

Xeff analysis method optimization to enhance IACTs performances

C. Trichard*

LAPP, Université Savoie Mont Blanc, CNRS/IN2P3, Annecy-le-Vieux, France

E-mail: trichard@lapp.in2p3.fr

A. Fiasson

LAPP, Université Savoie Mont Blanc, CNRS/IN2P3, Annecy-le-Vieux, France

E-mail: fiasson@lapp.in2p3.fr

G. Maurin

LAPP, Université Savoie Mont Blanc, CNRS/IN2P3, Annecy-le-Vieux, France

E-mail: maurin@lapp.in2p3.fr

G. Lamanna

LAPP, Université Savoie Mont Blanc, CNRS/IN2P3, Annecy-le-Vieux, France

E-mail: lamanna@lapp.in2p3.fr

The seek of high precision analyses in γ -ray astronomy leads to the implementation of multivariate combination, benefiting from several reconstruction methods. Such analysis, called X_{eff} , was developed for the H.E.S.S. data using three shower reconstruction methods. This paper presents the improvement granted to this analysis by refining the distribution calculation of discriminant variables, considering observation conditions, and adding new variables in the X_{eff} combination. The efficiency of the analysis is presented using simulations and real data. A comparison with the standard analysis *model++*, for a typical set of sources, shows a significant gain in sensitivity.

*The 34th International Cosmic Ray Conference,
30 July- 6 August, 2015
The Hague, The Netherlands*

*Speaker.

1. Introduction

Identification of photons against hadron cosmic rays is very challenging for Imaging Air Shower Telescopes such as H.E.S.S.. A multivariate method, X_{eff} , has been developed to improve this separation [1]. The aim is to collect the maximum information provided by the air shower reconstruction and the Monte-Carlo simulations for photons and hadrons. Originally the X_{eff} analysis combined three reconstruction methods, the reconstruction of Hillas parameters, and two 3D reconstructions: a semi-analytical one, referred hereafter as *model* and its improved version *model++* [3], and an analytical one, referred as *3D-model* [4]. Since the *model* reconstruction provides a very good angular resolution and energy reconstruction, it has been used and X_{eff} serves as an estimator to separate γ and hadron events. The performance resulting from the combination of *Hillas* parameters, *3D-model*, the latest version of *model* (*model++*) and other discriminating variables are reported below.

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

The three shower reconstruction methods applied so far in the HESS data analysis are briefly described in this section.

2.1 Hillas analysis

The first method applied to the H.E.S.S. data makes use of the *Hillas* parameters [7]. These parameters are extracted by fitting an ellipse to the image. The direction of the particle is determined by the orientation of the ellipse and its energy by both the total image amplitude and the reconstructed impact parameter of the shower. The hadron/ γ discrimination is performed using scaled variables. The length and width (geometric parameters of the fitted ellipse) are scaled with the mean value given by Monte-Carlo simulations.

2.2 3D model analysis

A 3D-modeling [4] of the air showers has been developed to benefit from the array of telescopes through the stereoscopic information. This parametrization leads to a quantity called "3D-width". This parameter of the fit refers to the transverse standard deviation of the Gaussian distribution of the shower with respect to the main axis. This variable has a discrimination power between γ rays and hadron induced showers. Scaling the 3D-width by the depth of shower maximum (*reduced 3Dwidth*), enhances the background rejection capability.

2.3 Semi-analytical model analysis

The semi-analytical *model* analysis has been developed for H.E.S.S. data analysis [3]. This method compares the images of the atmospheric showers to predicted images given by a semi analytical *model*. This prediction is stored in a look-up table and a log-likelihood maximization is done over all the pixels. The parameters of the primary particle are given by the most probable set of images. The background and signal separation is performed through the goodness-of-fit variable. Then this variable is rescaled using the same method as for the *Hillas* analysis.

3. The X_{eff} analysis

3.1 X_{eff} estimator

The discriminating parameters, described above, are already powerful by themselves and applied to the H.E.S.S. data give already good background rejection for bright sources. The aim of this approach is to combine the information produced by the three analysis methods. All image information will be collected in a single discriminating estimator, X_{eff} . This variable, associated to each event, has the power of an event-by-event gamma-mistag probability estimator. The definition of this estimator, introduced by [8]- [9], follows the relation

$$X_{eff} = \frac{\eta \prod_j h_j(x_j)}{\eta \prod_j h_j(x_j) + (1 - \eta) \prod_j g_j(x_j)}$$

where $h(x_j)$ and $g(x_j)$ are the probability density functions (*PDF*) of variable j for hadron and γ events respectively, $\eta = \frac{N_b}{N_b + N_\gamma}$ is the assumed relative background fraction events, N_b and N_γ being the number of background and γ events in data sample.

3.2 Input variables for combination

The first version of the X_{eff} analysis within the H.E.S.S. analysis framework [1] used four variables : the mean scaled width and length from *Hillas* reconstruction, the rescaled width from *3D-model* and the mean scaled goodness from *model* reconstruction. In this updated version of X_{eff} the latest variable from *model*, the mean scaled shower goodness (*MSSG* see [3]), has been substituted to the mean scaled goodness.

There are almost no correlation for γ rays and partial correlation for hadrons between the mean scaled shower goodness and the other variables described above. Thus the *MSSG* can be used efficiently as a variable in X_{eff} analysis. Similar results have been obtained by [1] between the remaining variables used in this analysis. Adding *MSSG* instead of Goodness increases significantly the discrimination power. In order to quantify the benefits of taking into account the reconstructions other than *model++* in this multivariate analysis, the performance of X_{eff} will be compared in the next sections to the *model++* analysis.

In order to improve the discriminant power of the analysis, 3 variables were added. The directions reconstructed are slightly different for one method to another. These differences spread to higher values for hadrons [1] providing a discrimination power. Therefore, the differences between reconstructed directions ($\Delta\theta_{Hillas-Model}$, $\Delta\theta_{Hillas-Model3D}$ and $\Delta\theta_{Model-Model3D}$) have been introduced as supplementary PDFs in the X_{eff} estimator.

3.3 Training strategy

The approach of multivariate analysis is to use all information provided by the various shower reconstructions. To reduce the simulation uncertainties, we used real data to determine the hadron distributions for each variable. A sample of data (obtained in 2004 and 2005) without γ -ray sources in the field of view ("*OFF*") and far from the galactic plane, to avoid the galactic γ -ray background, were taken. These data come from AGN observations at different zenith angles and different distances from the center of the telescope cameras (offset).

Unfortunately there are no free background γ -ray sources available in the H.E.S.S. data. However it has been shown that γ -ray simulations reproduce faithfully the data. We used Monte-Carlo simulations with an offset from 0.5° to 2.5° to compute γ -ray distributions.

The *PDFs* computation is different from the previous version of X_{eff} analysis. In order to compute the *PDF* as close as possible to the real data, our simulation and *OFF* data samples have been separated in 6 energy bins and 7 zenith angle bins. Air shower development slightly depends on the azimuth angle. The simulation sample has been further separated into two different sets corresponding to azimuth angle of 0° and 180° . The lack of statistic prevents making such separation for hadrons. Small differences were expected for the *mean scaled* variables but the $\Delta\theta$ distributions should strongly depend on the energy and zenith angle. The aim is to select the *PDFs* corresponding to the observation conditions to perform the analysis. Including these customizations in the training improves the discrimination power of X_{eff} . This training method increases significantly the number of *PDFs* to compute, up to 882, which requires an automated procedure for a proper derivation from the original MC or OFF data distribution. The kernel estimation method from the package *RooFit* has been used. It ensures a determination of *PDFs* close to the original distribution without assumption on their shapes.

3.4 Event preselection

A set of pre-cuts is usually applied in the H.E.S.S. data analysis for two reasons : to select events that will be correctly reconstructed and to reject a significant part of hadron events without rejecting γ events. The analysis presented here uses the *model* reconstruction. The same pre-selection cuts as applied for the *model* analysis were chosen. They comprise the distance from the image barycenter to the camera center, the minimal charge of the image and the likelihood in the pure NSB hypothesis. It has been shown that the primary interaction of the particle scaled in terms of photon radiation length could be used as a pre-cut to reject an important fraction of hadron events with a low impact on γ rays [1]. Events with a reconstructed primary depth between -1 and 4 were kept. Events with values out of selection range for at least one of the variables were flagged as background.

3.5 Simulated efficiencies

To probe the performance of X_{eff} applied on H.E.S.S. data we analyzed simulated γ events and OFF data. Figure 1 shows an example of γ -ray efficiency versus the background rejection for several Zenith and Energy bins. A comparison with the *model++* analysis, which is generally used for H.E.S.S. data, is added. X_{eff} has always greater γ efficiency than *model* for a given background rejection in each bin. The standard H.E.S.S. analyses reject around 95% of background. For such rejection, X_{eff} provides $\sim 10\%$ more γ events than *model++* (see Figure 1).

3.6 Determination of discrimination cuts

Once an event passed through the pre-selection described above, we still have to determine the nature of the primary particle. We use the X_{eff} value to flag this event as hadron/ γ -ray-like. The cut to apply on the X_{eff} estimator is arbitrary.

According to the definition of X_{eff} , η is a purity estimator. It extends from 0 for background-free to 1 for no γ events data samples respectively. Four values for η have been defined depending

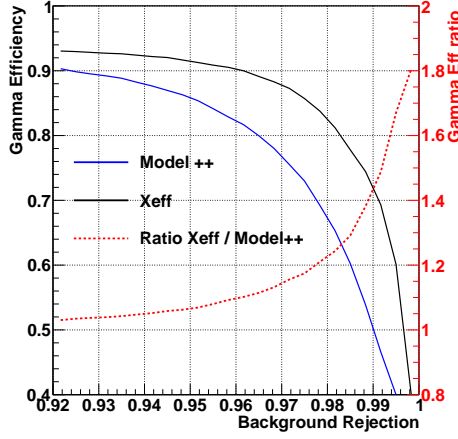


Figure 1: The blue and black lines show the γ efficiency of *model* [3] and X_{eff} respectively for a zenith angle between $15^\circ - 25^\circ$, an energy between 2 - 5 TeV and a muon efficiency of 100%. The ratio of X_{eff} γ efficiency over *model*'s is shown in red. The scale is at the right side of the plot.

on the brightness of the source. This provides 4 different analyses adapted for specific source type. A cut on the X_{eff} variable is then applied. A choice has been done to keep at least the same γ efficiency as *model++* across energy and zenith ranges as wide as possible. According to simulations, a single cut value of 0.3 is applied for all values of η (i.e. source brightness). This provides a good compromise between significance of the signal and γ -ray excess. Applying a unique cut for all η values leads to different background rejection and γ efficiency. Indeed, high value of η (e.g. 0.7) corresponds to high background rejection whereas for $\eta = 0.1$ the γ efficiency is at maximum. So for faint sources, the X_{eff} analysis focuses on signal significance enhancing the detection capability, while for bright sources the γ -ray excess is favored. The X_{eff} analysis parameters are summarized in Table 1.

Table 1: Summary of cuts applied in the X_{eff} analysis depending on the brightness of the source.

Brightness	$\Phi(1\text{TeV}).10^{-12}$ $cm^{-2}.s^{-1}.TeV^{-1}$	η	X_{eff}
Very Bright	> 10	0.1	< 0.3
Bright	1 ... 10	0.2	< 0.3
Medium	0.5 ... 1	0.4	< 0.3
Faint	< 0.5	0.7	< 0.3

3.7 γ - hadron efficiencies

We estimated the γ efficiencies using Monte-Carlo γ rays for the X_{eff} cuts defined in previous section. We used the bins defined above for the computation of the *PDFs*. The γ efficiency for $\eta = 0.1$ is $>85\%$ for most part of Zenith/Energy bins. The γ efficiency for standard H.E.S.S. analyses is generally $<70\%$. While for $\eta = 0.7$ the lower γ efficiency ($>60\%$ in most bins) is comparable to other H.E.S.S. analyses. For this configuration the high level of hadron rejection

(>96%) and a good γ efficiency provide a powerful analysis to extract faint signals and claim new detections.

3.8 Effective Area

Figure 2 presents the effective area of X_{eff} and *model++* analyses. The results for $\eta = 0.1$ and 0.7 and for 3 zenith angles are shown. X_{eff} analysis provides a better γ acceptance for almost all the H.E.S.S. energy range and for all zenith angles. As expected the X_{eff} effective area is lower for $\eta = 0.7$ but is still better or similar to *model++*'s. The splitting of data and MC samples during the training of X_{eff} , see section 3.3, does not introduce visible effects on the effective area shape. These results demonstrate the gain in sensitivity provided by the multivariate X_{eff} analysis compared to the standard most powerful H.E.S.S. analysis.

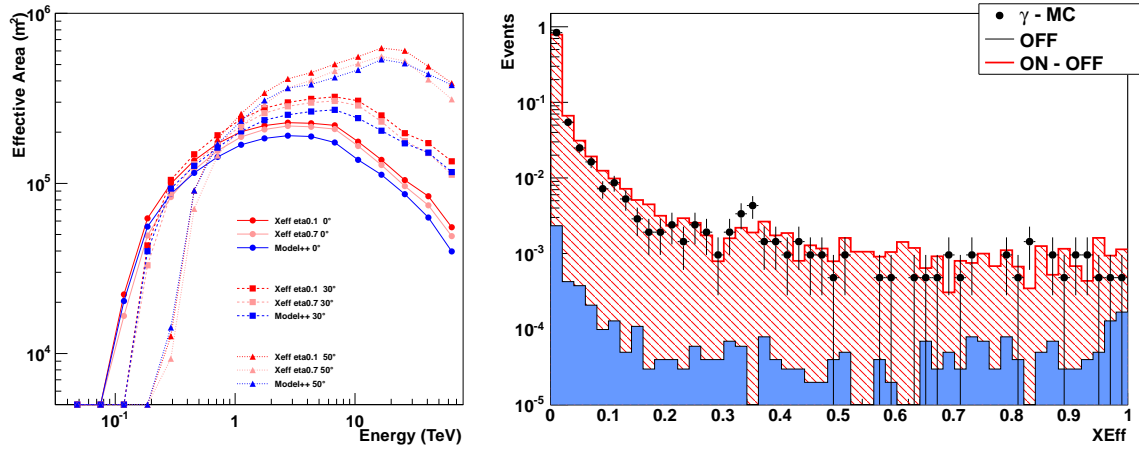


Figure 2: Left: Effective area of X_{eff} analysis for $\eta = 0.1$ and 0.7, and for 3 zenith angles. The effective area of *model++* analysis is added for comparison [3]. Right: X_{eff} distribution obtained using 3 runs of PKS 2155-304's flare and Monte Carlo simulations [6]. The red line represents the ON events normalized after background subtraction. The black points show the normalized distribution obtained with Monte-Carlo simulations corresponding to the same observation conditions. The blue histogram is the normalized background distribution.

4. Systematic studies

4.1 Comparison between simulation and real data

H.E.S.S. data are contaminated by hadron induced air showers. Furthermore, no background-free observations can be done. To control the behavior of the analysis on γ -ray data, we can use the July 2006 H.E.S.S. observation of PKS 2155-304's flare [6]. Thanks to the exceptional γ -ray flux, this flare produced the purest data set available which can be directly used for a comparison with Monte-Carlo simulations.

Three observations of July 28th 2006 corresponding to PKS 2155-304's flare were used. Figure 2 right shows the X_{eff} normalized distributions for this analysis. The very low level of background events provides a very high purity of the sample. These data were compared to simulations corresponding to the same zenith and target offset. However, given the available MC statistics,

simulations for azimuth angles of 0° and 180° and spectral index from 3.2 to 3.6 (a spectral index of ~ 3.4 was measured on the source with other analysis [6]) have been summed. The comparison shows that Monte-Carlo simulations are well describing the observed γ rays and the small differences can be understood as statistical fluctuations and possible bias due to Monte Carlo selection.

4.2 Comparison with *model* and published results on test sources

Simulations have shown a significant gain of X_{eff} analysis compared to standard analyses. Results obtained with X_{eff} and *model* analyses on several well known and/or published sources were compared. The sources have been arbitrarily chosen to be as representative of H.E.S.S. source type as possible : bright and faint, galactic and extragalactic, and soft and hard spectral index sources. As expected for faint sources, X_{eff} provides a gain of significance of $\sim 7\%$ up to $\sim 25\%$ while for bright sources the gain is significant for the γ -ray excess $\sim 10\%$, up to $\sim 17\%$. The differences from one source to another is due to the different observation conditions (in agreement with simulations).

4.3 Spectral comparison

The cuts introduced by X_{eff} analysis provided a better background rejection and/or a better γ efficiency than standard H.E.S.S. analyses. The spectral results may be affected by the event selection. A spectral analysis of the same set of sources as previously used was performed, assuming that the spectra are following a power law. Consistent results were obtained and no bias was introduced by the X_{eff} method.

The fits with exponential cut off power law model (for sources with such spectral shape) give consistent results as well. Finally, a comparison with published results using other analyses (*Hillas*, *3D-model* ...) shows consistent spectral parameters.

X_{eff} analysis depends on the free parameter η . Its value affects significantly the background rejection and the γ efficiency. The influence of the choice of this parameter on the results and in particular the spectral behavior of the analysis has been studied. Table 2 reports the results for 3 sources with different brightness (assuming the spectra follow a pure power-law). No evidence of significant bias or correlation between the value of η and the spectral parameters has been observed. This important feature leads to the conclusion that the parameter of X_{eff} analysis can be selected according to the purpose of the analysis. One can select hard cuts (high value for η) to get a good signal significance or loose cuts (low value for η) to increase γ -ray statistics.

5. Summary

The X_{eff} multivariate analysis, already used for H.E.S.S. data analysis, has been improved by taking into account the mean zenith and azimuth angles of the run and the energy of the event. Incorporating the differences between the reconstructed directions of *model*, *Hillas* and *3D-model* reconstructions increases the discrimination power of the X_{eff} combination. Simulation studies show that a higher γ efficiency and a better background rejection than standard H.E.S.S. analyses can be reached. The free parameter η allows to set different background rejection levels depending on the brightness of the target. This feature is a powerful way to increase the significance of faint sources enhancing the detection probability. Looking at various well known sources, it

Table 2: Results of X_{eff} analysis for 3 sources with different level of brightness. A pure power-law has been fitted on each spectrum. Various values of the free parameter η have been tested for each source. Results published with independent analyses are provided when available [10][11][12]. Only statistical errors are indicated.

Source	Method	η	N_γ	N_σ	Γ	$\Phi(1\text{TeV}) \cdot 10^{-12}$ $\text{cm}^{-2} \cdot \text{s}^{-1} \cdot \text{TeV}^{-1}$
LS5039	Xeff	0.1	3445	57.1	2.27 ± 0.03	1.74 ± 0.04
	Xeff	0.2	3360	57.8	2.27 ± 0.03	1.72 ± 0.04
	Xeff	0.4	3237	58.9	2.27 ± 0.03	1.70 ± 0.04
	Xeff	0.7	2991	59.8	2.25 ± 0.03	1.69 ± 0.04
G0.9+0.1	Xeff	0.1	607	18.4	2.34 ± 0.07	0.79 ± 0.05
	Xeff	0.2	563	17.9	2.30 ± 0.07	0.77 ± 0.05
	Xeff	0.4	533	18.2	2.30 ± 0.07	0.78 ± 0.05
	Xeff	0.7	469	18.3	2.26 ± 0.07	0.75 ± 0.05
	Pub	-	-	-	2.4 ± 0.1	-
Cen A	Xeff	0.1	410	8.6	2.6 ± 0.2	0.26 ± 0.03
	Xeff	0.2	427	9.4	2.7 ± 0.2	0.26 ± 0.03
	Xeff	0.4	414	9.9	2.7 ± 0.2	0.26 ± 0.03
	Xeff	0.7	385	10.9	2.5 ± 0.2	0.28 ± 0.03
	Pub	-	-	-	2.7 ± 0.5	0.25 ± 0.05

has been shown that X_{eff} produces a higher γ -ray excess for bright sources of $\sim 10\%$ and a higher significance of $\sim 7\%$ for faint sources, compared to the powerful *model* analysis. The present work provides a powerful analysis tool, suitable for all kind of physics analysis feasible with γ -ray imaging Cherenkov telescopes.

Acknowledgements: The authors would like to thank the H.E.S.S. Collaboration for the technical support and fruitful discussions, as well as for allowing us to use H.E.S.S. data in this publication.

References

- [1] Dubois F., Lamanna G. & Jacholkowska A., 2009, *Astropart. Phys.*, 32, 73
- [2] Hillas A.M., Proceedings of the 19th International Cosmic Ray Conference, August 11-August 23, 1985, La Jolla, USA, vol. 3, 445
- [3] de Naurois, M. and Rolland, L., 2009, *Astropart. Phys.*, 32, 231
- [4] Lemoine-Goumard M., Degrange B., Tluczykont M., 2006, *Astropart. Phys.*, 25, 195
- [5] Naumann-Godó M., Lemoine-Goumard M. & Degrange B., 2009, *Astropart. Phys.*, 31, 421
- [6] (H.E.S.S. Collaboration) Aharonian F. et al., 2009, *Astron. & Astrophys.*, 502, 749-770
- [7] (H.E.S.S. Collaboration) Aharonian F.A. et al., 2006, *Astron. & Astrophys.*, 457, 899-915
- [8] Buskulic D. et al., *Phys. Lett. B* 377 (1996) 205
- [9] Jaffe D.E., Le Diberder F., Schune M.H., LAL 94-67 and FSU-SCRI 94-101
- [10] (H.E.S.S. Collaboration) Aharonian F.A. et al., 2006, *Astron. & Astrophys.*, 460, 743-749
- [11] (H.E.S.S. Collaboration) Aharonian F.A. et al., 2005, *Astron. & Astrophys.*, 432, L25-L29
- [12] (H.E.S.S. Collaboration) Aharonian F.A. et al., 2009, *ApJ*, 695, L40