

# 1 Data Quality Monitoring system in the Baikal-GVD 2 experiment

---

A.D. Avrorin<sup>a</sup>, A.V. Avrorin<sup>a</sup>, V.M. Aynutdinov<sup>a</sup>, R. Bannash<sup>g</sup>, I.A. Belolaptikov<sup>b</sup>, V.B. Brudanin<sup>b</sup>, N.M. Budnev<sup>c</sup>, G.V. Domogatsky<sup>a</sup>, A.A. Doroshenko<sup>a</sup>, R. Dvornický<sup>b,h,\*</sup>, A.N. Dyachok<sup>c</sup>, Zh.-A.M. Dzhilkibaev<sup>a</sup>, L. Fajt<sup>b,h,i</sup>, S.V. Fialkovsky<sup>e</sup>, A.R. Gafarov<sup>c</sup>, K.V. Golubkov<sup>a</sup>, N.S. Gorshkov<sup>b</sup>, T.I. Gress<sup>c</sup>, R. Ivanov<sup>b</sup>, K.G. Kebkal<sup>g</sup>, O.G. Kebkal<sup>g</sup>, E.V. Khramov<sup>b</sup>, M.M. Kolbin<sup>b</sup>, K.V. Konischev<sup>b</sup>, A.V. Korobchenko<sup>b</sup>, A.P. Koshechkin<sup>a</sup>, A.V. Kozhin<sup>d</sup>, M.V. Kruglov<sup>b</sup>, M.K. Kryukov<sup>a</sup>, V.F. Kulepov<sup>e</sup>, M.B. Milenin<sup>e</sup>, R.A. Mirgazov<sup>c</sup>, V. Nazari<sup>b</sup>, **A.I. Panfilov<sup>a</sup>**, D.P. Petukhov<sup>a</sup>, E.N. Pliskovsky<sup>b</sup>, M.I. Rozanov<sup>f</sup>, E.V. Rjabov<sup>c</sup>, V.D. Rushay<sup>b</sup>, G.B. Safronov<sup>b</sup>, B.A. Shaybonov<sup>b</sup>, M.D. Shelepov<sup>a</sup>, F. Šimkovic<sup>b,h,i</sup>, A.V. Skurikhin<sup>d</sup>, A.G. Solovjev<sup>b</sup>, M.N. Sorokovikov<sup>b</sup>, I. Štekl<sup>i</sup>, O.V. Suvorova<sup>a</sup>, E.O. Sushenok<sup>b</sup>, V.A. Tabolenko<sup>c</sup>, B.A. Tarashansky<sup>c</sup>, S.A. Yakovlev<sup>g</sup>

<sup>a</sup> Institute for Nuclear Research, Russian Academy of Sciences, Moscow, 117312 Russia

<sup>b</sup> Joint Institute for Nuclear Research, Dubna, 141980 Russia

<sup>c</sup> Irkutsk State University, Irkutsk, 664003 Russia

<sup>d</sup> Institute of Nuclear Physics, Moscow State University, Moscow, 119991 Russia

<sup>e</sup> Nizhni Novgorod State Technical University, Nizhni Novgorod, 603950 Russia

<sup>f</sup> St. Petersburg State Marine Technical University, St. Petersburg, 190008 Russia

<sup>g</sup> EvoLogics GmbH, Germany

<sup>h</sup> Comenius University, Mlynska Dolina F1, Bratislava, 842 48 Slovakia

<sup>i</sup> Czech Technical University in Prague, Prague, 128 00 Czech Republic

E-mail: dvornicky@dnf.fmph.uniba.sk

The quality of the incoming experimental data has a significant importance for both analysis and running the experiment. The main point of the Baikal-GVD DQM system is to monitor the status of the detector and obtained data on the run-by-run based analysis. It should be fast enough to be able to provide analysis results to detector shifter and for participation in the global multi-messaging system.

36th International Cosmic Ray Conference -ICRC2019-  
July 24th - August 1st, 2019  
Madison, WI, U.S.A.

---

\*Speaker.

## 3 1. Introduction

4 There are several requirements to the monitoring system for the Baikal-GVD experiment [1].  
5 It should be as simple as possible to be fast in running and clear enough for use after reading the  
6 manual. But at the same time the system should be reliable and provide monitoring of all crucial  
7 parameters of the detector. In general, parameters to be monitored could be divided into two main  
8 groups. The first one monitors the statistical parameters of the detector to register muons. Since the  
9 trigger of the detector is extremely basic and firing if any neighbor pair of channels (or OM - optical  
10 module) generates signals exceeding at least high ( $\sim 5$  p.e.) and low ( $\sim 2$  p.e.) charge thresholds,  
11 the main source of events passing this trigger are atmospheric muons and random noise [2]. The  
12 second set of parameters is charge related. It analyses charge distributions of both trigger and  
13 non-trigger signals recorded by a given channel. The minimum  $\chi^2$  fit algorithm is applied while  
14 performing the fit of the distributions to a corresponding expected function. In addition to the  
15 parameters mentioned above the trigger efficiency of the cluster is also monitored.

## 16 2. Statistical parameters

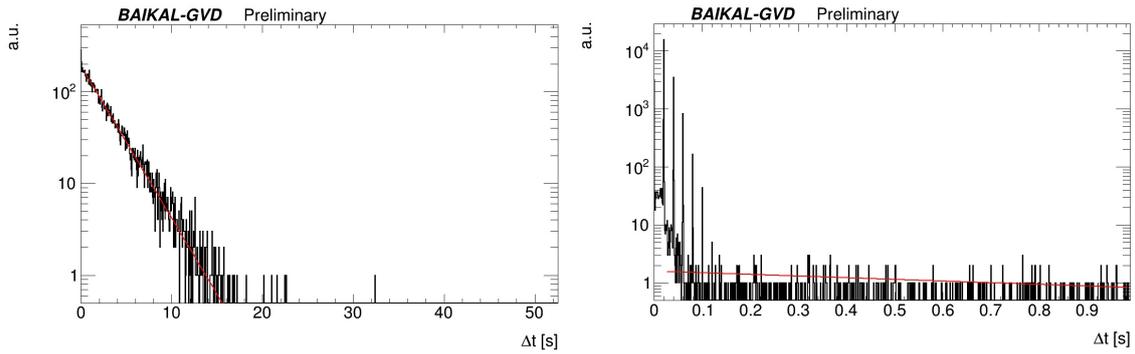
17 There are three parameters monitored within this set: exponential, uniformity and Poisson.  
18 For each of them the monitoring system analyses shapes of distributions for all events in the run  
19 under consideration. The shape analysis includes the fit of the distribution with the expected func-  
20 tion following by the estimation of the fit quality. These parameters are analysed for each channel,  
21 section, and the whole cluster separately [2]. Besides the fit quality estimation using the  $\chi^2/NDF$   
22 parameter, the amount of bins that deviate from the obtained fit-function is calculated. Only sta-  
23 tistical uncertainty is considered to estimate the number of standard deviations for the given bin of  
24 the distribution.

### 25 2.1 Exponential distribution

26 Since the main type of events passing the trigger are atmospheric muons and random noise the  
27 time difference between two neighbor events should be described by the exponential function (see  
28 Fig. 1 left). The bin width is chosen to suppress possible bin-to-bin statistical fluctuations. Due to  
29 the dead time of the trigger operation the maximum of the distribution is at  $\sim 1$  ms and events with  
30 a time difference lower than 1 ms corresponds to statistics recorded by the second event buffer of  
31 the trigger system. Therefore the next bin to the maximum one is chosen as the low edge of the fit  
32 range and till the last bin. The calibration systems used for PMT calibration such as LED matrices  
33 or laser sources due to the fixed operation frequency values should deteriorate the exponential  
34 shape of the atmospheric muons and random noise (see Fig. 1 right) and therefore decrease the fit  
35 quality and increase the number of bins with very high deviation value from the obtained with the  
36 fit-function.

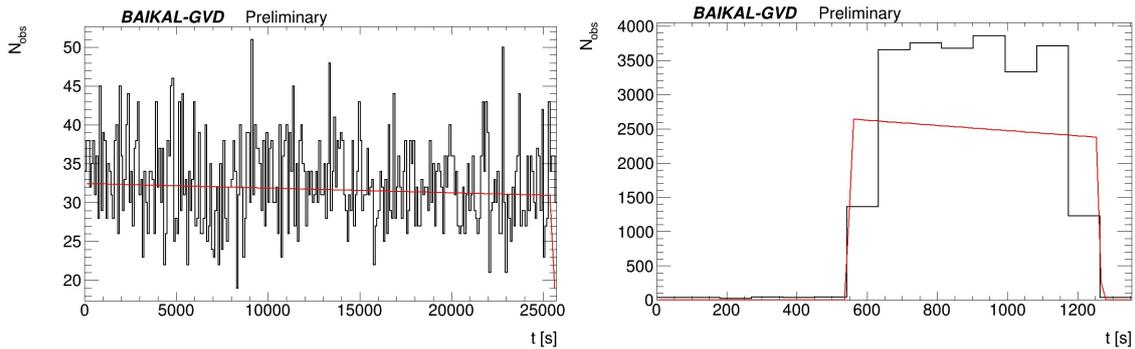
### 37 2.2 Uniformity distribution

38 It is expected that the distribution of the rate of atmospheric muons and random noise is linear  
39 with possible slope due to the environment (temperature, season, water transparency etc) or detector  
40 effects (current fluctuations etc.) and should be described by the first order polynomial function.



**Figure 1:** Exponential distributions for atmospheric muons and random noise for some selected channels. Right: for calibration run with LED matrix.

41 The expected cluster rate is  $\sim 50$ -200 events/sec. As well as for the exponential distribution the  
 42 bin width for the corresponding uniformity distribution is chosen to suppress possible bin-to-bin  
 43 statistical fluctuations (see Fig. 2 left). As it is seen in Fig. 2 (right) events with LED calibration  
 44 matrices applied show a significant increase in rate with very clear boundaries thus the fit quality  
 45 decreases.



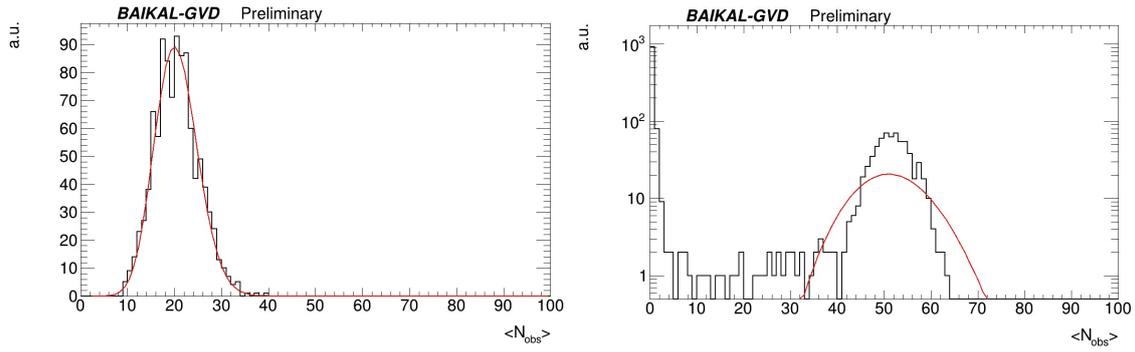
**Figure 2:** Uniformity distributions for atmospheric muons and random noise for some selected channels. Right: the same for calibration run with LED matrix.

### 46 2.3 Poisson distribution

47 The expected value of the event rate for the given cluster size ( $\sim 50$ -200 events/sec per cluster)  
 48 should follow Poissonian distribution for any fixed time interval. Thus, for each run the time  
 49 interval is chosen to have  $\sim 20$  recorded events on average and the distribution of the number of  
 50 events per this time range is tested for the poisson statistics (see Fig. 3 left). There is clear effect of  
 51 the calibration run with the LED matrix when the average number of events per given time interval  
 52 differs from poissonian shape and the mean value is lower than expected number ( $\sim 20$ ) of recorded  
 53 events (see Fig. 3 right).

### 54 3. Charge distribution

55 Another important set of parameters for the run quality monitoring is referred to the distribu-

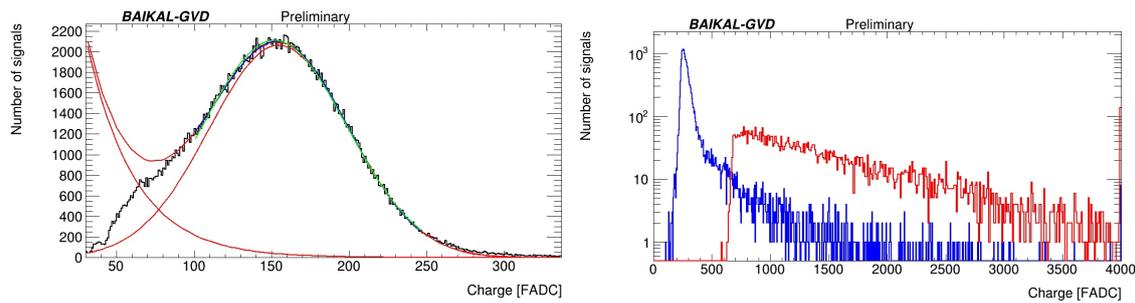


**Figure 3:** Rate distributions for atmospheric muons and random noise for some selected channels fitted with a Poisson function. Right: the same for calibration run with LED matrix.

56 tions of the charge from the PMT. Therefore the data quality is estimated for given channel only  
 57 and considered in summarizing algorithm. The low threshold for the charge is the so-called filter  
 58 threshold that is about half of the single photoelectron value. It is possible to analyse separately  
 59 channels that produced trigger in the given event from all other ones as well as full combination of  
 60 signals from channel.

### 61 3.1 Distribution for non-trigger signals

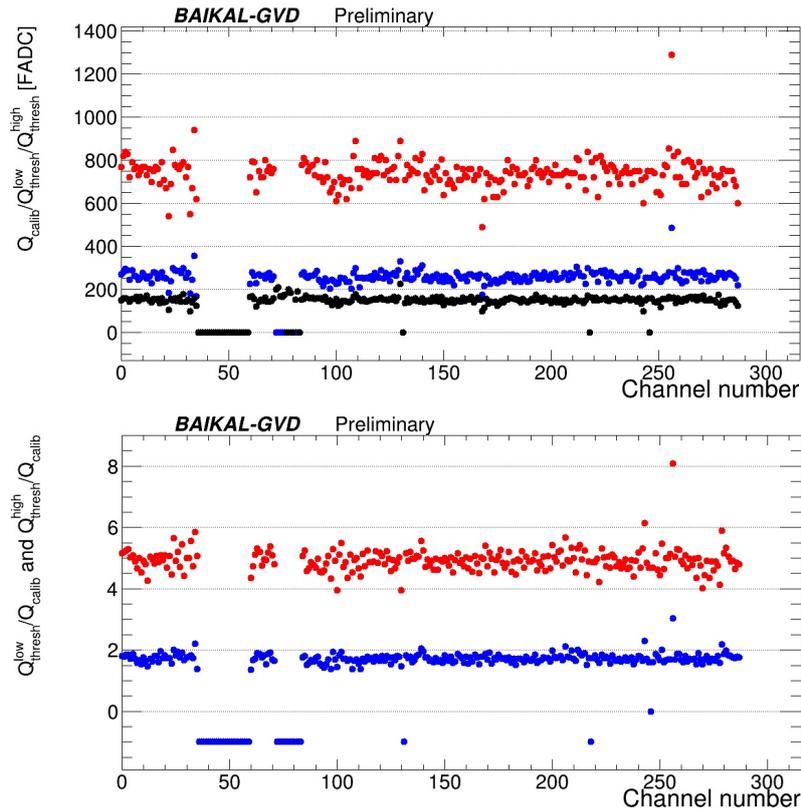
62 Any charge from the PMT exceeding the so-called "filter threshold" ( $\sim 0.5$  p.e.) is recorded.  
 63 Selecting only channels that did not fire the trigger for a given event we can determine the position  
 64 of the 1 p.e. peak for given channel in the current run produced mainly by chemi luminescence. As  
 65 one can see in Fig. 4 (left) along with very clear 1 p.e. gaussian distribution with the mean value at  
 66 around 160 FADC, there is some effect from dark currents of the PMT on the left side of the 1 p.e.  
 67 distribution. Also there is negligible impact from the 2 p.e. distribution that is seen on the right  
 68 side of the 1 p.e. one. To describe the combined distribution the sum of two gaussian functions for  
 69 the 1 and 2 p.e. distributions and an exponential function for the dark current part was used.



**Figure 4:** Left: the charge distribution for a some selected non-trigger channel; Right: charge values for some selected channel participating in the trigger.

70 Combined information for all channels in run is used for control of PMTs work stability. As  
 71 one can see in Fig. 5 (top) the 1 p.e. calibration values (black points) are rather stable over all  
 72 channels in the run. The spread of the values is due to the manual adjustment of the mean value for  
 73 each channel and the difference of the PMTs themselves. In addition to calibration behavior it is

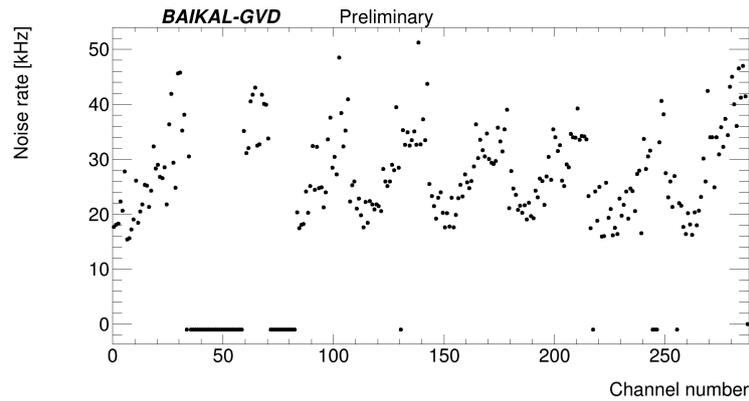
74 possible estimate the stability of the noise rate of the channel. Statistics recorded by given channel  
 75 is proportional to the noise rate of this channel. This noise rate is calculated as the distribution  
 76 integral from 0.5 p.e. for each channel (see Fig. 6). It is expected that this values should increase  
 77 towards the higher vertical position of the channel on the string [3]. The noise rate for channel is  
 78 closely correlated to the calibration value and thus it is expected that at every depth should be stable  
 79 from channel to channel. The deviation of noise related value for each channel from the average  
 80 over the given depth level is the parameter for the channel quality estimation.



**Figure 5:** Top: charge calibration (black), trigger low (blue) and high (red) threshold values for channels in some selected run; Bottom: Ratios of trigger low (blue) and high (red) threshold to charge calibration values for channels in some selected run.

### 81 3.2 Distribution for trigger signals

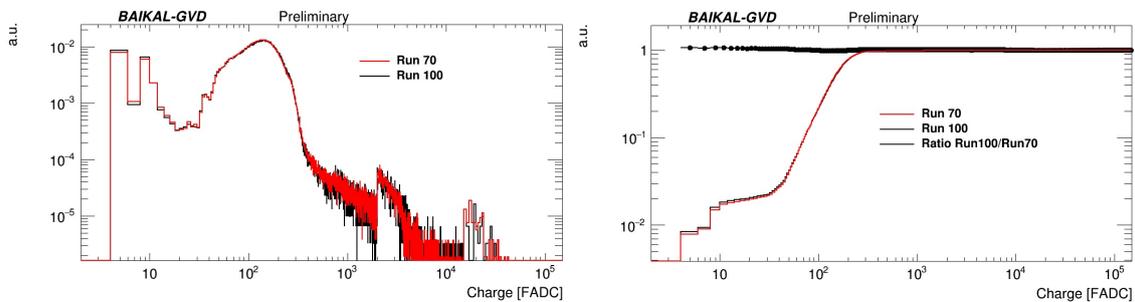
82 The values of low and high trigger threshold are set to  $\sim 2$  and  $\sim 5$  photoelectrons. It is very  
 83 important to make sure that these values are stable across the run and season. In Fig. 4 (right) it is  
 84 seen that the shapes of the corresponding distributions have thresholds that are extracted and used  
 85 for monitoring (see Fig. 5 (top) low (blues) and high (red) thresholds). Since the exact value of the  
 86 threshold in p.e. depends on the charge calibration of the PMT, the ratio of the threshold value in  
 87 the given channel to its charge calibration mean value are monitored. It is expected that these ratios  
 88 are stable across all channels within both run and season (see Fig. 5 bottom) and some deviations  
 89 are possible due to the same reasons as for the calibrations itself.



**Figure 6:** The noise rates for channels in some selected run.

### 90 3.3 Shape comparison for all signals

91 In addition to calibration and threshold values the general comparison of the charge distribu-  
 92 tion shapes are performed. Shapes of distributions of any charge deposited in the channel from the  
 93 filter threshold up to 1000 p.e. are compared. Distribution from the run under analysis is compared  
 94 to a so-called basic run that is well known to have good channel performance. The shapes are com-  
 95 pared via cumulative distributions normalized to 1. The relative impact of the integral in a given  
 96 range starting from 0 is the estimation parameter. Fig. 7 shows the example of charge distributions  
 97 (left) and cumulative distributions obtained from them (right) for basic (red) and considered (blue)  
 98 runs correspondingly. Number of events with charge value deposited in channel more than 100 p.e.  
 99 should not exceed 100 events, otherwise the channel is masked as flashed by LED matrix.

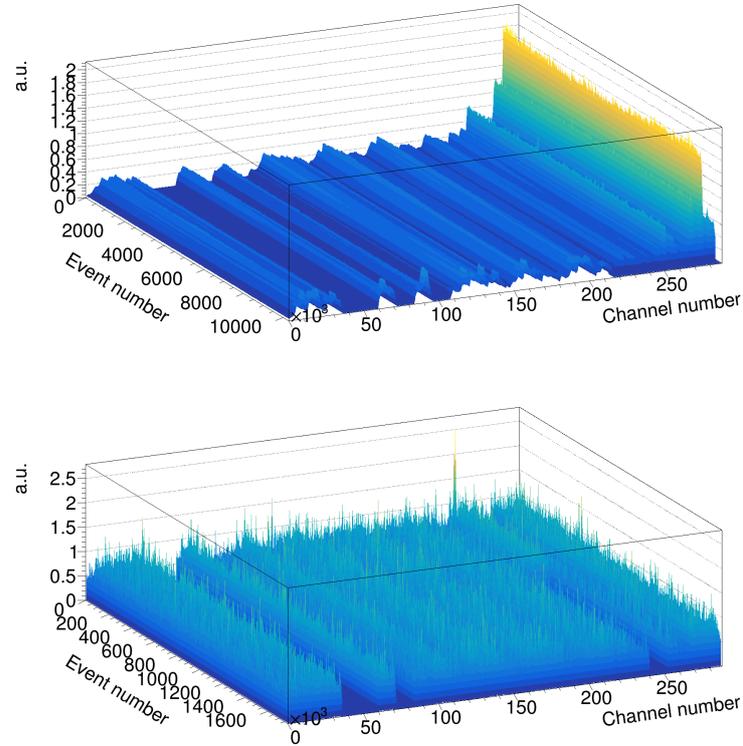


**Figure 7:** Charge distributions (left) and cumulative distributions obtained from them (right) for basic (red) and considered (blue) runs correspondingly. Ratio of charge distributions is shown in black on right panel.

## 100 4. Gantt-plots

101 To analyse as much events as possible with the given cluster configuration the Gantt-plots are  
 102 used for two main purposes. The first one is to avoid events with the LED matrix switched on. In  
 103 this case the number of events recorded by cluster increases by several orders of magnitude thus to  
 104 save the time processing the number events range tested for this issue is 10 000. The Gantt-plot  
 105 normalized to number of events is seen in Fig. 8 (top) that channels with LED matrix switched on

106 exceed 0.5 number of entries per bin. Thus in standard conditions we expect maximum a doubling  
 107 of the rate during the run.



**Figure 8:** Top: the Gantt-plot used for selection of events with the LED matrix switched on in each channel ; Bottom: the Gantt-plot used for selection events with stable event rate in each channel.

108 For the second purpose the other Gantt-plot was used. This plot contains 100 thousand bins  
 109 for the axis corresponding to event number and is also normalized to the total number of events  
 110 in the run. Such high granularity allows to select events range with the channel working properly.  
 111 Assuming that each channel should not have more than one signal higher than zero the upper limit  
 112 of the bin content is set to 1 (see Fig. 8 bottom).

## 113 5. Quality estimation algorithm

114 After estimation of the run using parameters described above the solution algorithm is applied.  
 115 The main point of this algorithm is to provide for the analysis procedure information about quality  
 116 status at each level of the detector from channel to the whole cluster. There are eight ranking  
 117 markers for data quality status (codes):

118 **0** – excluded by configuration (defined by detector performance experts)

119 **1** – empty data (<100 events in channel)

120 **2** – good (successfully passed criteria)

121 **3** – normal (passed criteria with some warnings)

122 **4** – bad (failed criteria)

123 **5** – good data but LED is detected

124 **6** – normal data but LED is detected

125 **7** – bad data but LED is detected

126 The decision chain is the following: channel → section → string → cluster levels. At any level  
127 (channel, section and cluster) for exponent, uniformity, Poisson and Charge (channel level only):

128 1. Fit quality is estimated via  $\chi^2/NDF$  with threshold values for good (<2), normal (<4) and  
129 bad (>4) data

130 2. Deviation of each bin value from fit-function using following condition with threshold values  
131  $R_1=25\%$  and  $R_2=1\%$  for good data and  $R_1=50\%$  and  $R_2=5\%$  for normal data correspond-  
132 ingly (any other case → bad data):  $N_{Normal}^{Bins}/N_{Total}^{Bins} < R_1$  and  $N_{Bad}^{Bins}/N_{Total}^{Bins} < R_2$

133 3. The deviation of the noise rate for each channel from the average over the given depth level:  
134  $<3\sigma_{noise}^{level}$  for good data,  $<5\sigma_{noise}^{level}$  for normal data and bad ( $>5\sigma_{noise}^{level}$ ) data

135 4. The relative impact of the integral in the [0-x] range of the distribution for non-trigger signals:  
136  $<15\%$  for good data,  $<25\%$  for normal data and  $>25\%$  for bad data data

## 137 6. Conclusions

138 The data quality monitoring system for the Baikal-GVD experiment was developed. Using  
139 some examples of the data recorded during the 2016 season it was shown that using various param-  
140 eters for data quality analysis allows to estimate the quality of obtained data from the level of the  
141 optical module and up to the whole cluster very efficiently.

## 142 7. Acknowledgements

143 This work was supported by the Russian Foundation for Basic Research (Grants 16-29-13032,  
144 17-0201237).

## 145 References

- 146 [1] A.D. Avrorin et al, *Baikal-GVD: status and prospects*, in proceedings of  $XX^{th}$  *International Seminar*  
147 *on High Energy Physics (QUARKS-2018)*, EPJ Web Conf. Volume 191, 2018.
- 148 [2] A.D. Avrorin et al, *Data acquisition system of the NT1000 Baikal neutrino telescope*, *Instruments and*  
149 *Experimental Techniques* **3** (2014) 262–273.
- 150 [3] A.D. Avrorin, R. Dvornický et al, *Luminescence of water in Lake Baikal observed with the*  
151 *Baikal-GVD neutrino telescope*, in proceedings *Very Large Volume Neutrino Telescopes*  
152 *(VLVnT-2018)*, EPJ Web Conf. Volume 207, 2019.