CTA-WP-MC Internal note:
A short description of an *evndisplay*-based CTA analysis

Gernot Maier
DESY *

January 10, 2012

Abstract

evndisplay is one of the main software packages developed and used for the analysis of data and simulations in the VERITAS collaboration. It was originally designed as display for the VERITAS prototype data, but evolved into a full analysis package with routines for calibration, FADC trace integration, image and stereo parameter analysis, response function calculation and high-level analysis steps (e.g. sky maps and spectral analysis). The package has been used for simulations of the response of the 36-telescope concept of AGIS and is well suited for the analysis of CTA-type arrays with a large number of telescopes. The *evndisplay* code is public and available to all members of the CTA collaboration. We describe in this short note the different analysis steps, the most important methods used in *evndisplay* and conclude with some results obtained with the described methods on the CTA subarray E.

1 Introduction

The analysis techniques described in the following are fairly standard and widely used in the field of VHE γ-ray astronomy. In summary, they consist of calibration, FADC trace integration (if necessary), image cleaning, second-moment parameterization of the recorded images [6], reconstruction of shower direction and impact parameter using stereoscopic methods (see, e.g., [9]), gamma-ray/hadron separation, calculation of effective areas, sensitivities and high-level analysis products. The performance of *evndisplay* is extensively tested on VERITAS data and simulations (see e.g. [11]). The sensitivity achieved is sufficient to detect sources with a flux of 1% of the Crab Nebula in less than 30 hr of observations (1 hr for 5% of the Crab Nebula flux) with the VERITAS observatory[1].

The VERITAS data evolved over time significantly: changes in data formats, detector geometry (number of telescopes, array layout) and in the parameters of the data acquisition and calibration. This led to a flexibility in the analysis which made it relatively easy to adapt the software for next-generation arrays with on the one hand a much larger number of telescopes (see e.g. its application on AGIS simulations [12]) and on the other hand many different telescope types.

The analysis steps in *evndisplay* relevant for the analysis of CTA Monte Carlo simulations are described in the following. They consist of

1. converter from sim_telarray (hessio/eventio) format to evndisplay DST format
2. calibration and FADC trace integration (optional), image cleaning, stereo reconstruction
3. training and use of lookup tables to estimate the energy and mean scaled parameters
4. cut optimization using TMVA methods

*gernot.maier@desy.de

[1] see http://veritas.sao.arizona.edu/about-veritas-mainmenu-81/veritas-specifications-mainmenu-111

5. estimation of effective areas, angular and energy resolution
6. calculate and plot sensitivities and instrument response functions
7. writer for CTA WP-Phys style sensitivity root files

It should be noted that only default analysis routines are described in the following, many options exists for different algorithms or option settings.

2 Converter

evndisp cannot read MC files in hessio/eventio format directly but uses an intermediate converter into the so called EVNDISP-DST format. This is essentially a simple root file with four trees: the main data tree with one entry per MC event containing the pixel information (integrated charge, timing, FADC trace (optional), trigger information) and the list of telescopes with data; a tree with all the necessary calibration information (e.g. pedestal and pedestal variation per pixel); one tree with one entry per telescopes containing the detector geometry (e.g. telescope positions, number of telescopes, etc) and one tree with all the MC events (one energy per MC events with thrown energy, direction, core position etc). The converter has to be run for each subarray once. There are no important choices or options for the user at this step.

3 FADC trace integration

The default FADC trace integration algorithm uses a two-pass method as described in [7]. In the first pass, a very wide integration window is applied to each FADC trace starting at a fixed position in time in order to calculate the integrated charge and the pulse arrival time. The resulting images are then cleaned as described in the next section. The gradient in the arrival times along the long axis of these cleaned images is then used in a second step to calculate the optimal placement of a much shorter integration window (see Figure 1 for an example of a time gradient across the long axis of an image). This method allows us to use short integration windows with optimal signal to noise ratios for small pulses and at the same prevent signal losses due to significant time gradients in large showers.
This step is optional and depend on the configuration used during the MC run.

4 Image and stereo analysis

Images are cleaned to remove noise pixels from the images. The cleaning consists of a two-level filter with user-defined thresholds q_1 and $q_2 < q_1$ removing all pixels with an integrated charge smaller than q_1 (image pixels) and any pixels that are adjacent to the remaining pixels and have signals smaller than q_2 (border pixels). Besides the described cleaning with fixed image and border thresholds, variable cleaning levels depending on the average noise level per pixel can be applied as an alternative cleaning method. Cleaning thresholds can and should be set differently for different telescope types and depend on the level of night-sky background.

The shower image is then parametrized with a second moment analysis [6]. A simple log-likelihood fitting algorithm is applied to recover partially contained images at the edge of the camera (the fitting method allows additionally to minimize the influence of dead or suppressed channels on the estimation of image parameters). The underlying assumption of the fitting method is that the image of a gamma-ray shower can be described by a two-dimensional normal distribution. The fitting routine is only applied to images at the edge of the camera with a loss value > 0 and < 0.5 (< 0.2 for LST; loss is defined as the fraction of image size contained in pixels on the edge of the camera). This extrapolation leads to a increase in sensitivity for showers with large energies or large directional offsets relative to the camera center.

Quality cuts are applied to the images before the direction of origin of the event on the sky and the impact parameter of the shower core on the ground are reconstructed. The default values of the cuts are: minimum number of image/border pixels per image (> 4 for LST and MST, > 4 for SST), no cuts on image size, maximum loss value (< 0.2 for LST, < 0.5 for MST and SST). These values have not been determined through a strict optimization procedure. A minimum of 2 images are required for the following stereo reconstruction.

Shower direction and core are reconstructed using techniques first used by the HEGRA collaboration. The stereo reconstruction is geometrical, using algorithm 1 from [8]. The shower direction is calculated from the mean intersection points between all possible pairs of telescope images. Image pairs are weighted by there image sum, width/length, and angular difference between the image pair ('parallelness'). Shower core positions are determined in a similar way.

The analysis results of this stage are written to disk in root format as input for the following analysis steps.

Last but not least a very important and useful feature: the camera display (see Figure 1), which allows us to check visually for example the effects of image cleaning or new trace integration routines.

5 Lookup tables

Energies and mean scaled with and length parameters are reconstructed using lookup table based on MC simulations.

The energy of each event in the source and background region is estimated assuming that the primary particle is a gamma-ray. The calculation determines the energy of an event as a function of the distance of the reconstructed impact parameter from the position of each telescope, image size, offset from the camera centre, level of night-sky background, zenith angle and telescope type. The median and 90%-width-values of the logarithm of the size parameters as function of the primary gamma-ray energy and distance of the impact parameter are filled into the lookup table. For each telescope with a valid image, an energy estimate is determined by inverting the lookup table. The energy of the primary particle is determined by averaging the energy estimates from all telescopes with a valid image. Images are weighed by the estimates from each single telescope proportional to one over the square of the statistical uncertainty on the estimate. The χ^2-value of the energy estimation is saved as well for later use in the gamma-hadron separation.

Mean-scaled width and length parameters are calculated in a analogical way. The lookup tables contain the median and 90%-widths of the image parameter width (w_{MC}, $\sigma_{width,MC}$) and length (l_{MC}, $\sigma_{length,MC}$) as a function of the distance of the impact parameter R, image size s, and zenith angle Θ:

$$mscw = \frac{1}{N_{Images}} \left(\sum_{i}^{N_{Images}} \frac{width_i - w_{MC}(R, s, \Theta)}{\sigma_{width,MC}(R, s, \Theta)} \right)$$

and similar for mean scaled length. A large number of lookup table can be produced (depending on the parameter space), values are then obtained by interpolating between the results from the different table (e.g. in zenith angle or background noise)[2]. There is one complete set of lookup tables for each telescope type.

One additional parameter calculated in this analysis step is the height of the maximum Cherenkov emission [2]: the distance c between the image centroids in the cameras of two telescopes pointing towards the source (parallel pointing mode) is related to the distance D of the telescopes to each other and the height of the Cherenkov emission maximum h by c = D/h. This parameter is calculated per pair of telescopes, the values used in later analysis steps is the corresponding mean value (weighted by the image size) and the χ^2-value (gamma-rays are expected to give a very similar emission height in contrast to background hadronic events).

[2]Note that the number of simulated showers is not always large enough to fill tables sufficiently. The analysis presented here uses for example no off-axis showers, but in all cases on-axis tables.

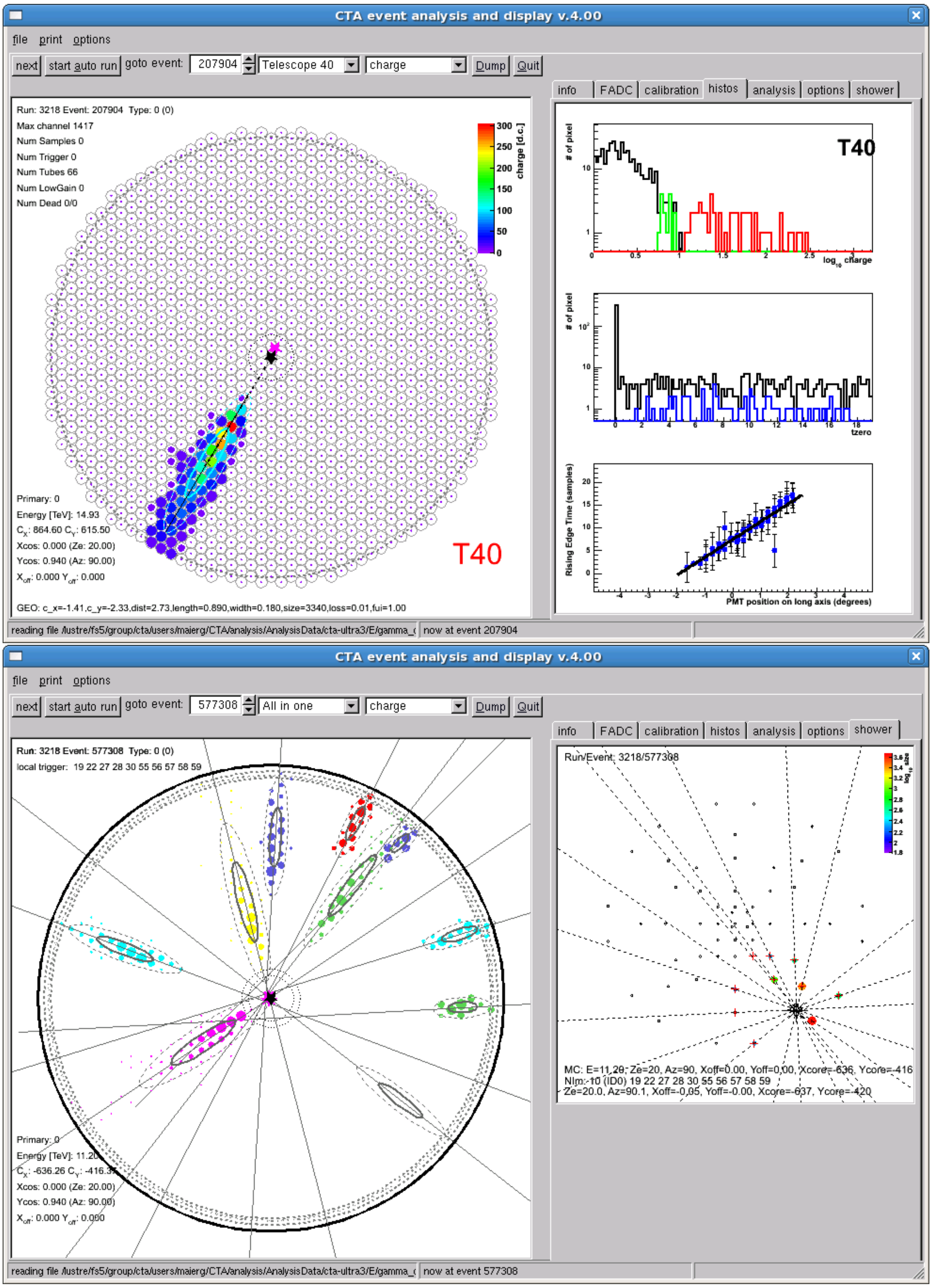

Figure 1: Camera and event display for a randomly selected high-energy event. The panel to the top right shows from top to bottom the distribution of integrated charges (black line: all; red: image pixel; green: border pixel), the distribution of pulse times (50% rise time) for image and border pixels, and the time gradient along the long axis of the image used for the 'double-pass' trace integration. The bottom figure shows the camera images (left) and the distribution of telescopes on the ground (right).

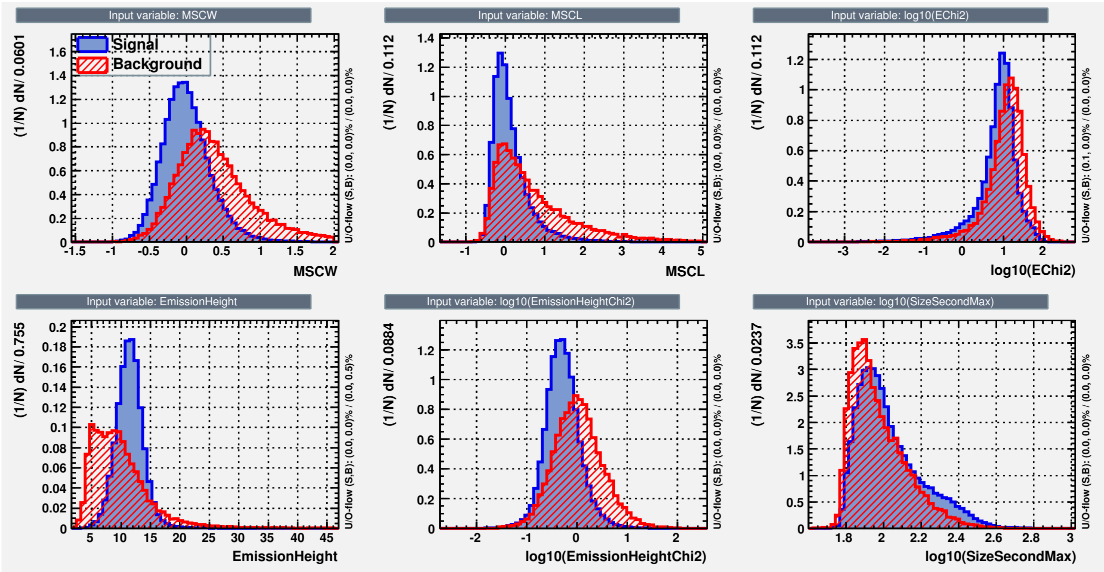

Figure 2: Distribution of cut variables for events with low energies (< 60 GeV). Signal events are simulated gamma rays, background events consist of proton events only.

6 Gamma-hadron separation and cut optimization

Six different variables are used for gamma-hadron separation (see listing below). The optimization of the gamma-hadron separation is difficult and not straight forward. In this approach we tried to minimize the number of parameters to be optimized. No 'low' level parameters like image cleaning parameters, minimum amplitude of images (not really, since the second largest amplitude is used as a cut value) and telescope multiplicity values are used in the optimization, all of them are set to minimum value (e.g. the telescope multiplicity required is 2). The variables utilized for gamma-hadron separation are:

1. shower direction (Θ^2): use always angular resolution (80% containment radius after some quality cuts)
2. mean scaled width (MSCW) and length (MSCL)
3. second largest image amplitude per event (SizeSecondMax)
4. χ^2-value of the energy estimation (EChi2)
5. height of the maximum Cherenkov emission (EmissionHeight)
6. χ^2-value of emission height estimation (EmissionHeightChi2)

Figures 2 and 3 show the parameter distributions for gamma-rays (blue) and proton events (red) for low (< 60 GeV) and high energies ($\approx$ 10 TeV). Mean scaled width and emission height (at low energies) are the most important cut parameters.

There are a variety of options for the gamma-ray hadron separation, ranging from box cuts using single telescope or stereo parameters to energy dependent cuts based on multivariate methods. Boosted decision trees (BDT) as implemented in the TMVA package[3] are used in the following for the gamma-hadron separation (400 trees are used in the training, all other BDT parameters are left at their default values). The BDT are trained and applied in several overlapping energy bins (typically 12 bins, the best suited energy bin is chosen for each event). The BDT output

[3] http://tmva.sourceforge.net/

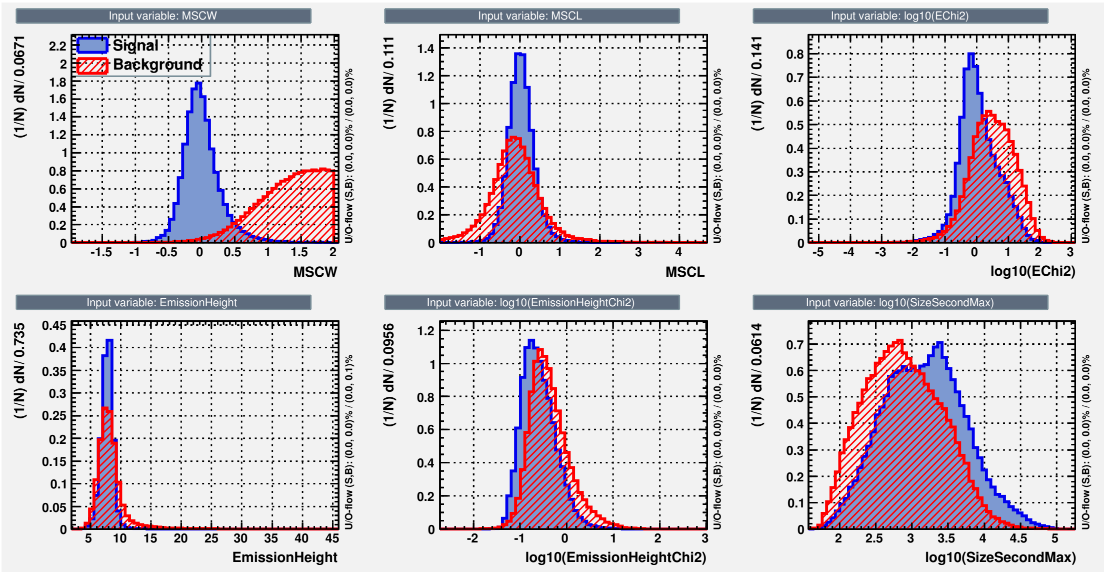

Figure 3: Distribution of cut variables for events with high energies ($\approx$ 10 TeV). Signal events are simulated gamma rays, background events consist of proton events only.

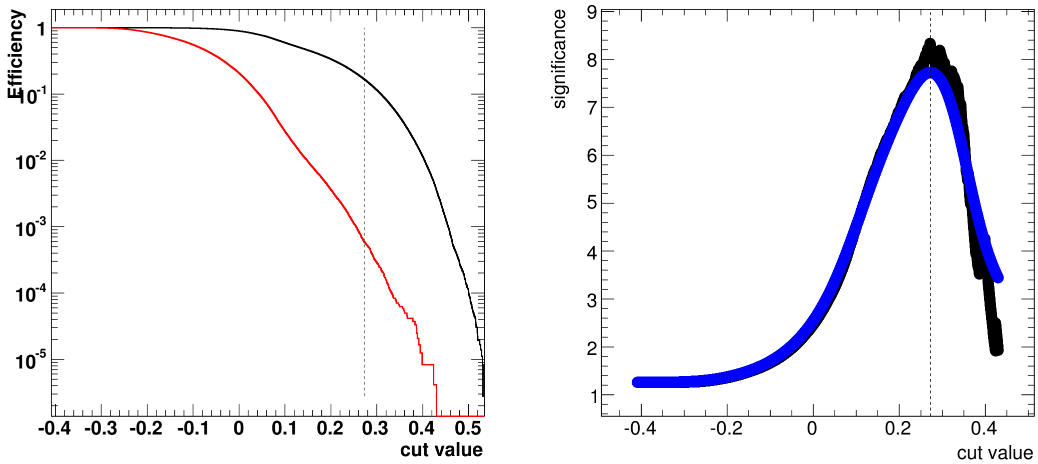

Figure 4: Left: Example of signal (black line) and background (red line) efficiencies as function of the BDT cut value. Right: Significance as function of the BDT cut value assuming a Crab Nebula-like spectrum and proton background events for a arbitrary energy bin. The blue line shows the smoothed version of the actual values (black points) and is used for the determination of the maximum of the curve (indicated in both graphs by the dashed vertical line).

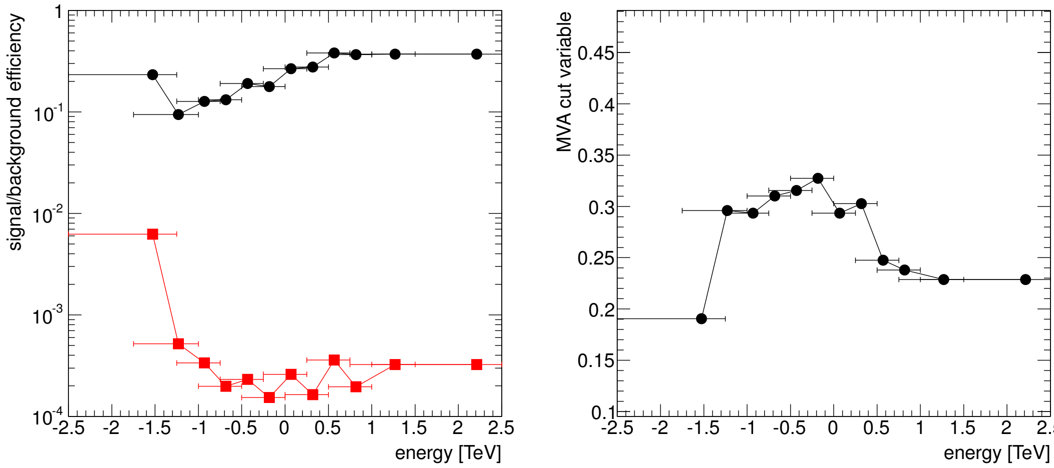

Figure 5: Example of an optimization of the BDT cut value for array E and an observation time of 50 hours. Left: Signal (black) and background (red) efficiency as function of energy. Right: BDT cut vale as function of energy. Error bars on the energy scale indicate the width of the energy bins.

is a single cut variable in the interval -0.5 to 0.5. The optimal BDT cut is selected at the value giving highest significance, assuming a Crab Nebula-like gamma-ray spectrum and proton background spectrum as listed in the appendix. Figure 4 shows examples of the signal and background efficiencies and the significance as function of the BDT cut value. Figure 5 shows the energy dependence of the resulting signal/background efficiency and the BDT cut value (note that this is the signal/background efficiency after pre-(quality) cuts). The optimal BDT cut value depends additionally on the observation time assumed and is determined for each energy bin before the gamma-hadron separation procedure. This procedure does not work at very-high energies (above several 10s of TeV) due to the limited statistics in protons events in the used MC sample. Here, the optimal cut value from the last bin with sufficient statistics is used.

7 Example results: instrument response functions and sensitivities for subarray E

In the following results on the subarray E are presented and compared to other analysis results from the CTA WP-Phys wiki page[4] (end of November 2011). The purpose of this section is to demonstrate the consistency of the evndisplay analysis, not to present a detailed comparison with other results.

The following parameters have been used for the evndisplay:

1. image cleaning levels: 11/5.5 (LST), 11/5.5 (MST), 9.2/5 (SST)

2. quality cuts before stereo reconstruction: minimum number of image/border pixels per image (>4 for LST and MST,>3 for SST), no cut on image size, maximum loss value (<0.2 for LST, <0.5 for MST and SST)

3. for off-axis results showers are binned in camera offset (degrees): [0,1], [1,2], [2,3], [3,3.5], [3.5,4], [4,4.5], [4.5,5], [5.,5.5], [5.5, 6]

[4]http://www.cta-observatory.org/ctawpcwiki/index.php/WP_MC#Interface_to_WP_PHYS

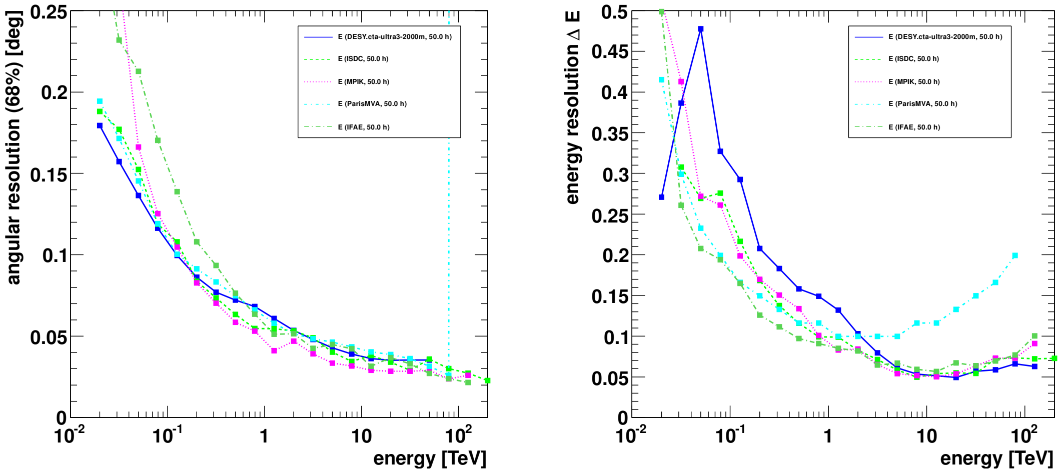

Figure 6: Angular (left) and energy resolution (right) for subarray E after gamma-hadron separation cuts from this analysis (blue line) in comparison to other CTA analysis results.

4. to increase the number of proton showers available for the sensitivity calculations, a constant response with camera offset of proton events in these angular bins is assumed. That means that for example for all on-axis results, protons with arrival directions in the bin 0-1 degrees are selected.

5. all results have been computed for 50 h of observations and the standard requirements: 5 bins per energy decade, with at least 10 events per bin, and with a significance above 5σ (after equation 17 in [10]), the number of gammas above 5% of the background rate and a on-to-off source ratio of 1/5[5].

Figures 6, 7, 8 and 9, show angular and energy resolution, effective areas, background rates and differential sensitivities. All results are on-axis, as long as not stated differently. Note that the optimization procedure and the loose quality cuts applied in this analysis favor high signal rates, both energy and angular resolution or not optimized (especially the energy reconstruction needs some refinement).

8 Conclusion

We described in this short note the different analysis steps and the most important methods used in evndisplay. It is a well tested package used for both data and MC analysis and open for any additions and development input. For a CTA analysis, the only dependency of evndisplay are root and hessio/eventio (no VERITAS software libraries are needed). The analysis package can be used by any member of CTA. Access to the DESY svn code repository is given at this stage on request, we are improving currently the documentation and will then provide a general release of the software to the consortium members (the source code is available from https://znwiki3.ifh.de/CTA/Eventdisplay%20Software).

[5] see http://www.cta-observatory.org/ctawpcwiki/index.php/Wp_mc_sensitivity_calc

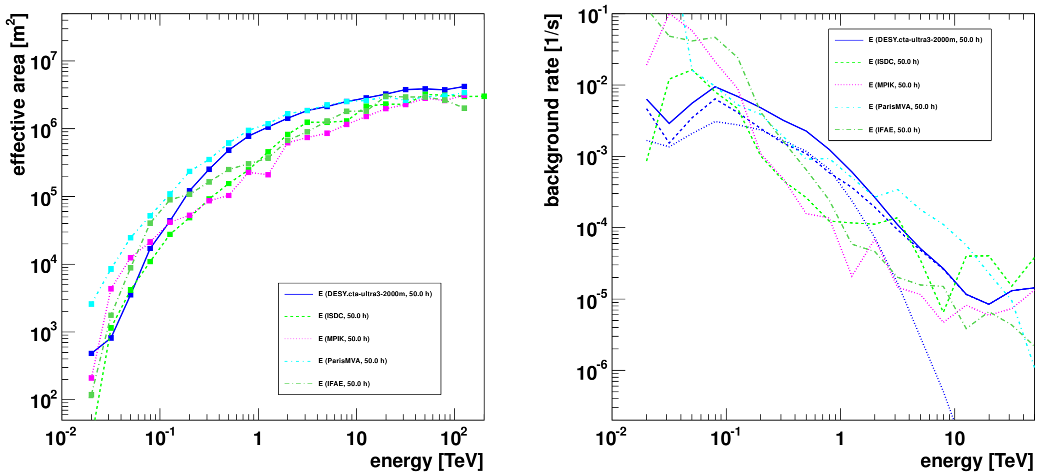

Figure 7: Left: Gamma-ray on-source effective area for subarray E from this analysis (blue line)in comparison to other CTA analysis results.. Right: Background rates for subarray E from this analysis (black solid line) in comparison to other CTA analysis results. Additionally drawn are the proton (dashed blue line) and electron (dotted blue line) background rates (DESY analysis only).

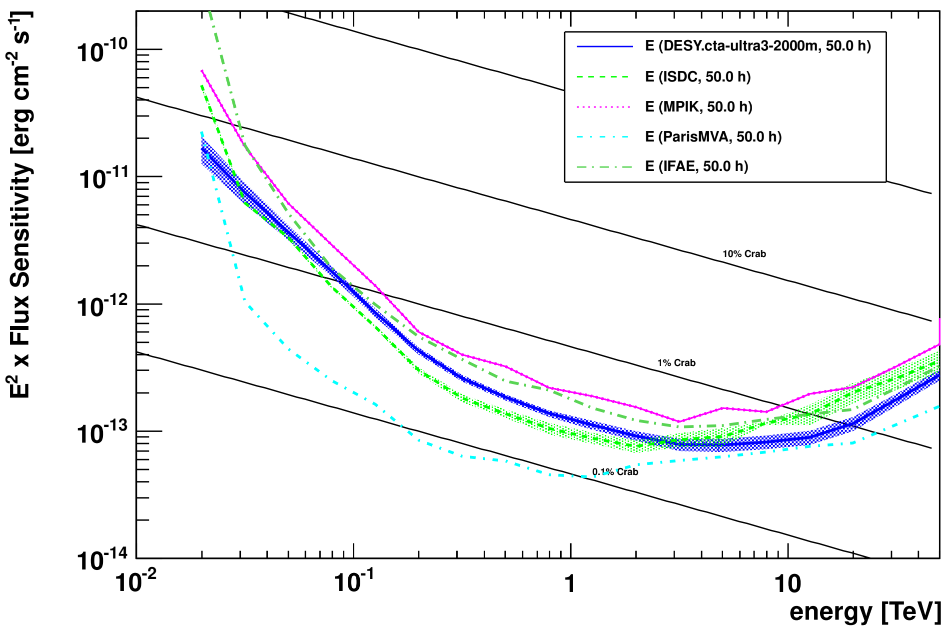

Figure 8: Differential sensitivity for array E for 50 h of on-axis observations from this analysis (blue line) in comparison to other CTA analysis results.

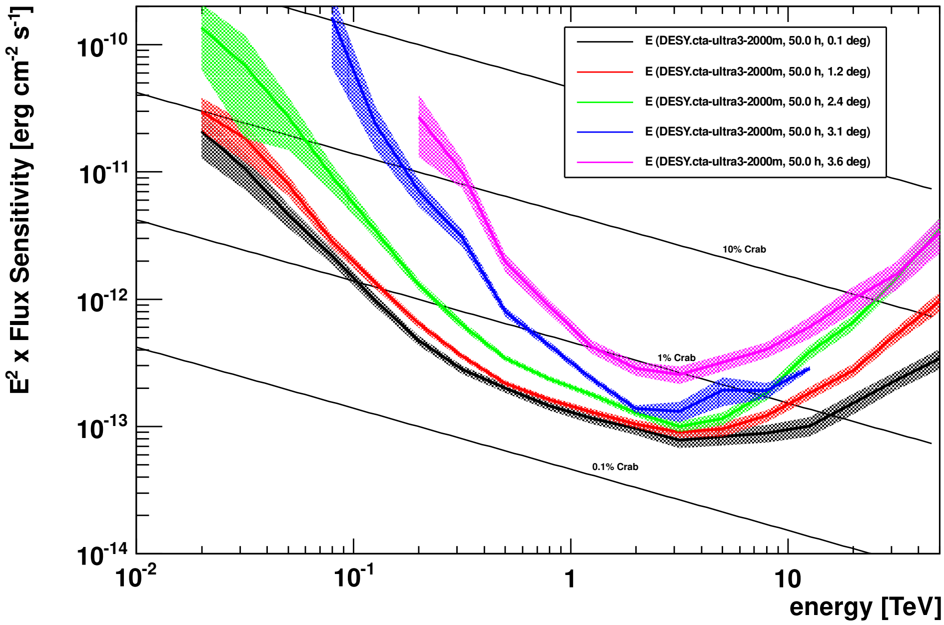

Figure 9: Differential sensitivity for array E for 50 h of observations from this analysis (blue lines) for different off-axis angles: $0^o, 1.2^o, 2.4^o, 3.6^o$ (counted from bottom).

9 Appendix

9.1 MC raw data

The CTA simulations used here are produced at MPIK in Heidelberg (CTA Prod-1). We made use of the full raw data (hessio/eventio format), for a complete listing of the data see the corresponding website[6].

9.2 Cosmic ray spectra

The cosmic ray and gamma-ray spectra used in this analysis are listed in the following. The Crab Nebula spectrum for gamma-rays (HEGRA spectrum [4]):

$$\Phi_{gamma} = 2.83 \times 10^{-7} \left(\frac{E}{\mathrm{TeV}}\right)^{-2.62} \mathrm{TeV}^{-1}\mathrm{s}^{-1}\mathrm{m}^{-2} \tag{1}$$

Proton spectrum according to ATIC measurements:

$$\Phi_{protons} = 0.098 \times 10^{-4} \left(\frac{E}{\mathrm{TeV}}\right)^{-2.62} \mathrm{TeV}^{-1}\mathrm{s}^{-1}\mathrm{m}^{-2}\mathrm{sr}^{-1} \tag{2}$$

Electron spectrum follows a fit to the Fermi LAT [1] and HESS [3] data applying a broken power law function:

$$\Phi_{electrons} = \Phi_0 \left(\frac{E}{E_b}\right)^{-\Gamma_1} \cdot \left(1 + \left(\frac{E}{E_b}\right)^{1/\alpha}\right)^{-(\Gamma_2-\Gamma_1)\alpha} \tag{3}$$

with $\Phi_0 = 2.3 \times 10^{-9}$ $\mathrm{TeV}^{-1}\mathrm{s}^{-1}\mathrm{m}^{-2}\mathrm{sr}^{-1}$, $\alpha = 0.4$, $\Gamma_1 = 3.07$, $\Gamma_2 = 5.$, $E_b = 1.7$ TeV.

The impact of other components of the hadronic background was tested with helium showers and found to be negligible.

[6] http://www.mpi-hd.mpg.de/hfm/CTA/internal/MC/

References

[1] Abdo, A.A. et al (LAT collaboration) 2009, Phys.Rev.Letters 102, 181101

[2] Aharonian, F. et al. 1997, Astroparticle Physics, 6 343.

[3] Aharonian, F. et al (HESS collaboration) 2009, A& A 508, 561

[4] Aharonian, F. et al. (HEGRA collaboration) 2004, ApJ 614, 897

[5] Farnier, C. and Lenain, J.-P., 2011, CTA-WP-MC Internal note

[6] Hillas, M. 1985, Proc. of the 19th ICRC (La Jolla, IL), 3, 445

[7] Holder, J. et al (VERITAS collaboration) 2006, Astroparticle Physics 25, 391

[8] Hofmann, W. et al 1999, Astroparticle Physics, 12, 135

[9] Krawczynski, H. et al. 2006, Astroparticle Physics, 25, 380

[10] Li, T.P. & Ma, Y.Q. 1983, ApJ, 272, 317

[11] Maier, G. et al (VERITAS collaboration) 2007, Proc. of the 30th ICRC (Merida, Mexico)

[12] Maier, G. et al (AGIS collaboration) 2009, Proc. of the 31th ICRC (Lodz, Poland)