29th International Cosmic Ray Conference Pune (2005) **00**, 101–104

A robust way of estimating the energy of a gamma ray shower detected by the MAGIC telescope

K. Berger[a], T. Bretz[a], D. Dorner[a], D. Hoehne[a], B. Riegel[a] for the MAGIC collaboration[b]
(a) Institut für Theoretische Physik und Astrophysik, Universität Würzburg, Am Hubland, 97074 Würzburg, Germany
(b) Updated collaborator list at: http://magic.mppmu.mpg.de/collaboration/members/index.html
Presenter: K. Berger (berger@astro.uni-wuerzburg.de), ger-berger-K-abs1-og27-poster

We have investigated the role of various image parameters for the measurement of the energy of primary gamma rays with the MAGIC imaging air Cherenkov telescope. We point out that the Size parameter turns out to be the only relevant parameter for a robust estimation of the primary energy.

1. Introduction

The MAGIC telescope is located on the Roque de los Muchachos at a height of 2200 m a.s.l. on the Canary islands. It has been designed to observe gamma ray showers below 100 GeV energy of the primary gamma ray particle. Images of the showers are obtained with a photo multiplier tube camera, and then parametrized for analysis purposes. The image parameters [1] are important for gamma-hadron separation and for the determination of the energy of the shower. The showers are simulated using the Corsika Monte Carlo package [2]. Image parameters are derived from these simulations and compared with image parameters from the experimental data. Vice versa, image parameters are also derived from the simulations and then compared with the experimental data. It is shown in this contribution, how this allows to calculate the energy of the primary gamma ray particle in a robust way.

2. Parameterization

Several image parameters have been tested for their dependencies on energy. Based on these image parameters the energy estimator can be trained. Very different dependencies have been discovered and not all image parameters showed a dependency on energy. In order to investigate these dependencies, every image parameter has been plotted with respect to energy. After a simple energy estimation the residual energy has been tested. Then a polynomial fit was applied on every two dimensional distribution. The derived plots and fits are very different and unique for each image parameter. In some cases a fit with six coefficients was needed to describe the distribution. Some of these coefficients were very small, below 10^{-4}. This results in poor fits with large errors for the fit coefficients. More fit parameters also mean more degrees of freedom, resulting in poorer results (larger errors) for the energy estimation.

Thus many parameters have been abolished. Among these are the width (major axis of the ellipse of the shower) and the length (minor axis of the ellipse of the shower). The only parameter that showed a very strong dependency of energy is the size of the shower. Figure 1 shows a linear correlation between energy and size. A sufficient amount of statistics in each energy bin is important. Tests with different amounts of data have shown that at least 100 000 showers are needed for a reliable estimation of the energy. This approximation is only accurate for low zenith angles, preferably below 30°. For higher zenith angles the following equation has been found to describe the estimated energy:

$$E_{est} = (c_0 + c_1 \cdot \theta^2) \cdot S^{c_2} \tag{1}$$

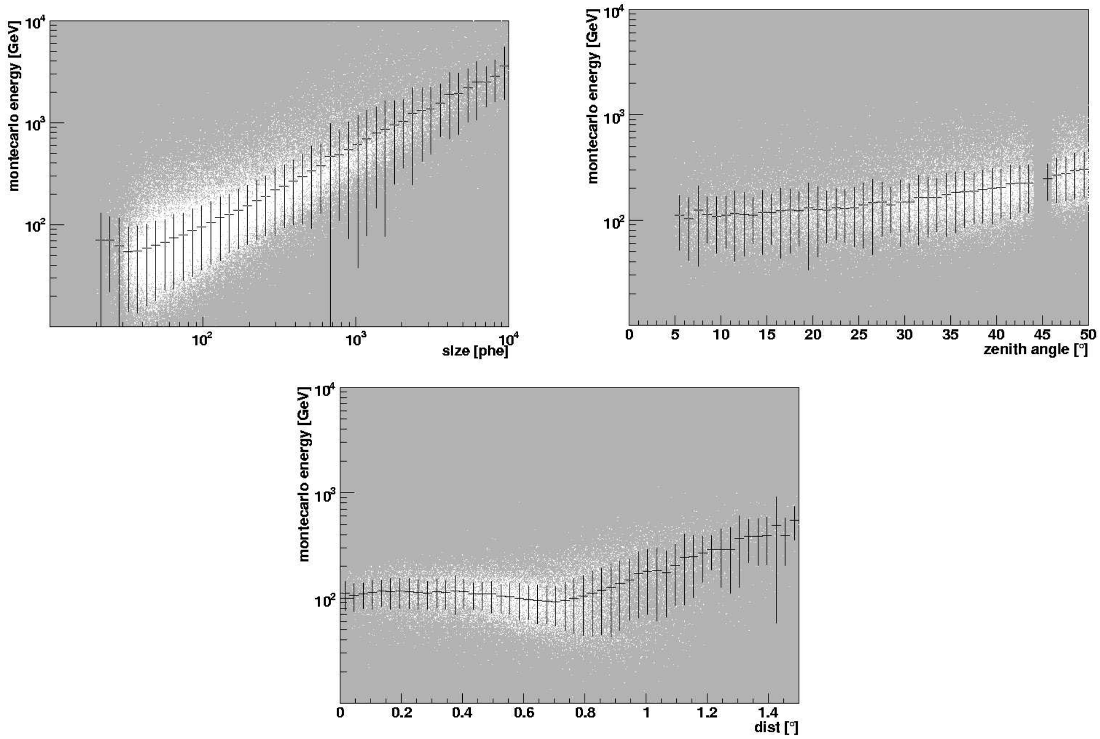

Figure 1. Energy distribution of simulated Monte Carlo events as a function of size, zenith angle and dist. The black crosses indicate the mean of each energy bin and the fluctuations in energy and size. The Monte Carlo sample contains gamma ray showers with zenith angles between 0° and 30° for the dist and size plot and 0° to 50° for the zenith angle plot. The dist and zenith angle plot are shown for a finite size bin.

With larger zenith angle the Cherenkov light from the gamma showers has a longer path through the atmosphere, resulting in stronger absorption. Thus it can be explained why high zenith angle showers have a higher energy than estimated and this must be corrected for. In order to optimize the estimator the MC sample has been randomly devided into two samples. One sample was used to run an optimization process using the simplex algorithm [3]. The result was then tested on the other sample. This was done to ensure that the energy estimator is not optimized on fluctuations in the Monte Carlo sample. Fluctuations and spurious correlations appear when there is insufficient statistics in the Monte Carlo sample. The optimization minimizes $\sum_i (log_{10}E_{i,est} - log_{10}E_{i,mc})$. This was chosen, since the energy is often plotted in logarithmic bins, e.g. in the spectrum of the gamma ray source. Fig.1 also shows a deviation of the straight line for energies below 100 GeV. This is a very important but only statistical effect due to fluctuations. To ensure that the Monte Carlo data and the experimental data are treated in the same way, the same standard analysis was used on both of them [4], [5]. The Monte Carlo simulations were only made for gamma ray showers between the energy of 30 GeV and 30 TeV. While there are always fluctuations to higher or lower energies at a given size bin, the lowest bins cannot have fluctuations to lower energies, since these have not been simulated or have been cut away by image cleaning. The same problem also occurs for energies above 1 TeV, since the energy fluctuations to higher energies are missing and there is not enough statistics. As a result the energy of these events should be extrapolated.

Most Monte Carlo events are simulated at lower energies due to the steep simulated spectrum. After the Monte Carlo data has been processed in the standard analysis chain there is still a larger number of events at low energies. The energy estimator is thus optimizing mainly on a small energy range with most of the events. Since a good fit is only useful over a wide energy range, weights should be applied that correct for this effect. Otherwise a flat Monte Carlo spectrum could be used. The key goal of the energy estimator is to reproduce the Monte Carlo spectrum of the data it has not been trained with. This should ensure that a spectrum of real gamma ray data will be reproduced correctly.

3. Random forest method

In a different approach the random forest method [6] was used for the energy estimation. The main application of random forest is to separate gamma ray showers from hadron showers. It therefore calculates a certain gamma probability for each event. If the gamma probability reaches a given limit, the event is classified as a background event, while the other events are classified as gamma events. This can also be used for energy estimation. The energy distribution is therefore devided into several logarithmic energy bins. For each bin random forest estimates the likelihood that a given gamma event can be assigned to the bin. The Monte Carlo sample was again devided into a test and training sample to check if the energy estimation is reproducing the MC spectrum correctly.

A dependence between the energy and size has been found as well as a dependence on the zenith angle at large values of the zenith angle, which is consistent with earlier results. In addition random forest was trained with dist as a parameter for energy estimation.

A correlation between dist and energy has been suspected (see Fig.1) but could not easily be implemented using an optimized fit. Other simple parameterizations are now in development. Dist corresponds to the distance of the shower to source position in the camera. Showers at the edge of the camera are being cut off. This results in a smaller size of these showers. Compared to a shower, that is completely within the camera, a smaller size corresponds to the same energy. Thus the energy is underestimated for all events, that are on the border of the camera. This effect produces a certain bias in the energy distribution, which can be corrected by using dist in the random forrest method. Other parameters such as leakage do not have this effect. Due to the possibility to include dist, random forrest provides slightly better results than the optimized fit method.

Other image parameters have also been tested. A combination of eleven image parameters (this includes the size, the width, the length, the disp, the dist, the third moments, alpha, theta, the leakage, the inner leakage and the zenith angle, see Fig.2) showed, that the three parameters described above are sufficient to reproduce the spectrum. Since more parameters introduce larger uncertainties and the possibility to optimize on special features of the MC spectrum, only three parameters are needed for a robust energy estimation.

There is still room for improvements. Since MAGIC is a single telescope, the height of the shower is unknown. But this information is very important for the energy reconstruction. More studies are therefore made to find a parameter, that can replace the height of the shower.

4. Conclusions

Size is the most important image parameter for the estimation of the energy of a gamma ray shower with a single-dish imaging air shower cherenkov telescope. The energy estimator becomes zenith angle dependent for higher zenith angles. Another important correction is the image parameter dist, since it allows to account for the many showers that are cut off at the border of the camera. This is especially important for wobble

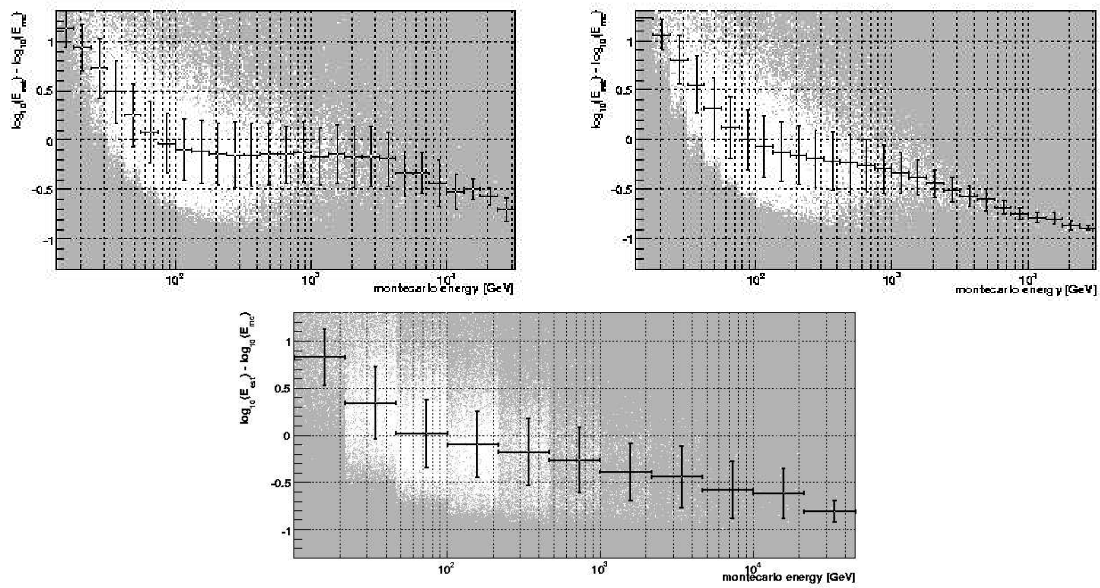

Figure 2. *Upper left and right:* results of the random forest method. *Below:* energy resolution obtained with optimized fits. The upper left plot includes only size, zenith angle and dist, while the upper right plot includes eight additional image parameters. All plots show the two dimensional distribution of Monte Carlo events (white) with mean and spreading for each bin (black crosses).

mode observations [7]. Including these corrections the energy estimator allows for a stable and reliable energy estimation. The random forest method provides better results, since it allows the use of dist. Further Monte Carlo simulations are planned, e.g. to include very high energy showers, as well as low energy showers and showers at large zenith angles, for the further investigation of statistical effects.

5. Acknowledgments

We acknowledge support by the german Federal Ministry of Education and Research (BMBF, 05 CMOMG1/3) and the Astrophysics Institute of the Canary Islands (IAC).

References

[1] A systematic study of the interdependence of IACT image parameters, T.Bretz et al, these proceedings.
[2] Monte Carlo simulation for the MAGIC telescope, Majumdar et al, these proceedings.
[3] Root description, http://root.cern.ch/root/html301/TMinuit.html.
[4] Standard Analysis for the MAGIC Telescope, T.Bretz et al, these proceedings.
[5] Data Management and Processing for the MAGIC Telescope, D.Dorner et al, these proceedings.
[6] Leo Breimann, http://www.stat.berkeley.edu/users/breiman/RandomForests/cc_home.htm.
[7] Comparison of On-Off and Wobble mode observations for MAGIC, T.Bretz et al, these proceedings.