

## Data Quality Monitoring system of the Baikal-GVD experiment

V.A. Allakhverdyan,<sup>a</sup> A.D. Avrorin,<sup>b</sup> A.V. Avrorin,<sup>b</sup> V.M. Aynutdinov,<sup>b</sup> R. Bannasch,<sup>c</sup> Z. Bardačová,<sup>d</sup> I.A. Belolaptikov,<sup>a</sup> I.V. Borina,<sup>a</sup> V.B. Brudanin,<sup>a,1</sup> N.M. Budnev,<sup>e</sup> V.Y. Dik,<sup>a</sup> G.V. Domogatsky,<sup>b</sup> A.A. Doroshenko,<sup>b</sup> R. Dvornický,<sup>a,d</sup> A.N. Dyachok,<sup>e</sup> Zh.-A.M. Dzhlkibaev,<sup>b</sup> E. Eckerová,<sup>d</sup> T.V. Elzhov,<sup>a</sup> L. Fajt,<sup>f</sup> S.V. Fialkovski,<sup>g,1</sup> A.R. Gafarov,<sup>e</sup> K.V. Golubkov,<sup>b</sup> N.S. Gorshkov,<sup>a</sup> T.I. Gress,<sup>e</sup> M.S. Katulin,<sup>a</sup> K.G. Kebkal,<sup>c</sup> O.G. Kebkal,<sup>c</sup> E.V. Khramov,<sup>a</sup> M.M. Kolbin,<sup>a</sup> K.V. Konischev,<sup>a</sup> K.A. Kopański,<sup>h</sup> A.V. Korobchenko,<sup>a</sup> A.P. Koshechkin,<sup>b</sup> V.A. Kozhin,<sup>i</sup> M.V. Kruglov,<sup>a</sup> M.K. Kryukov,<sup>b</sup> V.F. Kulepov,<sup>g</sup> Pa. Malecki,<sup>h</sup> Y.M. Malyshkin,<sup>a</sup> M.B. Milenin,<sup>b</sup> R.R. Mirgazov,<sup>e</sup> D.V. Naumov,<sup>a</sup> V. Nazari,<sup>a</sup> W. Noga,<sup>h</sup> D.P. Petukhov,<sup>b</sup> E.N. Pliskovsky,<sup>a</sup> M.I. Rozanov,<sup>j</sup> V.D. Rushay,<sup>a</sup> E.V. Ryabov,<sup>e</sup> G.B. Safronov,<sup>b</sup> B.A. Shaybonov,<sup>a</sup> M.D. Shelepov,<sup>b</sup> F. Šimkovic,<sup>a,d,f</sup> A.E. Sirenko,<sup>a</sup> A.V. Skurikhin,<sup>i</sup> A.G. Solovjev,<sup>a</sup> M.N. Sorokovikov,<sup>a,\*</sup> I. Štekl,<sup>f</sup> A.P. Stromakov,<sup>b</sup> E.O. Sushenok,<sup>a</sup> O.V. Suvorova,<sup>b</sup> V.A. Tabolenko,<sup>e</sup> B.A. Tarashansky,<sup>e</sup> Y.V. Yablokova,<sup>a</sup> S.A. Yakovlev<sup>c</sup> and D.N. Zaborov<sup>b</sup>

<sup>a</sup>Joint Institute for Nuclear Research, Dubna, Russia

<sup>b</sup>Institute for Nuclear Research, Russian Academy of Sciences, Moscow, Russia

<sup>c</sup>EvoLogics GmbH, Berlin, Germany

<sup>d</sup>Comenius University, Bratislava, Slovakia

<sup>e</sup>Irkutsk State University, Irkutsk, Russia

<sup>f</sup>Czech Technical University in Prague, Prague, Czech Republic

<sup>g</sup>Nizhny Novgorod State Technical University, Nizhny Novgorod, Russia

<sup>h</sup>Institute of Nuclear Physics of Polish Academy of Sciences (IFJ PAN), Kraków, Poland

<sup>i</sup>Skobel'syn Institute of Nuclear Physics, Moscow State University, Moscow, Russia

<sup>j</sup>St. Petersburg State Marine Technical University, St.Petersburg, Russia

E-mail: [sorokovikov@jinr.ru](mailto:sorokovikov@jinr.ru)

The main purpose of the Baikal-GVD Data Quality Monitoring (DQM) system is to monitor the status of the detector and collected data. The system estimates quality of the recorded signals and performs the data validation. The DQM system is integrated with the Baikal-GVD's unified software framework ("BARS") and operates in quasi-online manner. This allows us to react promptly and effectively to the changes in the telescope conditions.

37<sup>th</sup> International Cosmic Ray Conference (ICRC 2021)

July 12th – 23rd, 2021

Online – Berlin, Germany

<sup>1</sup>Deceased.

\*Presenter

## 1. Introduction

Baikal-GVD is a km<sup>3</sup>-scale neutrino telescope currently under construction in Lake Baikal [1, 2]. It produces about 100GB of data every day, and verification of large amount of data is necessary. The main purpose of the Baikal-GVD Data Quality Monitoring (DQM) system is to monitor the status of the detector and collected data. The system estimates quality of the recorded signals and performs the data validation. The DQM system is integrated with the Baikal-GVD's unified software framework "BARS" [3] and operates in quasi-online manner. The characteristics of data records under monitoring form two groups of parameters related to the distributions described the Poissonian character of events flow, and also the signal charge measurements. The first group consists of the exponential distribution, uniform distribution and Poissonian one. The second group comprises the single photoelectron (p.e) distribution, channel noise rate, and trigger thresholds. We fit the distributions with expected functions by means of the minimum chi square method, and  $\chi^2/NDF$  parameter is used to obtain quality estimations.

## 2. Estimation of the Poissonian character of events flow

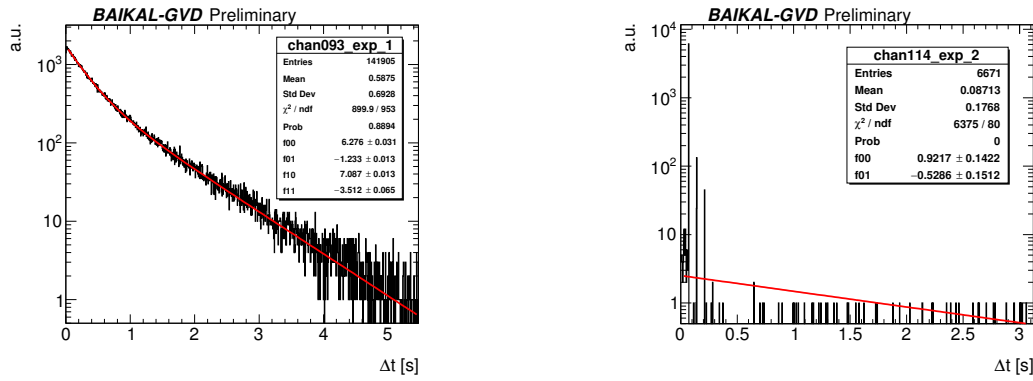
The signals generated by atmospheric muons and lake random noise are dominated in the total data flow of the telescope. Detection of this type of events is Poissonian-like process and recorded telescope data obey the next three distribution – exponential, Poissonian, and uniform. These distributions are checked for all levels of the telescope – for channel (single optical module (OM)), for section (section is a minimal structure unit of the detector, twelve linked optical modules), and for the whole cluster. The calibration systems used for PMT calibration, such as LED matrices or laser sources, operate with fixed frequency and emit the light that should worsen the shape of distributions, and therefore the fit quality decreases. On the other hand, unstable environmental conditions of the telescope [4], namely, a possible dynamical changing of the count rate during the seance of data taken (so-called "run") also lead to deterioration of distributions.

### 2.1 Exponential distribution test

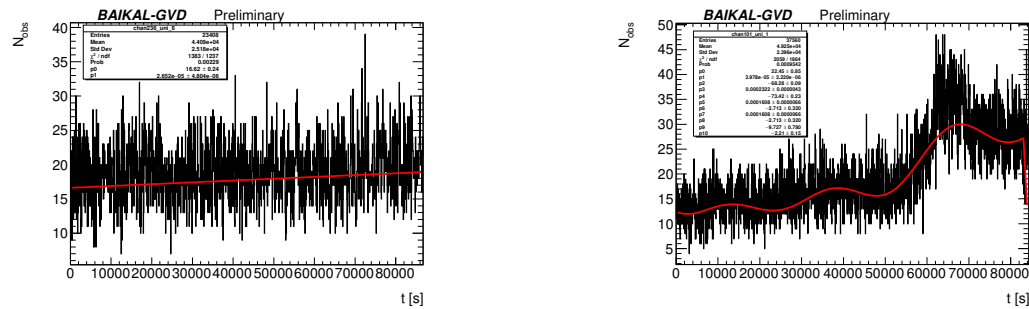
The time difference between two consecutive events is described by the exponential function. Figure 1 presents distributions for such time differences obtained for some selected channels. We take into account dynamical changing of the count rate during run, as it is shown in left panel. For such cases we fit the distribution by a set of exponential functions with different slopes, that correspond to different stable values of the count rate. There are clearly effect of laser source applied during calibration run, consequently the exponential distribution worsens and  $\chi^2/NDF$  parameter is rather high (Figure 1 (right)).

### 2.2 Uniform distribution test

Uniform distribution is described the count rate of the recorded events during a run. It is expected that the distribution of the rate of atmospheric muons and lake random noise in stable environmental condition is linear with certain slope and should be described by the first order polynomial function as it is shown in Figure 2 (left), where uniform distribution for some selected channel is shown. On the other hand, when environment during the run becomes unstable and the



**Figure 1:** Exponential distributions for some selected channels: standard run (left), calibration run with laser source (right).



**Figure 2:** Uniform distributions for some selected channels: stable (left), and unstable (right) environmental condition.

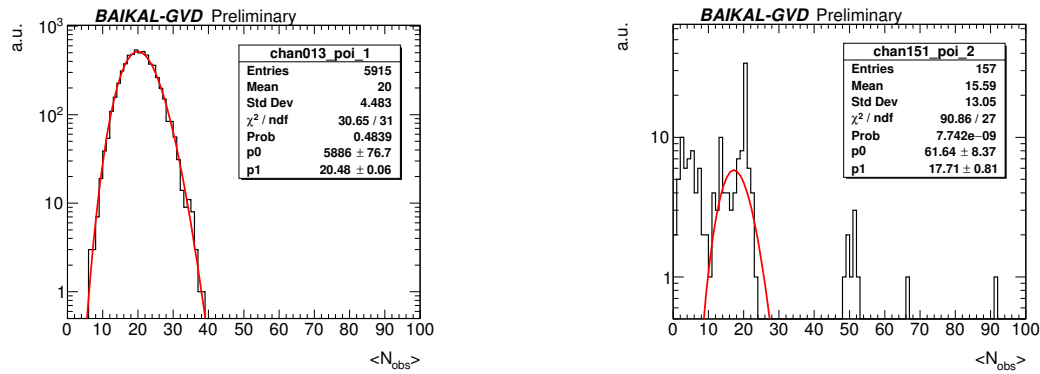
count rate quickly changes, we describe uniform distribution by a set of harmonic functions (see right panel in Figure 2).

### 2.3 Poissonian distribution test

The expected number of the recorded events should follow the Poissonian distribution for any fixed time interval (Figure 3). In order to decrease an impact of the unstable environment, we divide the whole run to parts with stable count rate. For each of run parts the fixed time interval is chosen to have approximately 20 recorded events on average. As one can see in right panel of Figure 3, there is clear effect of the calibration run with the laser source switched on, when the distribution differs from poissonian shape, and the bin content in statistically suppressed regions of the distribution is high due to laser events, as well as the mean value of distribution differs from expected number of recorded events.

## 3. Charge distributions test

The second group of parameters which monitors the charge characteristics of signals from every PMTs is referred to charge distributions. The quality of the charge related data is estimated only for channel level. The trigger condition of the detector is fulfilled if neighbor pair of channels produces signals exceeding low charge threshold ( $\sim 1.5$  p.e.) and high one ( $\sim 4$  p.e.). In order to

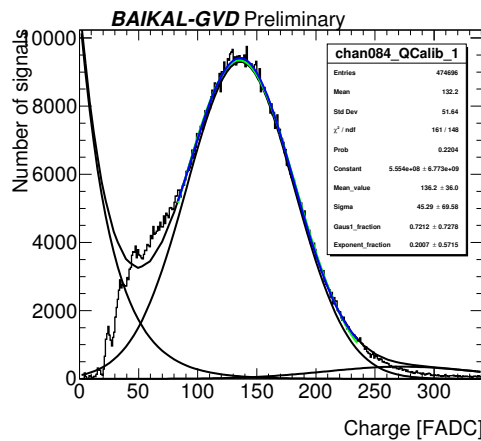


**Figure 3:** Poissonian distributions for some selected channels: standard run (left), calibration run with laser source (right).

analyze trigger signals separately from non-trigger signals, a special algorithm is applied to raw data waveforms which reconstructs the trigger implementation for each individual event.

### 3.1 Non-trigger signals

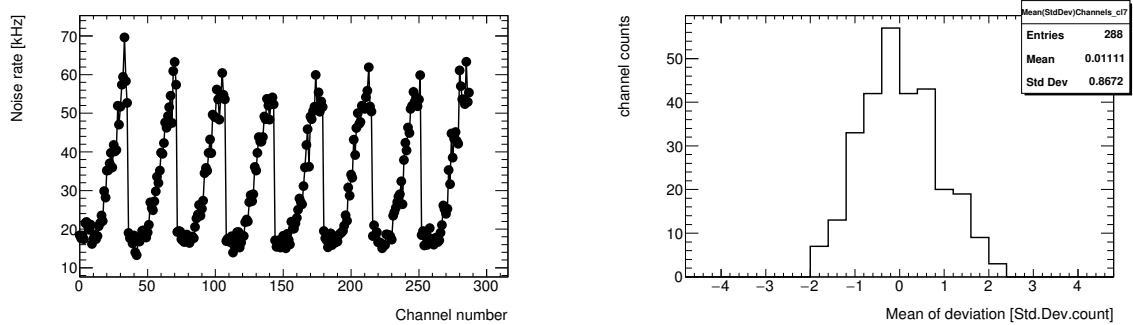
Any PMT signal exceeding the so-called "filter threshold" ( $\sim 0.5$  p.e.) is recorded. Using information from channels that did not fire the trigger for a given event we obtain the combined charge distribution of non-trigger signals (Figure 4). Combined distribution consists of dominated single p.e. gaussian distribution and unessential contributions from dark currents of PMT (to the left of 1 p.e. peak), and 2 p.e. distribution (to the right). After fitting the combined distribution by a sum of one exponential function and two gaussian functions we can determine the position of the single p.e. peak for given channel in the current run, produced mainly by random noise. This value is used as calibration factor for recorded charge from FADC counts to p.e. values.



**Figure 4:** The charge distribution of non-trigger signals for a some selected channel.

Non-trigger signals also give the opportunity to estimate the noise count rate. Figure 5 shows background noise characteristics of the collected data. It is expected that noise rate values should increase towards the higher vertical position of the channel on the string due to chemi luminescence

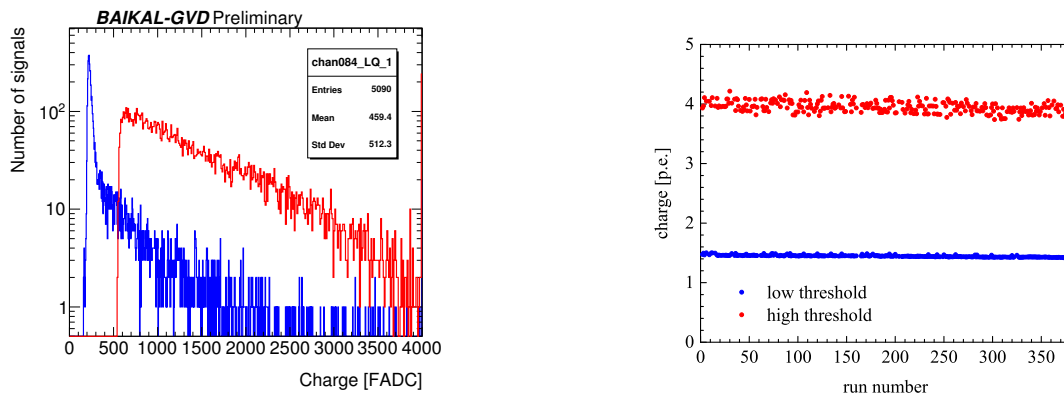
(left panel) [4]. We found that at every depth the channel noise rate is quite stable and practically do not depend on individual PMT, as it can be seen from the right panel of Figure 5, where the distribution of the mean noise rate deviations of each channel from depth-averaged rate obtained within over a half of season is shown.



**Figure 5:** The noise rates for channels in some selected run (left). Channels noise deviation from the depth-averaged value (expressed in standard deviation counts) during over a half of season (right).

### 3.2 Trigger signals

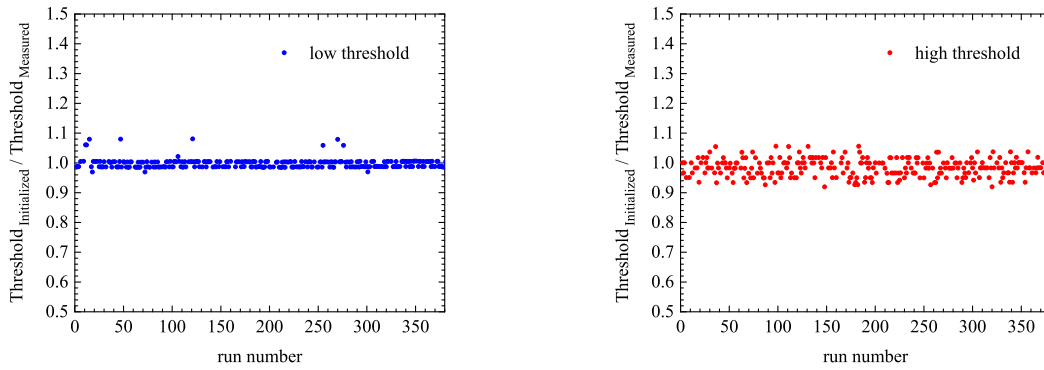
By using data on trigger signals, the DQM system performs monitoring of the measurements of the trigger charge thresholds. Low and high threshold values are set (or initialized) in the telescope configuration and are equal to  $\sim 1.5$  p.e. and  $\sim 4$  p.e., respectively. Initialized thresholds are set in 4 consecutive FADC counts, and it leads to an increase of the thresholds. We calculate the correction factor (that is equal to  $\sim 1.25$ ) and then obtain a precise initialized threshold values in photoelectrons. Figure 6 (left panel) shows the trigger thresholds that are measured in every run. The stability of these values is permanently monitored by the system for all channels during the run as well as the season.



**Figure 6:** The signal charge values for some selected channel participating in the trigger (left). Low and high thresholds within over a half of year for some channel (right). Blue color stands for low threshold, and red color stands for high threshold.

We found that the low and high thresholds expressed in terms of p.e. are quite stable for channels within season as it is shown in Figure 6 (right), where threshold measurements over half of

a year for some selected channel are plotted. Also, DQM system controls a deviation of measured thresholds from initialized values. Figure 7 shows the ratios of initialized to measured thresholds for some selected channel, which were collected during half of a year. We found a good agreement between initialized and measured thresholds within large data sample. As expected, these values are close to 1 and maximal deviations are smaller than 10%.



**Figure 7:** Ratios of the initialized to measured trigger thresholds collected within over half of a year for some selected channel. Left panel stands for low threshold, and right one stands for high threshold.

#### 4. Quality estimation algorithm

The purpose of the quality estimation algorithm is to obtain the quality status of the recorded data for each telescope unit. Channel, section and cluster levels are considered independently. Fit quality is estimated via  $\chi^2/NDF$  parameter with threshold values for good ( $<2$ ), normal ( $<4$ ), and bad ( $>4$ ) data. Fit quality estimations for each distribution are summed up as logical *and*. There are eleven ranking markers for data quality status (codes):

- 0 – excluded by configuration (defined by detector performance experts)
- 1 – empty data ( $<100$  events in channel)
- 2 – good data
- 3 – normal data
- 4 – bad data
- 5 – good data but periodic light source is detected
- 6 – normal data but periodic light source is detected
- 7 – bad data but periodic light source is detected
- 8 – good data but unstable environmental conditions are detected
- 9 – normal data but unstable environmental conditions are detected
- 10 – bad data but unstable environmental conditions are detected

## 5. Graphical user interface

The DQM system is integrated within our unified software framework "BARS" and operates automatically. The results obtained by system are placed into the central database, and are showed on the dashboard web service. Also, the DQM system has a graphical user interface, that allows use the DQM outputs in off-line mode (see Figure 8). There is opportunity to check quality estimations for selected parameters (for example for exponential distribution, or for exponential distribution with poissonian one and so on), as well as use combined result obtained by using all distributions. An explicit view of the distributions with fit function can be shown along with graphs which represent the deviation of distribution from the obtained fit-function for all bins. Also, by hovering the mouse one can see short statistic box with fit quality results. Any level of the telescope can be seen independently and supporting information like timestamps of start and stop run, total number of events in run, and so on is presented.

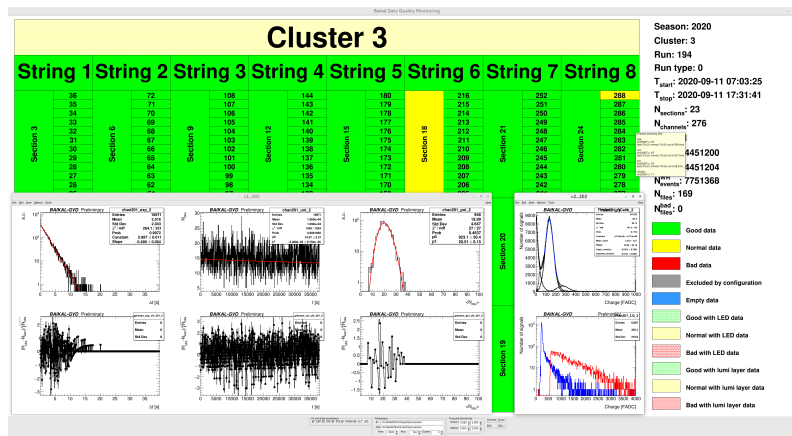


Figure 8: Graphical user interface of the DQM system.

## 6. Summary

We present the system which using to provide the quality estimations of recorded data of the Baikal-GVD. Data quality monitoring system operates automatically within Baikal-GVD's unified software framework "BARS" and it's graphical user interface allows to analyze the data quality in an off-line mode. DQM system takes into account a possible quick change of the telescope's background environment. By means of some examples of the data recorded during the 2020 season it was shown that applying various parameters for data quality analysis allows to estimate the quality of obtained data efficiently.

## 7. Acknowledgements

The work was partially supported by RFBR grant 20-02-00400. The CTU group acknowledges the support by European Regional Development Fund-Project No. CZ.02.1.01/0.0/0.0/16\_019/0000766. We also acknowledge the technical support of JINR staff for the computing facilities (JINR cloud).

**References**

- [1] A.V. Avrorin et al., Baikal-GVD coll., Baikal-GVD Experiment. *Phys. Atom. Nucl.* **83**, 916 (2020)
- [2] I.A. Belolaptikov et al., Baikal-GVD coll., Neutrino Telescope in Lake Baikal: Present and Nearest Future. *These proceedings*
- [3] B.A. Shaybonov et al., Baikal-GVD coll., Automatic data processing for Baikal-GVD neutrino observatory. *These proceedings*
- [4] R. Dvornický et al., Baikal-GVD coll., The Baikal-GVD neutrino telescope as an instrument for studying Baikal water luminescence. *These proceedings*