33RD INTERNATIONAL COSMIC RAY CONFERENCE, RIO DE JANEIRO 2013
THE ASTROPARTICLE PHYSICS CONFERENCE

# MARS, the MAGIC analysis and reconstruction software

ROBERTA ZANIN [1], EMILIANO CARMONA[2], JULIAN SITAREK[3], PIERRE COLIN[4], KATHARINA FRANTZEN[5], MARKUS GAUG[7], SAVERIO LOMBARDI[8], MARCOS LOPEZ[9], ABELARDO MORALEJO[3], KONSTANCJA SATALECKA[9], VALERIA SCAPIN[9], VICTOR STAMATESCU[3] FOR THE MAGIC COLLABORATION.

[1] *Universitat de Barcelona (ICC/IEEC), E-08028 Barcelona, Spain*
[2] *Centro de investigaciones energetícas, medioambientales y tecnológicas, E-28040 Madrid, Spain*
[3] *IFAE, Edifici Cn., Campus UAB,E-08193 Bellaterra, Spain*
[4] *Max-Planck-Institut fur Physik, D-80805 Munchen, Germany*
[5] *Technische Universitat of Dortmund, D-44221 Dortmund, Germany*
[7] *Fisica de les Radiacions, Departament de Física, Universitat Autonoma de Barcelona, 08193, Bellaterra, Spain*
[8] *INAF National Institute for Astrophysics, I-00136 Rome, Italy*
[9] *Universidad Complutense, E28040 Madrid, Spain*

*rzanin@am.ub.es*

**Abstract:** In the last two years, the MAGIC telescopes underwent an upgrade involving the readout system of both telescopes as well as the camera of the oldest MAGIC I telescope. The standard analysis package of the MAGIC collaboration, MARS, has been adapted to the new hardware configuration. MARS is a ROOT-based code, written in C++, which includes all the algorithms necessary to reconstruct the shower images and extract the physical information from them. The standard analysis chain is presented: from the raw data up to the extraction of the final scientific results. In particular, we describe in detail the newly developed algorithms for background estimation, event stereoscopic reconstruction and sky-mapping determination.

**Keywords:** analysis software, IACT, MAGIC

## 1 Introduction

MAGIC consists of two imaging atmospheric Cherenkov telescopes (IACTs) with a 17m diameter reflective surface and ultra-fast electronics. It is located on the Canary Island of La Palma at 2200 m above sea level at the Roque de los Muchachos observatory. MAGIC started operations as a single telescope, MAGIC I, in Fall 2004 and became a stereoscopic system in Fall 2009, when the second telescope ended its commissioning phase. At the end of 2011 the readout systems of both telescopes were upgraded with the domino-ring-sampler (DRS) version 4 chip [1]. In Summer 2012 the camera of MAGIC I was substituted by a clone of the camera of the second telescope which has a finer pixelization [2]. This upgrade led to an increase of the MAGIC-I trigger region by a factor 1.7. The performance of the recently upgraded telescopes is presented in [3].

The standard data analysis chain of the MAGIC collaboration is performed in several different steps (see the flow diagram in figure 1), each having a dedicated program included in the MARS (*MAGIC Analysis and Reconstruction Software*) package. MARS is a collection of ROOT-based [4] programs, written in C++, for the MAGIC data analysis. It also includes programs for the Monte Carlo (MC) production. The current MARS version 2.12.2 is designed to work with ROOT version 5.34 and it is supported for multiple Linux and Macintosh distributions.

The starting point of the data analysis are the raw data recorded by the telescopes, which consist of binary files containing the full information available per pixel and per triggered event, plus ASCII files (reports) containing auxiliary information from the different telescope subsystems (like the telescope drive system, the trigger system or the weather station). These data are converted into ROOT format by the MARS program *merpp*, and organized in ROOT-trees containing, for every event, a set of values (including the time stamp) stored in parameter containers.

## 2 Signal extraction and calibration

The raw data contain for each triggered event the waveforms of each readout channel, sampled by the DRS4 chip at the frequency of 2 GSample/s. Since the upgrade of the two readout systems and the MAGIC I camera in Summer 2012, the size of the recorded window is 60 time slices (30ns). The first step of the analysis chain consists of extracting and calibrating the signals of each pixel. Once the baseline offset is subtracted, estimated from interleaved "pedestal events" taken with a random trigger, the signal is extracted by using the 3ns sliding window extractor. The latter searches for the maximum integral content of a number of consecutive time slices inside the recorded time window. The signal arrival time is calculated as the average of the time slices, weighted with the sample value.

After this, a calibration of all pixels is performed to achieve an equalized response to a light pulse homogeneously illuminating the whole camera (flatfielding). The calibration is done by means of calibration events interleaved with ordinary data. An absolute calibration procedure, based on the F-factor method [5], is also applied to convert the digital counts of the extracted signals into photo-electrons (*ph.e.*) which are proportional to the number of Cherenkov photons reaching a pixel.

These procedures are performed within MARS by a program called *sorcerer*. This new program substituted the previous calibration program, *callisto*. *Sorcerer* was optimized to decrease the execution time of processing the DRS4 data, which require special analysis procedures such as the absolute time calibration or advanced cell offset

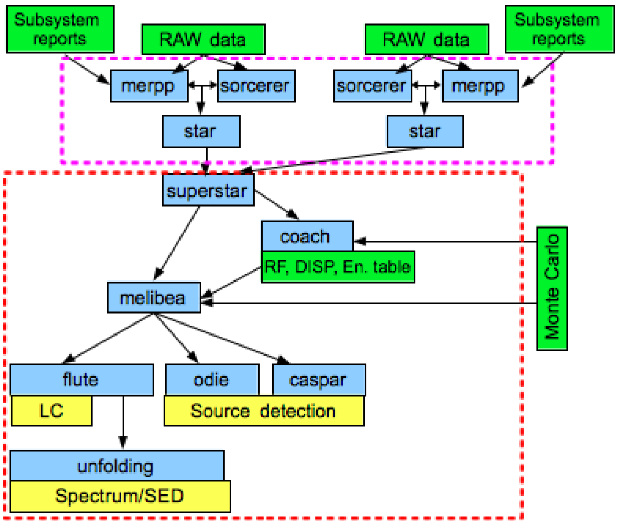


**Fig. 1**: Flow diagram of the MAGIC analysis chain. The color palette indicates input data in green, standard chain programs in blue, high level results in yellow. The magenta line shows the part of the analysis chain which is usually performed by automatic analysis, whereas the red line the part which is usually performed by individual analyzers.

correction [6]. It also provides a user-friendly access to various diagnostic plots for each individual pixel. Such a feature is useful for detection and investigation of hardware problems with pixels or readout channels.

## 3 Image cleaning and parameterization

After calibration, the image is reconstructed by identifying all those pixels that contain information about the showers, and rejecting those whose signal likely results from the fluctuations of the night sky background light. This procedure is performed by the image cleaning algorithm. Its basic idea consists of defining two absolute signal thresholds which allow to define core and boundary pixels (the core threshold is larger than the boundary one) in two successive stages. In addition, there are further constraints on the topology which exclude pixels unrelated to the image.
Each shower image is then parameterized by a small set of parameters describing the image orientation, and shape, and its timing properties. The basic quantities are described by the *Hillas* parameters [7], which are basically the moments up to the second order of the light distribution in the camera. The image reconstruction and parameterization is performed within the MARS package by a program called *star*.

The performance of the image cleaning algorithm can be improved by introducing more constraints which take advantage of the compactness in time and extension of the shower images. Such new constraints allow to lower the charge thresholds, thus to yield the analysis energy threshold closer to the trigger energy threshold. The default algorithm in the MAGIC collaboration is the so-called *sum image cleaning* [8]. It requires a core pixel to lie within a compact group of $x$ next-neighbor ($x$-NN, with $x$=2,3,4) pixels, and the total clipped charge of the $x$-NN group has to exceed the threshold of the corresponding group. In addition, the arrival time of a core pixel has to fall within a defined time window from the mean arrival time of the corresponding $x$-NN group. Boundary pixels must have at least one core neighbor pixel.

A new cleaning algorithm has been recently developed to improve the analysis sensitivity at energies above 1 TeV, the dynamic image cleaning, which is described in [9].

## 4 Stereoscopic shower reconstruction

Up to the image parametrization (*sorcerer* and *star*), the MARS analysis chain runs over the data of each telescope separately. At this point, there are two sets of data containing the same events observed from two different points of view. The task of the next analysis step, performed by the MARS program *superstar*, consists in combining the matching pairs of events in one data set. Therefore, events which did not survive image reconstruction in one telescope are removed from the analysis. *Superstar* computes also the stereoscopic parameters giving a 3-dimensional event description. The shower direction and the impact point are computed from the intersection of the major axes of the two single-telescope images (giving a first estimation of the shower direction). In the case of 2-telescope system, like MAGIC, the shower axis has a unique geometrical solution. Its accuracy depends on the relative position of the two telescopes and the shower: the more parallel the two images on the camera planes, the bigger the uncertainties. Given the limitations of this geometrical reconstruction, a more complex procedure has been developed to reconstruct the shower direction: it is a more sophisticated version of the *disp*-method [10], described in section 6.
Once the shower axis is determined, the height of the shower is derived from the angle at which the image center of gravity is viewed from each telescope. *Superstar* also calculates the impact parameter of the shower with respect to each telescope.

The standard procedure to estimate the shower energy makes use of look-up tables (LUTs). The LUTs are based on a simple model describing the distribution of the Cherenkov photons on the ground. In particular, it relies on the Cherenkov radius and Cherenkov density parameters, namely the radius of the Cherenkov light ring and the photon density on the ground, produced by a single electron track at the level of the shower maximum. The electron is assumed to be a typical electron of the electromagnetic shower with an energy equivalent to the critical energy in the air (86 MeV) and travelling parallel to the shower axis direction.
The LUT model assumes that the shower energy is proportional to the number of electrons in the shower, hence to the ratio of the total charge of the image (size, in *ph.e.*) and the Cherenkov density. It also assumes that, outside the Cherenkov light pool, the Cherenkov density depends on the impact parameter of the shower in units of Cherenkov radius.

The LUTs are built independently for each telescope starting from a sample of MC simulated $\gamma$-ray events, and the final estimated energy is the weighted average value obtained from the two telescopes, where the weight is given by the statistical uncertainty of the two measurements. The building of the LUTs is implemented in the MARS program *coach* which is also in charge of the background discrimination (section 5), and of the definitive shower direction definition (section 6). On the other, the assigment of an estimated energy value, to each event, on the basis of the already created LUTs, is done in the MARS program called *melibea*.

## 5 Background discrimination

The discrimination between background showers, produced by charged cosmic rays, and the $\gamma$-ray-induced events is performed by means of a multivariate classification analysis, based on the Random Forest (RF) algorithm [11]. RF is trained with events of known nature, i.e MC simulated $\gamma$-ray data and hadrons randomly extracted from real data recorded when the telescope is pointing at a region of the sky with no known $\gamma$-ray sources. In order to build the decision trees, it uses a set of discriminating parameters. Both single-telescope parameters and global ones obtained from the stereo reconstruction are used for this purpose. A part from the energy reconstruction, the program *coach* also produces the RF matrices, which are later used by the program *melibea* to assign a value of the global parameter *hadronness* to each event. The *hadronness*, spanning from 0 ($\gamma$-like) to 1(hadron-like), indicates if the event is a good gamma candidate. Only events with *hadronness* below a certain cut value will be used for the subsequent steps of the analysis.

## 6 Determination of the arrival direction

We revised the *disp*-method, proposed in [10], in order to have a more accurate estimation of the shower direction, and, consequently, of the source position on the camera plane, hence to improve the angular resolution of the experiment. *disp* is an estimate of the distance between the image center of gravity and the point on the camera which corresponds to the shower direction. It is computed by a RF algorithm, *disp*-RF [12], which is trained with a sample of MC simulated $\gamma$-ray events for which the correct source position, and hence the disp parameter, is known. The RF decision trees are built using parameters correlated to the image elongation, among which those obtained by the pure geometrical reconstruction of the shower direction.

Again, the program *coach* takes care of creating the *disp*-RF matrices, which are, then, used by the *melibea* program to assign a *disp* value to each single-telescope event. For each of them, the disp identifies two possible reconstructed source positions (head-tail ambiguity). At this stage for each stereoscopic event, *melibea* evaluates all four possible combinations of position pairs, and chooses the closest pair, but only if its distance is smaller than a certain value. If none of the pairs satisfies this condition, the event is rejected. The stereo-reconstructed position is determined as the weighted average of the chosen pair of positions.

## 7 Source detection

A source detection implies to search for and determine the statistical significance of a $\gamma$-ray signal towards the source direction in the sky. This can be done by counting, after a cut in *hadronness*, the number of $\gamma$-ray excess events within a maximum angular distance of the source direction, and subtracting from it an estimate of the number of background events. The MARS program which is in charge of it is called *odie*.

Given that all the observations are carried out in wobble mode (the telescope pointing wobbles around the source at 0.4° distance), the background is evaluated from the same data set containing the source signal. The background is evaluated from the same region of the camera where the excess events have been evaluated for events several minutes before. This is the *off-from-wobble-partner* method which works for any source position in the camera plane. It allows, by definition, for $(n\_wobble\_pointings - 1)$ background estimation regions. In order to have a better background statistics we increased to 4 the standard number of wobble pointings.

The significance of the excess events is usually evaluated following the Li&Ma (1983) equation 17 [13]. Nevertheless, we are starting to use the generalized likelihood ratio test statistic described in [14] instead, because it is more suitable for data taken in wobble mode.

## 8 Energy spectrum and light curve

The differential energy spectrum of the observed gamma-rays is estimated by the program *flute* of the MARS package. It computes the number of excess events in bins of estimated energy and divides it by the total observation time and by the $\gamma$-ray effective area for each energy bin.

The effective area, after all cuts, is calculated by using a test sample of MC $\gamma$-ray events (which must be statistically independent from the MC sample used by *coach* for the energy estimation, the training of the gamma-hadron separation, and the disp calculation). The effective area depends on the direction in local coordinates of the observed $\gamma$-rays, mainly due to the variation of the air mass along the pointing direction with the zenith angle. It also depends on the azimuth angle due to the effect of the geomagnetic field and to the variation of the "effective distance" between the two telescopes (i.e. distance of the telescopes in he plane orthogonal to the observation direction). The effective area in an energy bin is therefore obtained as a weighted average, where the weights are proportional to the observation time spent in each bin of zenith and azimuth angle traversed by the source.

The true energy of each $\gamma$-ray event remains undetermined. However, it can be obtained from the measured one by accounting for the natural distortions due to the biases and the finite resolution of the detector. Such information is included into the detector response matrix, i.e. migration matrix, $M$, whose elements indicate a sort of probability that the events with a considered true energy are measured with a different estimated energy. *Flute* provides differential energy spectra which are already corrected for the event migration across the energy bins, under the assumption that the tentative spectrum used in the calculation of the effective area roughly matches the actual spectrum (kind of forward unfolding, see below). The errors on the obtained spectral points, including the statistical errors on the number of excess events, and the uncertainty on the effective area, are therefore correlated.

*Flute* gives also the possibility to apply the unfolding procedure to its spectra, i.e. to correct for the finite energy resolution and bias in the most formal way. If the measured distribution $Y(y)$ can be expressed in terms of the true one $S(x)$ by:

$$Y(y) = \int M(y,x)S(x)dx \tag{1}$$

where $y$ indicates the estimated energy, $x$ the true energy, the aim of the unfolding is to determine $S$ given $M$ and $Y$. Being not analytically solvable, the problem is often discretized:

$$Y(y) = M \times S \tag{2}$$

There are different approaches to find the solution. One consists of inverting the matrix $M$. Although technically correct, the solution usually leads to large fluctuations making the distribution of the true energy differ from the physical one, hence useless. The remedy to this problem is the regularization: adding a regularization term (*Reg*) to the expression to be minimized, the least square expression ($\chi_0^2$), allows to smoothly suppress those non-significant coefficients of the solutions which are the cause of large fluctuations:

$$\chi^2 = \chi_0^2 + \frac{\tau}{2} Reg \tag{3}$$

Very small values of $\tau$, corresponding to no regularization, often produce noisy spectra which perfectly fit the data. Large values of $\tau$, indicating high level of regularization, yield to a smooth unfolded distribution which can strongly deviate from the data. Thus, selecting the appropriate value for $\tau$ is very important. In the MARS package, there are different unfolding methods available which differ for the way the regularization is implemented [15, 16, 17].
Another approach consists of assuming a parameterization of the true distribution $S$ and then comparing $M \times S$ with the measured distribution $Y$. This is called forward unfolding. The main difference to the previous approaches is that an assumption about the true distribution has to be made. It is not a real unfolding because its result is just the best fit, with corresponding errors, obtained using an a priori assumed parametrization. In the MARS analysis, the various unfolding methods are used for each observation, and the consistency of the results is checked.

The light curves are $\gamma$-ray fluxes, in a given energy range, as function of time. No unfolding procedure of the above-described type is used for them. A simple correction is applied to the effective area in the selected energy range to account for the spillover of the events with true energy outside it, under the assumption of a given spectral shape (usually a power-law type) for the energy spectrum.

## 9 Sky maps

Sky maps are bi-dimensional histograms containing the distribution of arrival directions of events after all cuts. They allow for blind searches of signal in the whole field of view of telescopes: this is needed in case the exact location of a $\gamma$-ray source is not known a priori, or the observed source is extended. The program which is in charge of producing sky maps within the MARS package is *caspar*.

The direction coordinates (in the camera plane) of all $\gamma$-like events are converted into celestial coordinates (Right Ascension, Declination). Given that these events include the residual cosmic ray background, obtaining an excess event sky map requires knowing the camera acceptance for residual background events, itself projected into the celestial coordinate system.
The background exposure map is modeled from the data. In wobble mode, this is usually constructed from the sky areas opposite to the source position in each wobble data set while rotating its coordinate system with the azimuth angle in order to remove this dependency from the camera acceptance. The background exposure map is then obtained by oversampling the acceptance model.
Before the background subtraction, a Gaussian kernel smoothing is applied to the data: the width of this kernel is an analysis parameter to be adjusted depending on the measured point spread function ($\sigma_{PSF}$) of the analysis. In a blind search for signal in the field of view, the kernel is chosen to be $1\sigma_{PSF}$: a good compromise between high resolution and low noise.

We compare the measured events to our estimated background exposure map by constructing a test statistic (TS) that applies the Li & Ma significance definition using an extremely low normalization factor, which is inversely proportional to the background model oversampling factor. The null hypothesis is approximately Gaussian, although some deviations from it can be found depending on the available statistics, above all at the edges of the sky window. We therefore construct out null hypothesis distribution by computing the TS from toy MC sky maps which use the background exposure map.
We note that the resulting TS map is not a physical quantity. We consider, instead, the $\gamma$-ray flux in arbitrary units (a.u.) which is calculated as the number of smeared excess events in units of residual background flux within $0.1^\circ$. Given that the background density is roughly proportional to the effective area, the relative flux is, in first approximation, proportional to the absolute flux.

## 10 Summary

This paper presents an overview of the methods implemented in the MARS package for the stereoscopic analysis of the MAGIC data. It goes through all the steps of the standard analysis chain needed to produce the scientific outputs, i.e. differential energy spectra, light curves, and sky maps.

## References

[1] http://www.psi.ch/drs/
[2] Mazin, D. et al., These proceedings
[3] Sitarek, J. et al., These proceedings
[4] Burn, R. & Rademarkers, F., Proceedings AIHENP'96 Workshop, Lausanne, 1996
[5] Mirzoyan, R. and Lorentz, E., Proceedings of the $25^{th}$ ICRC, Durban, 1997
[6] Sitarek, J. et al., Nucl. Instr. Meth. A, in press
[7] Hillas, et al. Proceeding of the $32^{th}$ ICRC, Beijin, 2011
[8] Lombardi, S. et al., Proceeding of the $32^{th}$ ICRC, Beijin, 2011
[9] Berger, K. et al., These proceedings.
[10] Fomin, V. et al., Astroparticle Physics 2 (1994) 137
[11] Albert, J. et al., Nucl.Instr.Meth. A 588 (2008) 424
[12] Aleksić, J. et al., A&A, 524 (2010) A77
[13] Li,T.P. and Ma, Y.Q., ApJ, 271 (1983)
[14] Klepser, S. Astroparticle Physics 36 (2012) 64
[15] Bertero, M., Advances in Electronics and Electron Physics, Vol. 75 (1989) 1
[16] Schmelling, M., Numerische Methoden der datenanalyse, MPI-K Heidelberg (1998)
[17] Tikhonov, A.N. and Arsenin, V.J., Methods of solution of illposed problems, Nauka, Moscow (1979)