrons (OFF events from Crab data sets) samples are compared.

The best $\gamma/h$ separation is then achieved by following the $X_{\text{eff}}$ method as described above and then optimizing a unique upper limit on the $X_{\text{eff}}$ acceptable value, e.g., an $X_{\text{eff, Cut}}$, to clean-up the $\gamma$-like reconstructed events from the bulk of hadron-like ones.

The optimum selection cuts yields the maximum significance for a given source correlated to the maximum value of the Signal-to-Background ratio. The cuts’ selection criteria generally depend on the energy spectrum of the sources, however as a rule, in order to preserve the a priori nature of the analysis an universal $X_{\text{eff, Cut}} = 0.3$ is always applied. Then in order to better approximate the relative background contamination for a given source the value of $\eta$ is fixed to 0.2 for a Crab-like source and for those sources with flux which is $>20\%$ the flux of the Crab while the choice $\eta=0.1$ is adopted for the case of flaring events as it is the case of the 2006 exceptional flare of PKS2155-304. In the majority of all other cases: source whose flux is some $\%$ of the Crab flux and in particular when conducting source searches in the Galactic sky survey, the $\eta$ value is also universally fixed to 0.5.

RESULTS

In order to establish and summerize the competitive results achieved through the application of the $X_{\text{eff}}$ analysis, the receiver operator characteristic diagram showing the fraction of accepted $\gamma$ as a function of the corresponding fraction of accepted hadron is presented (see Fig. 4). Same diagram is produced for both the standard H.E.S.S. Hillas analysis and the $X_{\text{eff}}$ analysis. The results here presented concern the H.E.S.S. data sample of the Crab observation. This comparison shows a general superiority achieved in the $\gamma/h$ separation by combining the three different reconstruction methods (through the $X_{\text{eff}}$ approach) against the most applied selection based exclusively on optimised cuts on the Hillas scaled parameters.

![FIGURE 3](image1.png)

FIGURE 3. Distributions of the absolute dispersions of the direction reconstruction for any pair of analysis methods, e.g. $|\theta_{\text{Model}} - \theta_{\text{Hillas}}|$, $|\theta_{\text{Model}} - \theta_{3D}|$ and $|\theta_{\text{Hillas}} - \theta_{3D}|$. The preselected (within the chosen lower and upper limits of the corresponding discriminant variables) Gamma (Monte Carlo) and Hadrons (OFF events from Crab data sets) samples are compared.

![FIGURE 4](image2.png)

FIGURE 4. Receiver operator characteristics diagrams applied to the $\gamma/h$ separation for the case of the standard H.E.S.S. Hillas analysis and the $X_{\text{eff}}$ analysis.
TABLE 1. Results from standard and $X_{eff}$ analysis on Crab data

<table>
<thead>
<tr>
<th>Data set</th>
<th>Method</th>
<th>$N_{ON}$</th>
<th>$N_{OFF}$ ($\alpha$)</th>
<th>Excess</th>
<th>Significance</th>
<th>S/B</th>
<th>$\Phi$ ($10^{-11}.cm^{-2}.s^{-1}$)</th>
<th>$\Gamma$</th>
<th>$E_{Cut}$ (TeV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>II</td>
<td>std</td>
<td>1976</td>
<td>1579(0.2)</td>
<td>1667</td>
<td>53.2</td>
<td>5.2</td>
<td>2.48 ± 0.16</td>
<td>2.30 ± 0.06</td>
<td>8.4 ± 1.2</td>
</tr>
<tr>
<td>II</td>
<td>$X_{eff}$</td>
<td>1551</td>
<td>248(0.2)</td>
<td>1501</td>
<td>64.8</td>
<td>30.3</td>
<td>2.55 ± 0.19</td>
<td>2.46 ± 0.09</td>
<td>10.7 ± 3.7</td>
</tr>
<tr>
<td>III</td>
<td>std</td>
<td>4759</td>
<td>2417(0.2)</td>
<td>4283</td>
<td>94.2</td>
<td>8.8</td>
<td>2.31 ± 0.10</td>
<td>2.41 ± 0.04</td>
<td>15.1 ± 2.8</td>
</tr>
<tr>
<td>III</td>
<td>$X_{eff}$</td>
<td>4238</td>
<td>733(0.2)</td>
<td>4091</td>
<td>106.3</td>
<td>27.9</td>
<td>2.34 ± 0.10</td>
<td>2.54 ± 0.05</td>
<td>17.7 ± 5.5</td>
</tr>
</tbody>
</table>

The $X_{eff}$ analysis method gives a better hadron rejection with the same $\gamma$ efficiency as it is shown on Fig. 4. Comparing results from Crab data analysis (see Tab. 1), we obtain significance and signal to background ratio improvements, losing few excess events.

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 Pr. W. Hofmann, spokesperson of the H.E.S.S. Collaboration for allowing us to use H.E.S.S. data in this publication.

REFERENCES