Primary Processing of Cosmic Ray Station Data: Algorithm, Computer Program and Realization.

A.V.Belov, Ya.L.Blokh, E.G.Klepach, V.G.Yanke

Institute of Terrestrial Magnetism, Ionosphere and Radiowave Propagation (IZMIRAN), 142092, Troitsk, Moscow region, Russia

ABSTRACT

In this work we analyze a quality of ground based cosmic ray detector performance (neutron monitors, muon telescopes). Different types of instrument variations, their possible causes and selection methods are considered. For comparison of similar channels we propose a method of the ratio logarithms that has clear advantages in comparison with the ordinary method of ratios. The algorithm of primary processing of multi-cannel detector data is discussed in details. This algorithm uses a notion of efficiency for particular channel and the entire detector as well. A reference is given for electron publication of the editor-program. The program allows determining efficiency for each channel, selecting "out of order" channels, and editing data of these channels. Besides, statistical errors of obtained characteristics for each particular channel and the whole detector are estimated. The procedure of data control and edition are especially important for data publication in real time, when the requirements for data quality are extremely high.

1. Purposes and Characteristic Features of Primary Data Processing

Data of cosmic ray ground based observations are processing to get, as a final result, geophysical and astrophysical information, firstly, information on cosmic ray distribution in the near Earth space. Data of many stations and rather complex models of primary cosmic ray interactions with the Earth magnetosphere and atmosphere are used. For successive results of this processing (high level processing) data of particular stations should be preliminary verified and prepared, i.e. should be processed preliminary.

So, the main purpose of preliminary processing of cosmic ray data is a preparation of full value experimental material for the high level processing. The primary processing itself does not provide any physical information except that on the instrument performance. However, the quality of primary processing determines an amount and quality of information on the interplanetary space, the Earth magnetosphere and atmosphere obtained from cosmic rays.

The preliminary processing of ground based cosmic ray observations is arbitrary performed at the station, where the data are obtained. Data of each station is processed independently and a large amount of auxiliary information is used, but it is not necessary to store this information after the primary processing.

Besides, the primary processing and preparing of experimental data in real time is very important now. More and more experimental data are presented on the internet in real time.

The primary processing includes:

1) Data verification, checking of their reliability, estimates of instrument performance quality;
2) Searching and removing of instrument variations;
3) Introducing correction coefficients for meteorological effects;
4) Introducing corrections for random coincidences;
5) Estimate of statistical errors of data processed preliminary;
6) Data preparing in the standard form for their exchange, publication, storage and further processing.

Some problems of the primary processing allow as hardware well as software solution. Here we will discuss only software solutions. The main attention will be devoted for searching and removing of instrument variations, tests of experimental data by control of detector efficiency. The term of detector (or registration channel) "efficiency" is intuitively clear and often appeared. We will use this term as well, but its definition will be given later.

2. Methods of Data Quality Control in Ground Based Cosmic Ray Observations

The main detector of cosmic radiation was the ionization chamber (Compton,1934;Shafer,1958) during the first stage of continuous observations. The physical principle of the detector itself helps maintaining constant its efficiency. The chamber current caused by the cosmic radiation was compared with the current initiated by the internal radioactive source. A difference between these two currents was measured permanently and a summary of the currents was measured for control after equal time intervals. Besides, the detector was calibrated periodically by external source of ionizing radiation.

After the IGY neutron monitors (Simpson,1955) and later the NM64 super-monitors (Carmichael, 1962) appeared, the problem of their efficiency control was attempted to be solved in the same manner, the detector was calibrated periodically by the source of neutrons. However, in general this method was not satisfied. It demanded storing of radioactive sources at stations, wasting of additional operation time. The calibration disturbed continuous data registration, did not allow controlling continuously the detector efficiency. A composition, energy spectrum, angular distribution of particles from the calibration detector did not correspond to the similar characteristics of the cosmic radiation. The reasons mentioned above and some other disadvantages of the method have resulted that it is not used at present at all.

Methods of data quality control were developed in another manner. Two independent and practically identical sections constituted the IGY monitor. This allowed continuous data registration and control of their quality by comparing count rates of the sections with each other (Janossy,1965). So, by this way it was possible only justify a normal work of the detector (variations registered by two sections coincided), but it was impossible without data of other detectors to distinguish, which section was well performing and which not.

The NM64 neutron monitor of three sections allowed much better control of data quality. Comparing data of different identical sections it is possible to select one (not more) bad section. In a case of four section detector even malfunction of two sections is not dangerous.

Mentioned, four identical sections form the standard vertical muon telescope, so methods for selection of instrument variations elaborated for neutron monitors are suitable for the vertical muon telescope with small corrections as well. The data quality control of inclined muon telescopes needs other approaches.

Therefore, modern methods for internal control of ground based cosmic ray detectors and their data quality are based on dividing of the detector into several (≥ 3) practically identical sections (the section method) and comparing of their data with each other. This method has several undoubted advantages: data are registered and controlled continuously, it can be easy automated. The algorithms of primary data processing described in are different variants of the section method. The method efficiency increases for increasing number of sections. Providing a channel for data of each neutron counter is, apparently, the best realization of the method. An amount of information should be rejected decreases by several times and, correspondingly, decreases the statistical error of detector data. Another important obstacle is that a probability to have instrument variations with one sign and close values simultaneously in two sections is rather high. In this case, the section in order would be identified as out of order and the detector efficiency would not be calculated correctly. Clear, the probability of similar instrument variations in, for instance, ten channels from eighteen is statistically negligible. Discussing problems of multi section registration of the neutron component, it should be mentioned that counters of the neutron monitor are not identical and their count rates are not independent. These differences and dependence could be selected and accounted by controlling performance of the multi channel detector.

With increasing number of information channels considerably increases an amount of information for primary processing, but it is not so important for modern computers. The modern computers provide new abilities for the instrument control and open new ways for problems and methods of data processing. To our opinion, an amount of auxiliary information will increase in future. Possibly, coordinates and time for each registration of particle in the detector, an amplitude and shape of corresponding pulses would be used.

We know only one disadvantage of the multi-section method, it can not select efficiency changes of the detector as a whole, i.e. the changes occurred simultaneously in all channels. These changes are not always of instrument origin, they might be associated with some changes of the detector environment (for instance, a change of snow cover or building reconstruction). They make worse the data quality as well as the instrument variations. Therefore, along with methods of the internal control methods based on comparing of data from different detectors should be developed.

3. Types of Instrument Variations

All variations of instrument origin can be divided into three main groups: peaks?,jumps? and drifts?. Figure 1 presents these types of variations.

Possible causes of the peak are: malfunction of high voltage supply, short power switch-off; interference, malfunction of the register, bug of data storage on intermediate carrier and etc.

The peaks can be solitary or they may follow each other during a short time. The peak is a spoilt information, which should be removed.

The jump is arbitrary observed if some element of the information channel has been replaced or its working regime has been changed dramatically. As a rule, it is possible to remove the jump and save the information by editing the data.

An origin of the drift is arbitrary not clear. Its possible causes are: incoordination, growing older or lack of temperature stability of some element of the information channel. Sometimes there are several causes of the drift and distinguish them is difficult. Periodic and quasi-periodic instrument variations, for example, season and daily variations can be considered as a special case of the drift. In principle, data with instrumental drifts can be edited without losing of their fruitfulness, but this is a more difficult problem than in a case of jumps.

All types of instrument variations can be treated as efficiency changes of particular channel or the entire detector. We can give following short definitions. The peak is a short and considerable change of the efficiency with quickly recovery to its initial value. The jump is a sharp change of the efficiency for a rather long period. The drift is a gradual and rather slow change of the efficiency.

The proposed definitions as well as dividing of instrument variations into three types are conventional, but they are useful in practice.

The instrument variations were discussed above for count rates averaged (or accumulated) for equal time intervals. It is common primary processing hourly data. Measuring average count rates we deal with realizations of some definite distribution (the poison distribution in the simplest case). We have discussed only yet how the instrument variations effect on the expectation of a given distribution. However, the instrumental factors may effect on the entire distribution. They may increase or decrease its dispersion or effect on a distribution shape in some other way. We can speak about instrument variations of systematic and random origin and instrument variations of count rate fluctuations. Instrument variations of dispersion σ_{device} are mostly important. We need know them for correct estimates of statistical errors of detector data. Besides, it is apparently impossible to determine precisely instrument variations and edit data correctly data without simultaneous determination of their dispersion.

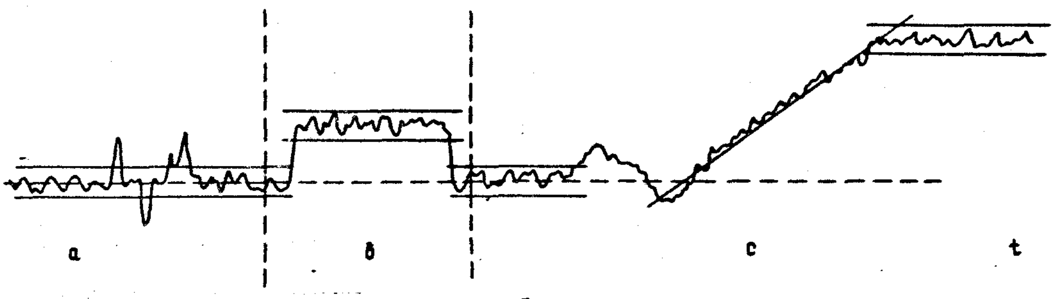


Fig. 1 Classification of instrument variations: a - peak, b - jump, c - drift.

As for average count rates one can speak about instrumental peaks, jumps and drifts of dispersion and consider instrumental fluctuations as fluctuations of the detector efficiency. A definition of the detector efficiency is given below.

4. Detector Efficiency

Determining the detector efficiency we should account that a count rate of the detector depends not only on a number of incident particles, but also depends on their composition and energy distribution, as well as, that a particle registration in the detector is random process, so the efficiency would be a random function of many variables.

Let $S(t) = \Sigma p_i(t)$ - is a probability of registration of incident particles by the detector in a given time interval and $S(t_0)$ is a sum of registration probabilities existed in a previously defined time interval for identical particles. Then, the expectation of ratio

$$\varepsilon(t) = \frac{S(t)}{S(t_0)} = \frac{\Sigma p_i(t)}{\Sigma p_i(t_0)} \tag{4.1}$$

is the detector efficiency in a given time. The efficiency of particular channel can be defined similarly.

Another definition of the efficiency, which directly speaks about its usage, is possible. The detector efficiency $\varepsilon(t)$ - is a number on which the observed count rate N(t) should be divided to remove variations associated with changes of the detector itself, i.e.

$$\varepsilon(t) = 1 + \delta^{device}(t) \tag{4.2}$$

where $\delta^{device}(t)$ are instrument detector variations. Therefore, knowing the detector efficiency

$$N'(t) = N(t) / \varepsilon(t) \tag{4.3}$$

is enough for removing instrument variations (data editing). The efficiency is a relative value, so, besides its value in a given moment we need know for how long period the efficiency has not changed $\varepsilon(t_0) = 1$.

5. Methods for Comparison of Channel Efficiency

The purpose of primary data processing is a selection of real variations (i.e. natural variations of cosmic rays) and removing of falls (instrument) variations from the data. However, on the first stage of primary processing we should oppositely select pure instrument variations removing cosmic ray variations. For this purpose, comparing similar channels it is common to use a ratio of their count rates

$$r_{ij} = N_i / N_j \tag{5.1}$$

or their differences

$$d_{ij} = N_i - N_j \tag{5.2}$$

Let the variation is considered from the time moment t_0, when count rates and are known N_i^0 and N_j^0 ., then for a current moment we have

$$N_i = N_i^0(1+\delta)(1+\delta^{device}) \tag{5.3}$$

where $\delta = \delta(t)$ - is the cosmic ray variation expected to be common for all channels, and $\delta^{device} = \delta^{device}(t)$ - is the instrument variation, which is different for different channels.

It is more convenient to use in the ratio method not ratios themselves, but following values (ratios of ratios):

$$r_{ij} = \rho_{ij} / \rho_{ij}^0 = \frac{N_i}{N_j} / \frac{N_i^0}{N_j^0} \tag{5.4}$$

Substituting the expression (5.3) into (5.4) we get

$$r_{ij} = \frac{1 + \delta_i^{device}}{1 + \delta_j^{device}} = \frac{\varepsilon_i}{\varepsilon_j} \tag{5.5}$$

So, the method of ratios allows removing of cosmic ray variations, and the ratios of ratios are associated in a simple manner with the instrument variations and channel efficiency. The detector efficiency can be easy calculated by using values r_{ij} . and the connection $\varepsilon_i = r_{ij}\varepsilon_j$.

Disadvantages of the method of ratios are associated with non-linear transformation of count rate values for ratio calculations. Therefore, efficiency changes occurred in the i and j channels effects on values ρ_{ij} end r_{ij} differently. For 100% of efficiency decrease (i.e. channel switch-off) the value r_{ij} would decrease by one, the j- channel switch-off would lead to increasing r_{ij} up to ∞ . It is more important that the distribution of ratio fluctuations ρ_{ij} and r_{ij} is strongly asymmetric and different from the normal law. It is not so important, if instrument variations and random statistical fluctuations are small. However, for increasing of instrument variations or statistical fluctuations (for example, if small time intervals are used for averaging) such properties of the ratio distribution create serious problems for statistical data processing and make worse an accuracy of determination of instrument variations and efficiency. Using the ratio variations instead of the ratios, i.e. values $1 - r_{ij}$, does not differ considerably from the case discussed above, so we will not describe this here.

Differences in the form (5.2) can be used only for equivalent channels, i.e. , if $N_i^0 = N_j^0$. In practice normalized differences are used:

$$d_{ij} = c_i N_i - c_j N_j \tag{5.6}$$

where c_i и c_j - are normalized coefficients, which account differences between channels known apriori and satisfying a condition $c_i N_i^0 - c_j N_j^0 = 0$.

If the expression is valid (5.3), then accounting the last condition we have

$$d_{ij} = (c_i N_i^0 \delta_i^{device} - c_j N_j^0 \delta_j^{device})(1 + \delta) = 0 \tag{5.7}$$

The expression (5.7) will be simplified, if the normalized coefficients are obtained from conditions

$$c_i N_i^0 - c_j N_j^0 = 0 \quad (1 \le i \le k, 1 \le j \le k) \tag{5.8}$$

$$\sum_{i=1}^{k} c_i N_i^0 = k \tag{5.9}$$

then

$$c_i = \frac{1}{N_i^0} \quad \text{and} \quad d_{ij} = \frac{N_i}{N_i^0} - \frac{N_j}{N_j^0}$$

or

$$d_{ij} = (\delta_i^{device} - \delta_j^{device})(1 + \delta) = (\varepsilon_i - \varepsilon_j)(1 + \delta) \tag{5.10}$$

However, for any choice of normalized coefficients for differences d_{ij} (and the efficiency determined by using them) would depend on a value of cosmic ray variation and observed detector count rate. By this reason the method of differences can be recommended, if instrument variations and statistical fluctuations are much greater than cosmic ray variations, or, if cosmic ray variations can be removed from the data by some another way. If the above requirements are satisfied, then benefits of the difference method (equivalence of channels, symmetry of the distribution) in comparison with the ratio method would become, than its disadvantages.
The efficiency of particular channel can be controlled by calculating a ratio of a given channel count rate to a number of channels remained in order. The channel efficiency is:

$$\varepsilon_i = \frac{N_i}{N_i^0}\frac{\Sigma N_i^0}{\Sigma N_i} \quad (j \neq i, j \neq j_m) \tag{5.11}$$

where j_m - are numbers of channels in order. This allows removing of real variations, but the method is not convenient. Possibly, it is mostly important that the method uses the principle of multi-channels. For instance, it is not clear how estimate a statistical accuracy of channel efficiencies obtained by this method. Cosmic ray variations are not presented in values calculated by an expression:

We have

$$\eta_{ij} = \frac{\delta_i^{device} - \delta_j^{device}}{2 + \delta_i^{device} + \delta_j^{device}} = \frac{\varepsilon_i - \varepsilon_j}{\varepsilon_i + \varepsilon_j} \tag{5.12}$$

The method discussed above is rather huge and not convenient, so we will not recommend it for practice.
To our opinion for comparison of identical detector channels and selection of instrument variations it would be appropriate to use the method of ratio logarithms, i.e. values

$$l_{ij} = \ln r_{ij} = \ln N_i - \ln N_j + \ln N_j^0 - \ln N_i^0 \tag{5.13}$$

In this case, if the condition (5.3) is satisfied, we get

$$l_{ij} = \ln(1 + \delta_i^{dev}) - \ln(1 + \delta_j^{dev}) = \ln \varepsilon_i - \ln \varepsilon_j \tag{5.14}$$

and for small instrument variations $(\delta_i^{device} << 1)$:

$$l_{ij} = \delta_i^{device} - \delta_j^{device} = \varepsilon_i - \varepsilon_j \tag{5.15}$$

Using the logarithms of ratio we save all advantages of the ratio method: remove cosmic ray variations, get a simple relation between instrument variations and values measured in the experiment, calculate simply the channel efficiency.

Besides, the distribution of fluctuation values l_{ij} is symmetric and looks more like the normal distribution. Count rates, instrument variations and efficiencies of particular channels are presented similarly in the expression l_{ij}. So, substituting logarithms of ratios instead of ratios we simplify the statistical analysis of auxiliary information, preliminary processed, and make more accurate and reliable a process of determination of efficiencies and instrument variations of particular channels and their statistical errors.
We have supposed above that the instrument interference lead to multiplying of the count rate by some factor and the expression (5.3) can be used. Apparently, in most cases it would be so. However, we can imagine the instrument interference, which is an additive value in the expression for count rate (electronic noise, radioactive background and etc.). Then, instead (5.3) we should write

$$N_i = N_i^0(1+\delta)+\delta_i^{device} \tag{5.3a}$$

where δ_i^{device} would be already the additive instrument variation. Clear, in this case the method of ratios (and logarithm of ratios)does not allow removing of real variations δ, the method of differences would be more effective. Really, substituting (5.15) into (5.6) we get:

$$d_{ij} = c_j\delta_i^{device} - c_j\delta_j^{device} \tag{5.16}$$

The efficiency can be determined with an accuracy of the factor $\delta_a\delta/(1+\delta_m+\delta_a)$, if the ratios of count rates are used, where δ is the cosmic ray variation, δ_m and δ_a are respectively multiplicative and additive instrument variations.

In summary of the above discussion, we may say. The simplest and mostly reliable method for search of instrument variations and control of particular channel efficiency is the log-ratio method. This method should be recommended for practice. For search of additive instrument variations (for example, if would appear that the efficiency correlate with cosmic ray variations) normalized differences or ratios (difference)/(sum) should be used.

6. The Algorithm for Control of Multi Channel Cosmic Ray Detector

The algorithm proposed here for search of instrument variations and data quality control is used at cosmic ray stations: Kill, Moscow and Cape Shmidt. This algorithm processes data of the 18 and 24-channel neutron monitors and the 12-channel muon telescope.

Main characteristic features of the algorithm are:

1) A large number of identical (or near identical) channels is used;
2) Efficiencies of particular channels and the whole detector are calculated continuously;
3) Logarithms of count rate ratios are used for this purpose.

Count rates $N_i(i=1,...,k)$ measured in k different channels for equal time interval are processed. Here we will discuss hourly data. Data of each hour are processed gradually and independent from others. Current efficiencies ε_i of channels and the whole detector are calculated in respect to the initial efficiencies existed during some earlier definite period (the initial period. Adopting the initial efficiencies equal 1 is convenient. Initial count rates corresponding to the initial N_i^0 efficiencies are given for the initial period. Besides, efficiencies ε_i^0 just before the considered period (so-called, control efficiencies) are supposed to be known. An average estimate obtained by (5.14) using channels in order is adopted as an estimate of logarithm of the current efficiency of the channel - i

$$\ln\varepsilon_i = \sum_j g_{ij}(l_{ij} + \ln\varepsilon_j^0)/\sum g_{ij} \tag{6.1}$$

Here and below we sum over j for , $j \neq i$, $j \neq j_m$, where j_m - are numbers of channels out of order, g_{ij} - is a weight of given item depending on accuracy of l_{ij} value. A large number of channels allow estimating of instrument variation of the channel and its statistical error simultaneously. The last value is determined by the channel count rate and operation quality of all other channels, used for calculations of a given efficiency, i.e.

$$S_i^2 = S_{i1}^2 + S_{i2}^2 \tag{6.2}$$

Here we have

$$S_{i1}^2 = \frac{C_k}{C_F N_i^0 (1+\delta)} \tag{6.3}$$

in the denominator there is a count rate of given channel in a given moment, C_F - is a recalculating coefficient, the C_k coefficient determines a part of associated pulses in the count rate and depends on a type of registered secondary cosmic ray component, a detector construction and registration dead-time. A following expression gives the S_{i2} value:

$$S_{i2}^2 = \frac{\Sigma\, g_{ij}(l_{ij} + \ln\varepsilon_i^0) - \ln\varepsilon_i \Sigma\, g_{ij}(l_{ij} + \ln\varepsilon_j^0)}{(k-m-1)\Sigma\, g_{ij}} \tag{6.4}$$

where m – is a number of channels out of order.

All channels are considered in order in the beginning of processing, except channels not working $N_j = 0$. An estimate of the efficiency logarithm is calculated by (6.1) for all channels. Then, these estimates are compared with control efficiencies and the maximum module of difference is selected. We compare

$$\max_{i=1,\dots,k}\left(\left|\ln\varepsilon_i - \ln\varepsilon_i^0\right|\right) > q\sigma_i^0, \tag{6.5}$$

where $q\sigma_i^0$ - is a value depending on the adopted criteria of certainty. If the condition (6.5) is fulfilled, then the channel j_m for which a difference of efficiency logarithms has been maximum is supposed to be out of order and the procedure of efficiency calculations is repeated, but data of the channel are not used for calculations by (6.1) and (6.2).

If the condition (6.5) is not fulfilled, i.e. the current efficiencies are equal to the control efficiencies within statistical errors, then the efficiencies ε_i and their statistical errors σ_i calculated at a given step are considered as final estimates for current efficiencies and their errors. All remain channels are supposed to be in order.

The statistical error σ_i^0 used in (6.5), generally speaking, is different for different channels and varies in time along with the count rate:

$$\sigma_i^0 = \frac{C_k}{C_F(1+\delta)}\left(\frac{1}{N_i^0} + \frac{1}{(k-m)^2}\sum_j \frac{1}{N_i^0}\right) \quad (j \neq i,\, j = j_m) \tag{6.6}$$

The analysis of detector performance should be repeated twice per hour, because the $(1+\delta)$ factor is not known till instrument and natural variations would be separated. At first we suppose $\delta = 0$ in (6.6). Later, after calculations of the corrected count rate of the detector, a real value of $(1+\delta)$ is substituted in (6.6) and the analysis is performed once more with the corrected values σ_i^0.

A quality of the algorithm performance depends on a choice of the certainty criteria. If the q value is too small, then useful information would be rejected often. Oppositely, if the q-value is too large, then there is a possibility to channels out of order for calculations of the efficiencies. In this case the detector count rate would be determined wrongly as well as channel efficiencies and instrument variations. We adopted q=3.

The weight factors g_{ij} can be determined analyzing dispersions of count rate ratio $(\ln N_i - \ln N_j)$ for channels i and j during rather long time interval. The control efficiencies

supposed to be constant during this period ε_i^0 and ε_j^0. It is reasonable to introduce the weight factors, if the ratio dispersions are obtained with a good accuracy and the detector channels are not considerably different from each other. In many cases we may reasonably suppose $g_{ij} = 1$.
After the channels are divided into two groups – in order and out of order, the count rates of out of order channels are replaced by the corrected count rates

$$N_i^* = N_i / \varepsilon_i \qquad (6.7)$$

A special case occurs if the channel is not working at all and $N_i = 0$. The channel efficiency is $\varepsilon_i = 0$ and for data correction instead of (6.6) we should use an expression

$$N_i^* = N_i^0 \frac{\sum_j g_{ij} \frac{1}{\varepsilon_j^0} \frac{N_j}{N_j^0}}{\sum_j g_{ij}} \quad (j \neq i, j \neq j_m) \qquad (6.8)$$

The expressions (6.6) and (6.7) are equivalent if $\varepsilon_i = 0$.
We may calculate several summarized count rates characterizing the detector performance. If data of all channels are summarized without any correction, then we get the real count rate:

$$N = \sum_{i=1}^{k} N_i \qquad (6.9)$$

The sum of corrected count rates of the channels provides the corrected count rate (normalized to the initial efficiency:

$$N^8 = \sum_{i-=1}^{k} N_i / \varepsilon_i \qquad (6.10)$$

By comparison of real and corrected count rates we can easy get the current efficiency of the detector

$$\varepsilon = N / N^* \qquad (6.11)$$

If one or several channels are supposed to be out of order during a given hour, then only a part of the real count rate N (its "useful" part) is used for determination of the normalized count rate N^*

$$\overline{N}^* = \sum_{i=1}^{k} N_i \quad (i \neq j_{mm}) \qquad (6.12)$$

A ratio

$$P = \overline{N} / N \qquad (6.13)$$

provides a coefficient of the detector efficiency. However, we should not forget that the coefficient would be determined not only by a quality of the instrument, but also by characteristics of the instrument control and data editing. If for one hour it is not so important do we know a number of channels in order *k-m* or the coefficient P, then for averaging over a long period (day, month, year) it is more suitable to use the characteristic like (6.12).
The statistical error of the detector count rate is much more important characteristic of the detector. We can estimate an expected error by the sum N^*

$$S_1^2 = \frac{C_k^2}{C_F N^*} \qquad (6.14)$$

where C_k^2 - is determines by a part of associated with each other pulses in the detector count rate. The same statistical error we can estimate by dispersion of detector count rate estimates, using multi-channel detector, obtained by each corrected channel

$$N_i^* = \varepsilon_i^0 \frac{N_i}{N_i^0} \sum_{i \neq j_m} \frac{N_i^0}{\varepsilon_i^0} \qquad (6.15)$$

If weight coefficients are $g_{IJ} = 1$, then

$$S_2^2 = \frac{1}{m(m-1)} \sum_{i \neq j_m} \left(1 - \frac{N_i^0}{\varepsilon_i^0}\right)^2 = \frac{1}{m-1}\left[\frac{1}{m(N^*)^2} \sum_{i \neq j_m} (N_i^*)^2 - 1\right] \qquad (6.16)$$

If by some accidental reasons we have $S_2 < S_1$ during a given hour, then as an estimate of the statistical error should be adopted S_1.

7. The Program - SuperEditor.

You may take freely the program "SuperEditor" on the IZMIRAN server. The program corresponds to the algorithm described above for primary processing of data from multi-channel cosmic ray detector.

The operator calls the sub-program:

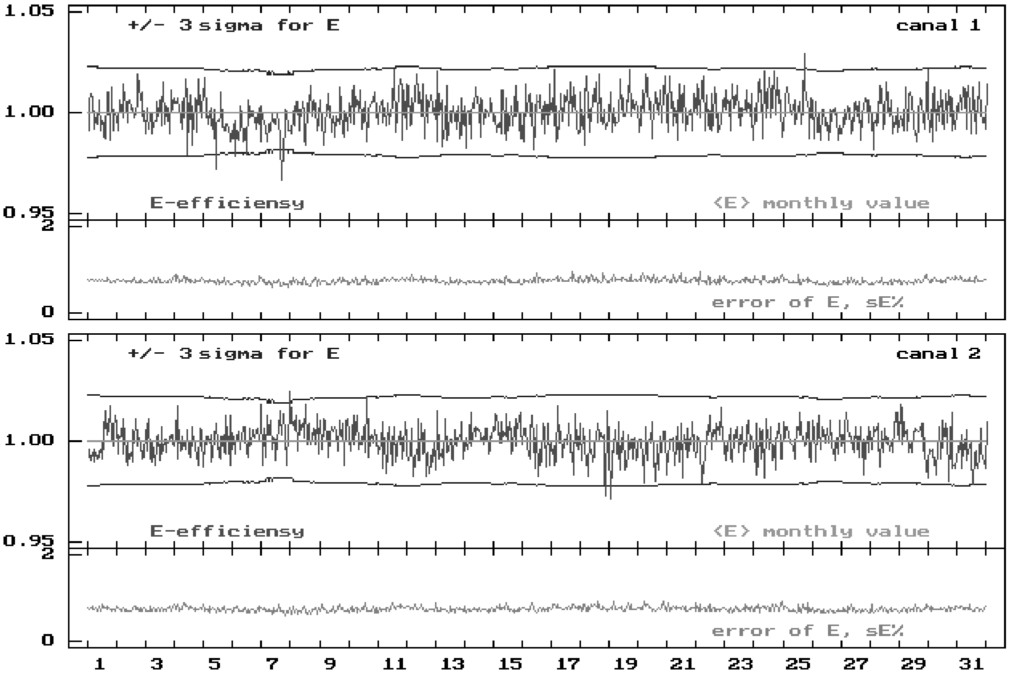
SuperEditor(K,N0,E0,N,N1,E,SB,SH,S,NS,SN,ST,ES,P,IN,SF,SK,Q)

Input parameters of the sub-program are:

κ	-a number of channels;
N0	real vector of initial count rates (This and all other vectors presented in the list of formal parameters have a dimension κ);
E0	- vector of control channel efficiencies;
N	- real vector of processed count rates;
SF	- recalculation coefficient;
CK	- coefficient showing how greater should be a dispersion of count rates in absence of instrument variations in comparison with the poison dispersion;
Q	- a number of expected standard deviations, this limit should not be exceeded for channels in order.
E	- vector of calculated channel efficiencies (for channels out of order this value is negative);
S	- vector standard mean-square errors of the channel efficiencies;
N1	- vector of corrected channel count rates;
SH,SB	- vectors of lower and upper boundaries of the confidence interval, where may vary an efficiency of channel in order;
NS	- corrected count rate of the detector;
SN	- standard mean-square relative error of the corrected detector counts rate;
ST	- its theoretical estimate;
ES	- calculated detector efficiency;
P	- coefficient of the detector efficiency;
IN	- integer vector, its components show a condition of channels: 1- for channels in order

and 0 – for channels out of order.

If a number of channels in order is less than к/2 or less 2, then the entire exposition is rejected. In all other cases the program calculate following output parameters:



8. Some Results

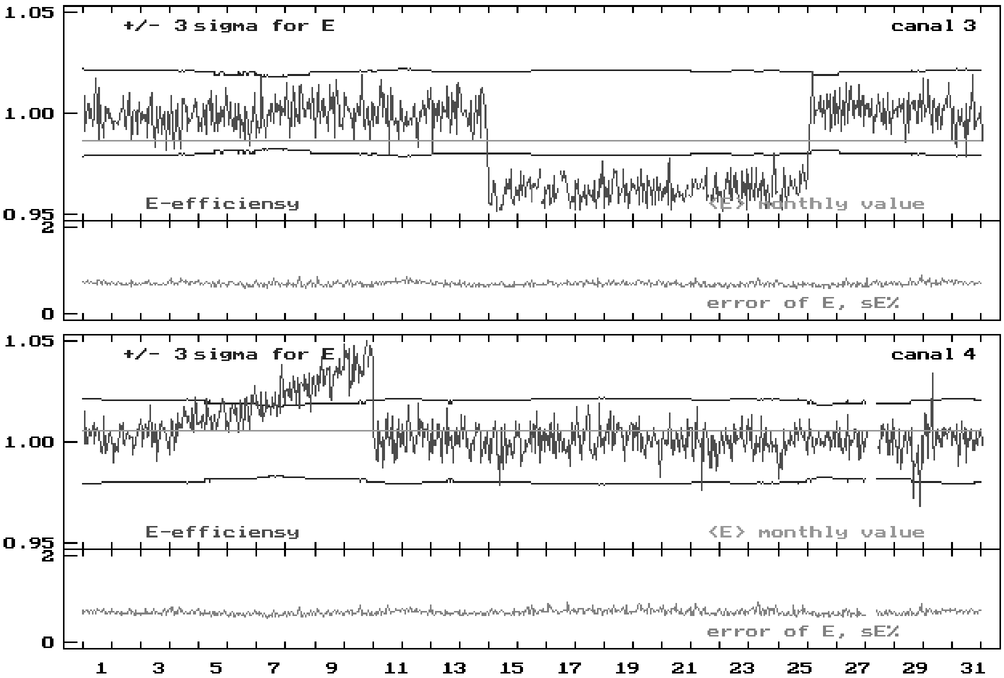
The program was tested by imitations of different instrument variations and changes of number of processed channels. Here we discuss an application of the program to data of one section of the 24NM-64 Moscow neutron monitor, i.e. data of six channels. We choose for illustration the period of January 1-31, 1999. Instrument variations were generated artificially in the 3-rd and 4-th channels. Figure 1 shows all types of instrument variations discussed above in the time-history of channel efficiencies. Data gaps occurred on January 28 in the 4th channel. Negative выбросы are observed, for instance, in the 4th channel on January 29 at 14.00 and 20.00 UT . Positive выбросы were in data of the 4th channel on January 30. In the 3rd channel a jump of count rate (1.03) has been generated artificially from January 15 to January 25, that is clear seen in the plot of efficiencies.

The instrument drift generated in the 4th channel from January 3 to January 10 (the count rate drift from 1 to 0.96) is well pronounced. Efficiencies of the 4th channel for a part of the considered period (January 6 and 7) are considerably different from the adopted control efficiencies. These differences are not so large to suppose the channel data are wrong, so during a long period the corrected count rate of the detector has been calculated by using count rates of all channels, including the 4th channel. Therefore, the dispersion of detector count rate on January 6 and 7 is considerably greater than the expected dispersion. The obstacle, that data of the 4th channel are rejected for some hours and assumed to be correct for others, causes additional fluctuations of the detector count rate, so the dispersion may appear to be greater than the estimates presented here.

From another hand, considerable instrument variations are not observed in data of channels 1,2,5,6 besides some gaps and several short peaks. The distribution of efficiency fluctuations is close to the normal distribution and correspond well to the expected values of mean square deviations (for hourly channel efficiency ~0,6%).

Changing control efficiencies in the 3rd channel (basing on the obtained data) we decrease considerably an amount of rejected information, moreover, the dispersion of detector count rate decreases as well as fluctuations of the detector efficiency and count rate.

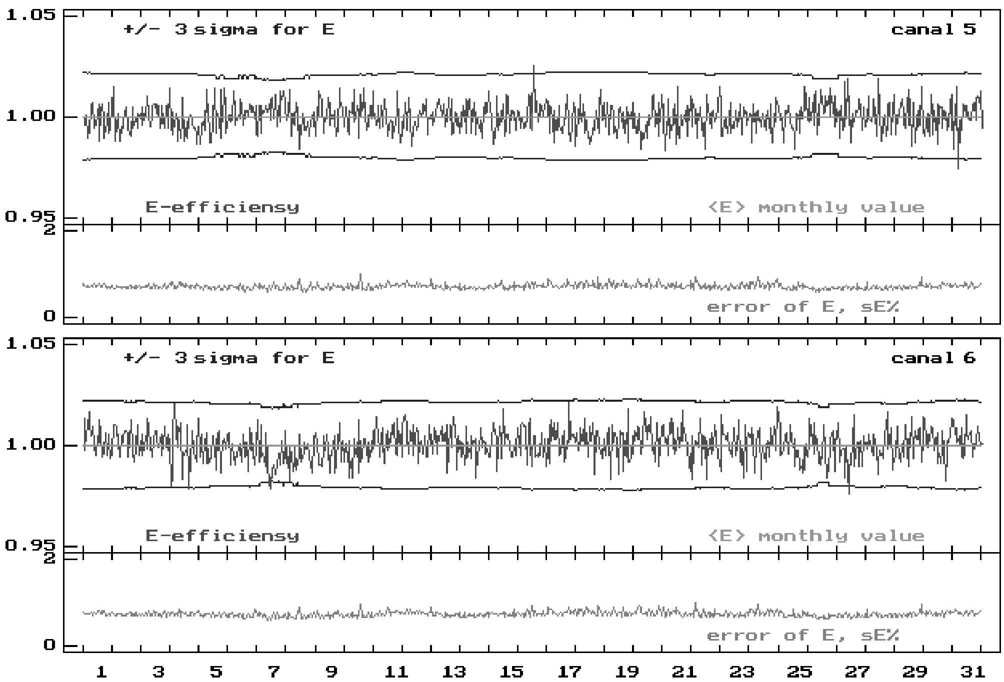
Figure 2 shows the corrected count rate and neutron monitor efficiency calculated by the formulae (6.11), the theoretical estimate (6.14) for the statistical error of count rate and its experimental value (6.16), which are practically coincide with each other.



9. Possibilities for Developing of Primary Data Processing

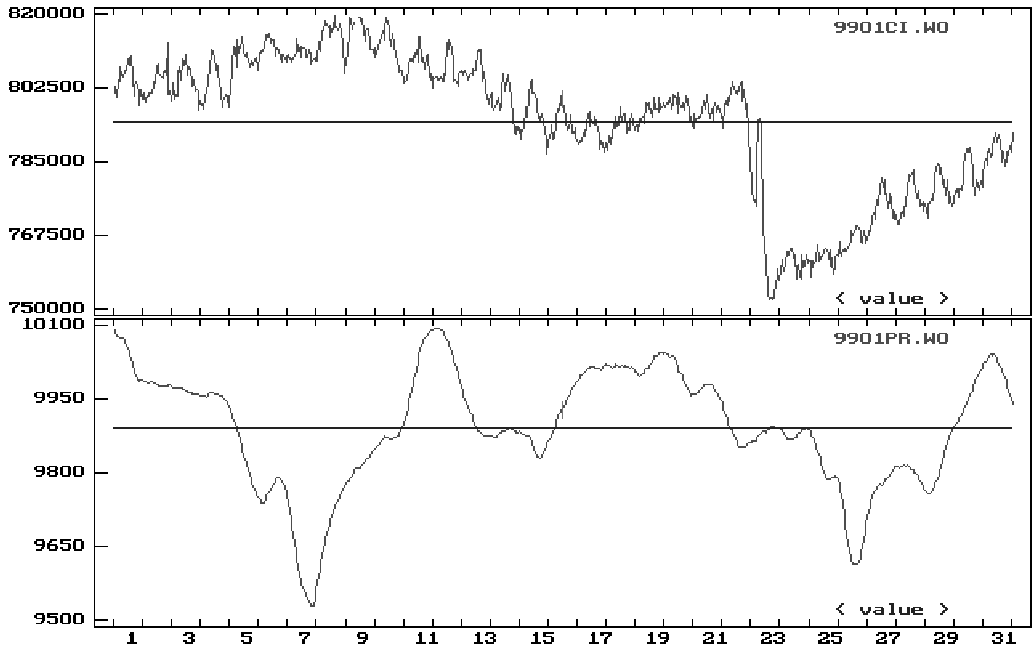
Let us discuss benefits and disadvantages of the algorithm presented here for instrument control and processing of cosmic ray detector data.

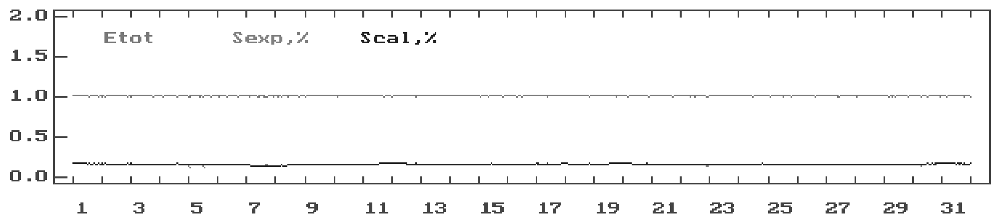
Main dignities of the algorithm, to our opinion, are its relative simplicity, consequent usage of the multi-channel principle and the detector efficiency. Using of logarithms of channel count rate ratios provides a rather accurate and reliable method for determination of the channel efficiencies and instrument variations, even when instrument variations and data statistical fluctuations are rather high. The algorithm allows selecting in a given moment channels out of order according to the



criteria adopted previously. It is possible continuously determining of current efficiencies of each channel and the entire detector. The algorithm provides corrected detector count rate; data of channels out of order do not effect on this value. The algorithm provides statistical errors for current channel efficiencies and corrected detector count rate based on data dispersion of particular channels in order. Information obtained by the algorithm allows estimating a quality of detector

performance and efficiency of the algorithm.
Therefore, we can consider the algorithm as a good base for primary processing of data of any multi-channel cosmic ray detector. However, disadvantages of the algorithm are clear.
If the instrument variation in a channel is high, but not so high to adopt the channel as out of order, then the channel effects unfavorably on calculated efficiencies of other channels. Using slow values instead of averages from channels in order we may eliminate this problem.





Other disadvantages of the algorithm are:
1. It is not fully automated, the control efficiencies should be changed sometimes;
2. It is not suitable for searches of instrument jumps and drifts. Really, small jumps and drifts the algorithm does not consider, but large jumps and drifts, in fact, are adopted as peaks. Moreover, useful information is often removed;
3. It is not suitable for search of instrument dispersion variations and higher momentum of the count rate statistical distribution.
The next step of developing of primary data processing algorithm should be time-history analysis of obtained channel efficiencies. That would allow to select instrument jumps and drifts in any particular channel, change in time control efficiencies, and? Therefore, to conserve necessary information and increase an accuracy of detector data.
There are two principally different approaches for the analysis of efficiency time-history: retrospective and adaptive. The retrospective analysis assumes that firstly efficiency estimates should be obtained, then a time-history of efficiencies would be approximated. The count rate would be corrected by using this approximation. So, for data correction of a given hour information obtained before and after the hour is used. Obviously, this approach is not suitable for data processing in real time.
The adaptive method is for work in real-time regime. The adaptive approach at the expense of less accuracy allows fully automated data processing and provides large opportunities for operative control of instruments and quickly access to useful information.
Another possibility of developing of primary data processing is associated with analysis of data fluctuations in particular channels and searches of correlation between them. We have supposed above that data of particular channels are independent and their dispersion varies with count rate in

all channels similarly. Both statements are clear simplifications. The operative control of fluctuation distribution in channels and channel correlation's would allow to select dispersion instrument variations, correct criteria for channels in order, determine more accurately instrument variations common for several channels and etc.

Using simultaneously two approaches of channel comparison (by the ratio logarithms and the normalized differences) we have a possibility to separate multiplicative and additive instrument variations.

Acknowledgements. This work is supported by Russian Federal Program "Astronomy" and the Russian Foundation for Fundamental Researth, grant №. 98-02-17315 and № 99-02-18003.

References

Blokh, Ya.L., Pimenov, I.A., Standard Algorithm for Cosmic Ray Station, Geomagnetizm i Aeronomiya, 1969, vol. 9, no. 1, pp. 53-55.

Blokh, Ya.L., Poznyak, A.S., Simsar'yan, R.A.:1976, "Some Principals and Algorithms of the System for Primary Processing of Solar Cosmic Ray Data", *Kosmicheskie Luchi*, Moscow: Nauka, no. 16, pp. 45-50.

Blokh, Ya.L., Poznyak, A.S., Simsar'yan, R.A.:1976, "Adaptive Algorithm for Search of Istrument Mulfunction during Soalr Cosmic Ray Data Processing", *Kosmicheskie Luchi*, Moscow: Nauka, no. 16, pp. 51-55.

Blokh, Ya.L., Libin, I.Ya., Simsar'yan, R.A., Starkov, F.A.:1976, "Algorithm for Search of Mulfunction in Supermonitor Data", *Kosmicheskie Luchi*, Moscow: Nauka, no. 16, pp. 51-55.

Carmichael H.:1962, "NM-64 Neutron monitor specitication. U.S.A. Deep River Laboratory", 70 p.

Compton A.H.. Wollan E.0. Bennet R.D.:1934, "A precision recording cosmic ray meter", *Rev. Scient Instrum.*, v. 5, N 12, p. 415-422.

Dorman, L.I., Pimenov, I.A., Satsuk, V.V.:1978, "Mathematical Support of Geophysical Research by using Example of Cosmic Rays", Moscow: Nauka, p. 152.

Dyring E., Sporre B.:1970, "Statical reliability tests of cosmic ray intensity data", *Acta Physica Academial Scientiarum Hongarical*, . .Suppl. 2., p.. 95-98.

Janossy, L.:1965, "Theory and Practice of the Evaluation of Measurements", Oxford University Press.

Kudela, K.:1974, "Some Criteria for Quality of Multi-channel Neutron Supermonitor Measurements", *Izv. AN SSSR, ser. fiz.*, vol. 38, no. 9, 1982-1985.

Shafer, Yu.G.:1958, "Continuous Registration of Cosmic Ray Intensity Variations by the Ionization Chamber with Automate Control", *Izd. AN SSSR, Trudy YaFAN, ser. fiz.*, no. 2, pp. 7-22.

Simpson I.A.:1955, "Neutron monitor". Chicago, 60 p.

Vergpib, A.B., Luzov, A.A.:1974, "On Estimates of Data Reliability by Cosmic Ray Intensity Variations, in Investigations of Geomagnetizm", *Aeronomie and Solar Physics*, Moscow: Nauka, vol. 31, pp. 161-165.

Vergpib, A.B., Luzov, A.A., Sergeev, A.V., et al.:1975, "Practical Method for Estimates of Cosmic Ray Data Reliability, in Investigations of Geomagnetizm", *Aeronomie and Solar Physics*, Moscow: Nauka, vol. 35, pp. 161-166.

Belov A., Blokh Ya., Klepach E., Yanke V.Primary Processing of Cosmic Ray Station Data:
Algorithm, Computer Program and Realization. сб. Космические лучи, N25, 113-134, М., Наука, 1988.