

A general data quality inspection for Gamma-Ray Bursts searches with HAWC

Cederik de León*, Humberto Salazar, Luis Villaseñor

Benemérita Universidad Autónoma de Puebla.

Av. San Claudio y 18 sur, colonia San Manuel Puebla México.

E-mail: cederik@gmail.com

for the HAWC Collaboration[†]

The High Altitude Water Cherenkov (HAWC) is a wide field-of-view gamma-ray observatory sensitive to gamma-rays in the 300 GeV – 100 TeV energy range, located in Mexico at an altitude of 4,100 m above sea level. The detector consists of 300 Water Cherenkov Detectors with a volume of 200,000 l each, having a footprint of 22,000 m^2 , a duty cycle of >95% and a field of view (FoV) of 2sr [1]. In this work we present a general data quality inspection of HAWC data for the years between 2014 and 2017 and also for two specific data periods selected on the base that GRB 150416A and GRB 160301A detected by FERMI occurred within the HAWC FoV, allowing to search for them in the HAWC data.

35th International Cosmic Ray Conference — ICRC2017

10–20 July, 2017

Bexco, Busan, Korea

*Speaker.

[†]For a complete author list, see <http://www.hawc-observatory.org/collaboration/icrc2017.php>

1. Introduction

Many events can produce excesses in the counts in the HAWC detector, which could have different origins, such as: electrical atmospheric activity (i. e. thunder storms, lightning), noise due to the experiment electronics and/or technical operations (i. e. maintenance and/or upgrades), unknown fluctuations and signals from astrophysical objects (i. e. transients like GRBs) [1], [3].

The understanding of the data quality behavior of a certain period of interest, regardless the nature of the analysis to be performed, is essential. Lack of data or bad data periods will lead to a biased analysis or even make it impossible. A GRB analysis (like in [3]) and other astrophysical phenomena (like in [4]), needs data of enough quality to ensure unbiased variables to perform the search of signal associated to this astrophysical events.

For this work, we used the HAWC Quality and Monitoring Data Base (QMDB) and custom made computer scripts and programs in order to retrieve and analyze the related data quality variables. Here we show a method to estimate the stability for the duration, the number of events and zenithal angle from a subset of HAWC data (sub run) acquired and stored in a HAWC custom file format and electric field activity for selected data periods: 2014 – 2017 and two specific dates 2015-04-16 and 2016-03-01, where two GRBs occurred, triggered an external alert, were within the HAWC FoV and HAWC was in normal operation.

2. HAWC datasets

The data acquired by the HAWC DAQ [1] is processed in a certain way that events are reconstructed and stored on–line, other information relevant for experiment, like scalers and TDC are as well stored. The data is organized in general blocks called runs, each run is labeled with a unique consecutive serial number that serves as identifier. A run contain sub runs (data files) as well identified by a name and a unique serial number. Sub run dataset file names are: reconstructed events (*reco*), TDC data (*tdc*), triggered data (*trig*) and raw data (*raw*), all of those files are generated by their own on–line process.

A typical run is expected to be a complete cycle of 24 hours with continuous data, nevertheless, due experiment operations (i. e. calibration, maintenance, repairs or outages) or external environmental condition (Weather conditions in general, particularly thunder storms or extreme electric field variations near to the HAWC Site) [2], some runs last less than expected. When a run ends the experiment control start a new one (just in case of explicit operations of the experiment require to stop the data acquisition the automatic run stop–start process are not performed).

Each sub run file stores ~ 125 seconds of information; this is ~ 691 sub run files per dataset (~ 3000 files per run, $> 2\text{TB}$ per day).

2.1 Quality Monitoring Data Base

A way to track the experiment acquired data is implemented in a process that the statistics extraction, from the reconstructed files, is performed by a computer program *diagnostics-rec-stats* that uses the HAWC AERIE Framework.

diagnostics-rec-stats gets information related to 54 variables for a given reconstructed data file; some values are directly extracted from the file but in the majority are statistics computed

from the entire data (events) contained into this file. The results are stored into the Quality Monitor Data Base (QMDB), a MySQL data base. The data can be accessed for the HAWC collaboration.

There is two main groups of data sets with exactly the same variable definition in QMDB, the HAWC On–line and the HAWC Production. The main distinction between this datasets is the way that the HAWC data are reconstructed.

3. GRB external alerts and HAWC Data

HAWC have a data base that stores information related to alerts triggered from other experiments (external alerts). We retrieve information from external alerts in order to explore the possibility of analysis. The selection was done using the conditions shown in Table 1.

Conditions	
Date range	2014–2017 (May)
Zenith angle (HAWC)	$[0 \leq \theta \leq 45]$ degrees
Instrument	FERMI, SWIFT

Table 1: GRBs external alerts conditions applied to search and retrieve its relevant information as: Instrument, Date and Time (UTC), Zenith Angle (relative to HAWC)

After applying the conditions mentioned above, we got a the fraction of external alerts within the HAWC FoV as shows in Table 2, the plot in Figure 1 shows the GRB external alerts distribution within $0 \leq \theta \leq 90$ degrees, relative to HAWC (top plot). The open red circles represents the SWIFT alerts, blue dots represents the FERMI alerts, the HAWC instantaneous Field of View (FoV) is in the $0 \leq \theta \leq 45$ degrees in shaded region (bottom plot).

Year	Fraction of GRB within HAWC FoV
2014	~ 0.14
2015	~ 0.26
2016	~ 0.23
2017	~ 0.39

Table 2: Fraction of GRB external alerts within the FoV ($\text{\#external alerts (Within HAWC FoV)} \ 0 \leq \theta \leq 45 \text{ [degrees]} / \text{\#external alerts } 0 \leq \theta \leq 90 \text{ [degrees]}$)

For each GRB external alert between the date range (see Table 1) we ask for a data duration of 24 hours, with a tolerance ≤ 0.03 , will be used for the GRB external alert and always will contain the triggered alert Date–Time. In this work we centered the GRB external alert, starting the dataset period 12 hours before the GRB external alert trigger and 12 after the GRB external alert trigger.

For the sub run duration and number of events stability we used the σ/μ ratio, this value was estimated using a computer simulation considering that the data follows a uniform distribution, without discontinuities. We made 1000 simulations for a 24 hours of data. Every simulations contains a set of ~ 691 sub run with their duration values 125 ± 0.001 in seconds, noise was

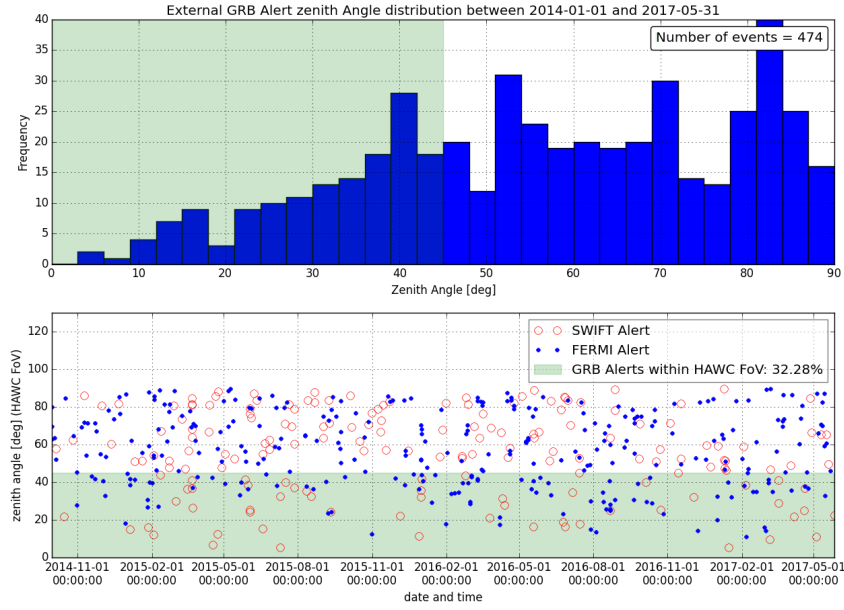


Figure 1: Distribution of the GRB external alerts within $0 \leq \theta \leq 90$ degrees relative to the HAWC zenith (Top plot). GRB external alerts, the open red circles represents the SWIFT alerts, the blue dots represents the FERMI alerts, the HAWC instantaneous Field of View (FoV) is in the $0 \leq \theta \leq 45$ degrees in shaded region (Bottom plot).

Cuts	
Duration required	± 12 hours (GRB alert centered)
Duration ratio ($\frac{\text{data duration}}{\text{duration required}}$)	≥ 0.97
Data duration stability (σ/μ ratio)	< 0.05
Number of events stability (σ/μ ratio)	< 0.05
Zenith angle stability	$20 < \text{med}(\theta) < 25$

Table 3: GRBs external alerts dataset quality cuts applied in the period related at every external alert between 2014 and 2017 (may)

introduced randomly and consist a fraction of the total number of sub run entries: $[0.0025 - 0.05]$. The “number of events” in a sub run is considered uniform as well in the 24 hours period and its simulation was performed in the same way, described above.

σ/μ ratio values ≥ 0.05 are considered as poor quality period data set, it implies a “noisy” dataset (noise is considered as extreme values far from the mean and in a proportion $N/S \geq 0.0075$).

We seek for the duration ratio condition mets, if not, the requested total duration are then reduced symmetrically (e. g. 1000 seconds both sides, before and after. This is an arbitrary value and could change), this process continues until the ratio (≥ 0.97) is reached. If the duration ratio is never reached the event is considered as “poor data quality”.

4. HAWC data quality inspection for GRBs external alerts

We selected and inspected the 24 hours data periods for two GRB external alerts: GRB150416A and GRB160301A, their results are shown in Table 4, Figure 2 contains information for every sub run in the GRBs 24 hours data period.

Instrument	date and time (UTC)	Best duration found [s]	duration ratio	σ/μ -value
FERMI GBM	2015-04-16 18:33:25	43200	~ 0.97	~ 0.08
FERMI GBM	2016-03-01 05:10:18	43200	~ 0.99	~ 0.03

Table 4: GRB150416A and GRB160301A data quality inspection

Comparing datasets, we seek for another two (of nine) ‘‘poor data quality’’ GRBs external alerts: GRB150323A and GRB150520A, their results are shown in Table 5, Figures 3 contains information for every sub run in the GRBs 24 hours data period.

Instrument	date and time (UTC)	Best duration found [s]	duration ratio	σ/μ -value
SWIFT BAT	2015-03-23 02:49:14	ND	ND	> 0.18
FERMI GBM	2015-05-20 21:25:34	ND	ND	> 0.10

Table 5: GRB150323A and GRB150520A data quality inspection (ND = No-Data).

The environmental conditions at the HAWC site have special importance, every GRB external alert and its associate data period includes, if available, the electric field (EF) related data, in order to consider and explore a possible bias on data acquisition; we observed that only on periods of extreme electric field variations (i. e. thunder storms near to the site) the data stability is affected.

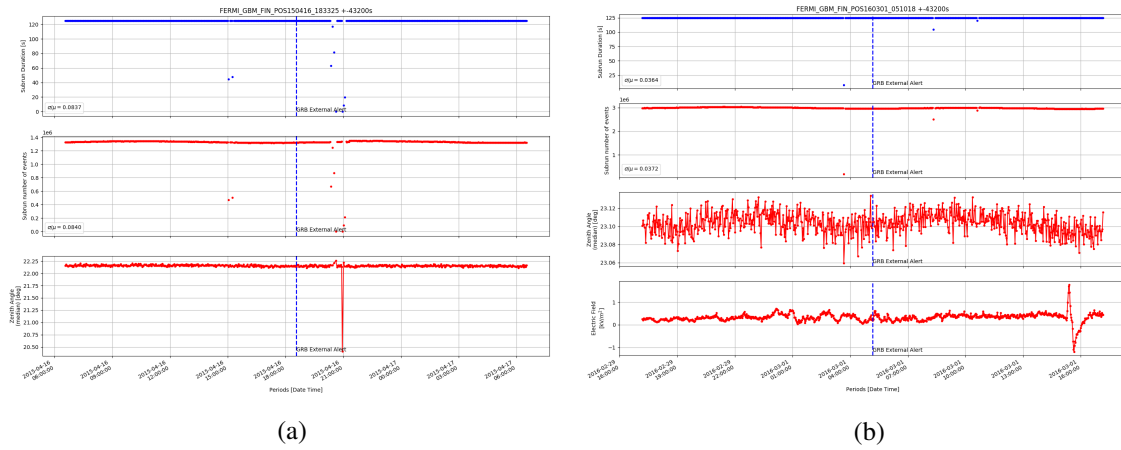


Figure 2: (2a) GRB150416A data quality inspection: (Top to down) sub run duration, number of events per sub run, sub runs median of the zenith angle, for this period HAWC has no electric field records. (2b) GRB160301A DQ Inspection: (Top to down) sub run duration, number of events per sub run, sub runs median of the zenith angle, electric field activity.

Figures 3a and 3b shows the sub run duration and number of events are spread and the median of zenith angle drops dramatically, for the GRB150520A an increment in the electric field is present.

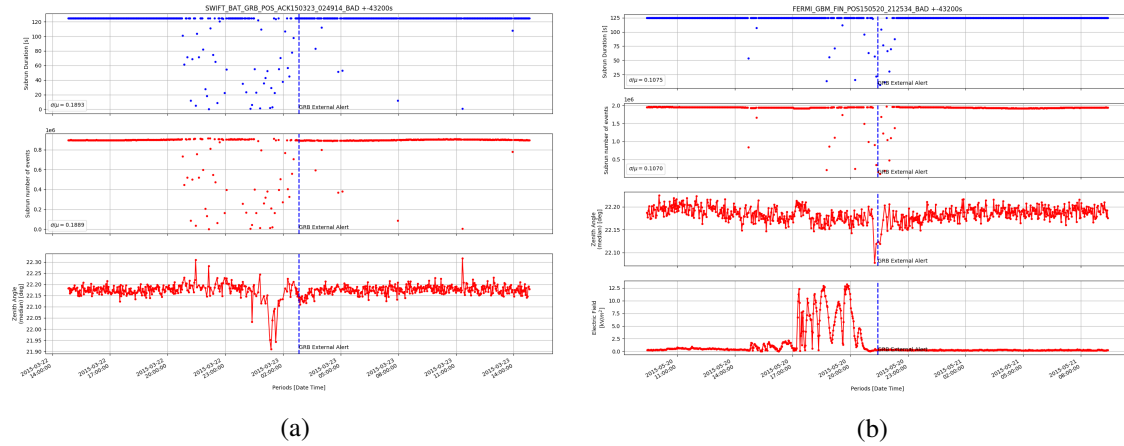


Figure 3: (3a) GRB150323A data quality inspection: (Top to down) sub run duration, number of events per sub run, sub runs median of the zenith angle, for this period HAWC has no electric field records. (3b) GRB150520A data quality inspection. Top to down: sub run duration, number of events per sub run, sub runs median of the zenith angle, electric field activity.

5. GRB150416A and GRB160301A analysis trial with data quality inspection

As a preliminary analysis we applied a Moving Average (MA) analysis on a set of TDC files belonging to GRB external alerts for GRB150416A and GRB160301A; the data duration stability behavior for the TDC files shows the same behavior as mentioned in Table 4. We use the same data period of 24 hours to perform a preliminary Moving Average analysis on them. We also used the information from Table 5 (poor datasets) to perform the same analysis, nevertheless due their poor data quality (mainly in duration and stability) we did not get any results.

For the GRB150416A, Figure 4 shows a preliminary analysis using the MA (this trial version uses a rebin of $x = 20s$ and considered 10 samples for the Moving Average calculation, Figure 4a background estimation (top), excesses histogram in the blind region $\pm 1000s$ (bottom). Figure 4b signal behavior in the dataset used to estimate the background (top), excesses distribution in the blind region $\pm 1000s$, shaded area are $T_0 + 100$ (bottom). In this signal we saw an artifact near the second $55000 + 1.1132 \times 10^9$, this need to be studied as an special case ($T_0 =$ Date and time when the GRB triggered the external alert reported).

For the GRB160301A, Figure 5 shows a preliminary analysis using the MA (this trial uses a rebin of $x = 20s$ and considered 10 samples for the Moving Average calculation, Figure 5a background estimation (top), excesses histogram in the blind region $\pm 1000s$ (bottom). Figure 5b the signal behavior in the dataset used to estimate the background (top), excesses distribution in the blind region $\pm 1000s$, shaded area are $T_0 + 100$ (bottom) ($T_0 =$ Date and time when the GRB triggered the external alert, reported by its experiment).

6. Summary

The HAWC experiment datasets analyzed for the GRBs external alerts between 2014–2017 (may) has been explored and shows a reliable source of data to perform analysis with them.

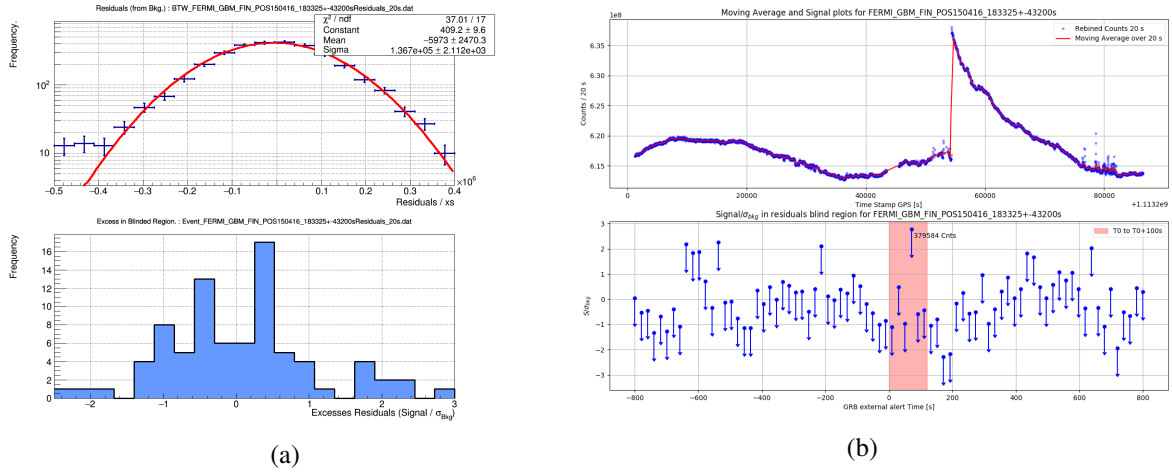


Figure 4: GRB150416A: (4a) Background estimation, gaussian fit (top), excesses histogram (bottom), (4b) Moving Average signal (top), excesses distribution, shaded area are $T_0 + 100$ (bottom),

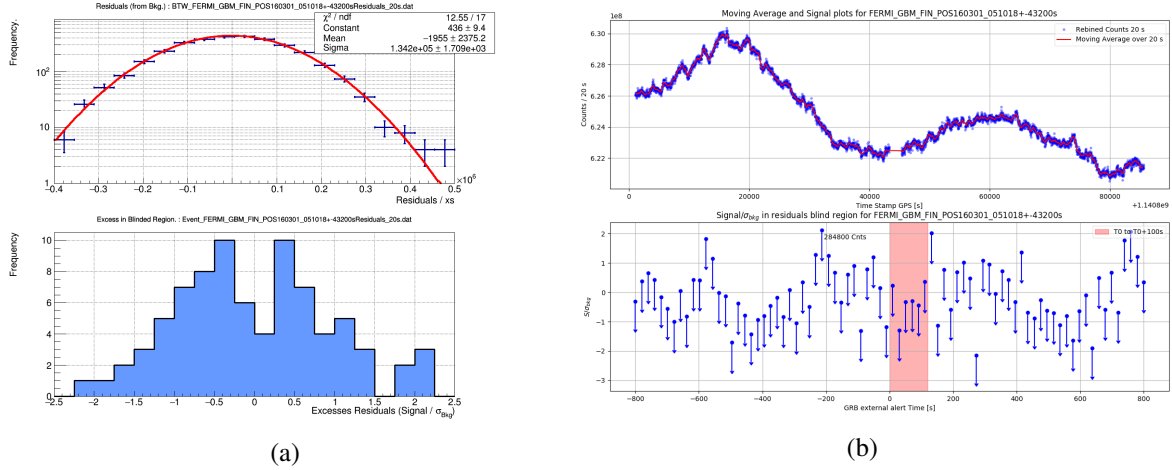


Figure 5: GRB160301A: (5a) Background estimation, gaussian fit (top), excesses histogram (bottom), (5b) Moving Average Moving Average signal (top), excesses distribution, shaded area are $T_0 + 100$ (bottom)

Between 2014–2017 (may) a total of 147 GRB External alerts in the HAWC FoV has been registered, 26 of them (< 18%) were marked with this method and its cuts as “poor data quality” due a lack of duration, 9 (< 7%) never reached the requested duration ratio nor the σ/μ threshold even with the period reduction process applied.

This general data quality inspection method, for the HAWC experiment, is an easy and fast way to find stable data periods. Environmental conditions needs to be taken into account, if activity of this kind is present in the selected periods, it should be treated in the most appropriate way for their purposes.

7. Acknowledgments

We acknowledge the support from: the US National Science Foundation (NSF); the US Department of Energy Office of High-Energy Physics; the Laboratory Directed Research and Development (LDRD) program of Los Alamos National Laboratory; Consejo Nacional de Ciencia y Tecnología (CONACyT), México (grants 271051, 232656, 260378, 179588, 239762, 254964, 271737, 258865, 243290, 132197), Laboratorio Nacional HAWC de rayos gamma; L'OREAL Fellowship for Women in Science 2014; Red HAWC, México; DGAPA-UNAM (grants IG100317, IN111315, IN111716-3, IA102715, 109916, IA102917); VIEP-BUAP; PIFI 2012, 2013, PROFOCIE 2014, 2015; the University of Wisconsin Alumni Research Foundation; the Institute of Geophysics, Planetary Physics, and Signatures at Los Alamos National Laboratory; Polish Science Centre grant DEC-2014/13/B/ST9/945; Coordinación de la Investigación Científica de la Universidad Michoacana. Thanks to Luciano Díaz and Eduardo Murrieta for technical support; National Laboratory; Polish Science Centre grant DEC-2014/13/B/ST9/945; The authors thankfully acknowledge the computer resources, technical expertise and support provided by the Laboratorio Nacional de Supercomputo del Sureste de México, CONACYT Network of national laboratories. Special thanks to: Eduardo Moreno and Humberto Martínez.

References

- [1] A.U. Abeysekara et Al, *On the sensitivity of the HAWC observatory to gamma-ray bursts*, *Astroparticle Physics*, 35, 10 641-650, [arXiv:1108.6034]
- [2] A.U. Abeysekara et Al, *VAMOS: A pathfinder for the HAWC gamma-ray observatory*, *Astroparticle Physics*, 62, 125-133, [arXiv:1408.3477]
- [3] A.U. Abeysekara et Al, *Search for Gamma-Rays from the Unusually Bright GRB 130427A with the HAWC Gamma-Ray Observatory*, *The Astrophysical Journal*, 800, 2, [arXiv:1410.1536]
- [4] A.U. Abeysekara et Al, *Daily Monitoring of TeV Gamma-Ray Emission from Mrk 421, Mrk 501, and the Crab Nebula with HAWC*, *The Astrophysical Journal*, 841, 2, [arXiv:1703.06968]
- [5] Sparks, Kathrynne, *Search for high energy emission from GRBs with the HAWC Observatory*, *Proceedings, 33rd International Cosmic Ray Conference (ICRC2013): Rio de Janeiro, Brazil, July 2-9, 2013*, pp 0783.
- [6] Hüntemeyer, Petra and Albert, Hugo and Solares, Ayala, *HAWC Sensitivity to Diffuse Emission*, *Proceedings, 33rd International Cosmic Ray Conference (ICRC2013): Rio de Janeiro, Brazil, July 2-9, 2013*, pp 0975.
- [7] Alejandro Lara, *HAWC and Solar Energetic Transient Events*, *Proceedings, 33rd International Cosmic Ray Conference (ICRC2013): Rio de Janeiro, Brazil, July 2-9, 2013*, pp 0929.
- [8] Dirk Lennarz, *Sensitivity of the HAWC Observatory to Gamma-ray Bursts Using the Scaler System*, *Proceedings, 33rd International Cosmic Ray Conference (ICRC2013): Rio de Janeiro, Brazil, July 2-9, 2013*, pp 1160.