This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier’s archiving and manuscript policies are encouraged to visit:

http://www.elsevier.com/copyright
Median filtering algorithms for multichannel detectors

A. Hovhannisyan *, A. Chilingarian

Alikhanyan Physics Institute, Alikhanyan Brothers 2, Yerevan 375036, Armenia

Received 12 March 2010; received in revised form 28 December 2010; accepted 13 January 2011
Available online 20 January 2011

Abstract

Particle detectors of worldwide networks are continuously measuring various secondary particle fluxes incident on Earth surface. At the Aragats Space Environmental Center (ASEC), the data of 12 cosmic ray particle detectors with a total of ~280 measuring channels (count rates of electrons, muons and neutrons channels) are sent each minute via wireless bridges to a MySQL database. These time series are used for the different tasks of off-line physical analysis and for online forewarning services. Usually long time series contain several types of errors (gaps due to failures of high or low voltage power supply, spurious spikes due to radio interferences, abrupt changes of mean values of several channels or/and slowly trends in mean values due to aging of electronics components, etc.). To avoid erroneous physical inference and false alarms of alerting systems we introduce offline and online filters to “purify” multiple time-series. In the presented paper we classify possible mistakes in time series and introduce median filtering algorithms for online and off-line “purification” of multiple time-series.

Keywords: Median algorithms; Cosmic rays; Time series

1. Introduction

The networks of ground-based particle detectors measure time series of secondary particles produced in the interactions of primary ions in the terrestrial atmosphere. Galactic Cosmic Rays (GCR, mostly protons and heavier fully stripped nuclei), are accelerated in our Galaxy by shock waves originated in supernova explosions and by other exotic stellar sources. After traveling millions of years in our Galaxy, GCR arrive in the solar system as highly isotropic and stable flux. On the other hand, our Sun is a variable object changing intensity of radiation and particle fluxes several orders of magnitude in a few minutes. During high-energetic events at the Sun solar cosmic rays (SCR) are produced. The SCR particles can travel away from the Sun along the open magnetic field lines. If the magnetic field lines are connected with the Earth and SCR particles with enough energy are present, the SCR particles produce secondary cosmic ray particles in interactions with the atoms of the Earth atmosphere. SCR have a major influence on the Earth, safety of satellites, electrical lines and pipelines on and under the Earth’s surface and also probably climate (Svensmark, 1998). The influence of the Sun on GCR flux can be described as a modulation of the usually stable “background”. The Sun modulates GCR in several ways. The solar wind “blows out” the lowest energy GCR from the solar system, thus changing the GCR flux intensity inverse proportionally to the Sun activity. Huge magnetized plasma clouds and shocks initiated by Coronal Mass Ejections (CME) are traveling in the interplanetary space tens of hours before reaching the Earth, changing intensity and direction of the GCRs. On arrival at the Earth the magnetic field of the plasma cloud depletes the GCRs, measured as a decrease in the count rates of the ground based cosmic ray detectors (so called Forbush decrease), or the magnetic cloud interacts with the Earth’s magnetic field and causes geomagnetic effects, measured as an enhancement in count rates of secondary particles for several hours. Besides modulation effects SCR also influence on count rate of secondary particles. The explosive flaring processes on the Sun result in the ejection of...
huge amounts of solar plasma and in acceleration of the copious electrons and ions. The SCR may reach the Earth and initiate secondary elementary particles in the terrestrial atmosphere, increasing the count rates of ground based particle monitors from several up to several thousand percents. This effect is called Ground Level Enhancement (GLE).

Hybrid particle monitors at Aragats Space Environmental Center (ASEC, Chilingarian et al., 2003, 2005) measure both charged and neutral components of secondary cosmic rays; ASEC data provides good coverage of the violent Solar Energetic Particle (SEP) events of the solar activity cycle 23. The multivariate correlation techniques applied upon detected fluxes of charged and neutral particles are used to study geo-effective events, i.e. GLEs, Forbush decreases, Geomagnetic Storms; and for reconstruction of the energy spectra of SCR (Chilingarian and Reymers, 2007).

The particle monitors are located in the two research stations on the slopes of Aragats Mountain at altitudes 2000 and 3200 m above sea level and are connected with the data analysis center in Yerevan by means of a radio network. Additionally, there is an ongoing process of establishing a SEVAN worldwide network of detectors operating at different latitudes, longitudes and altitudes (Chilingarian et al., 2008a; Chilingarian and Reymers, 2008).

During solar activity cycle 23 (1997–2008) the old type DAQ (Data Acquisition) electronics used in ASEC had often malfunctioned and there were many errors in the time series (see Fig. 1).

For the physical analysis we need to “purify” (correct, filter, smooth) the raw data. Filtering algorithms are usually based on the comparisons of data from identical measuring channels. NMs (Neutron Monitor) (Clem and Dorman, 2000) located at the stations Aragats and Nor-Amberd consist of three sections, six identical proportional counters in each. If the ratio of count rates of different sections is changing within defined limits the detector overall count rate is performed by simple summation of all sections. If one section of NM is defective, the ratios of count rate that contain data of defective section will be out of defined limits, and one should exclude it from summation and the NM count rate should be properly normalized. The same method can be applied for the counters within one section. However, this approach has several disadvantages. It is not fully automated, control parameters should be currently tuned, and the algorithm does not correct abrupt jumps.

The algorithms based on median filtering are currently widely used in pattern recognition and in multimedia technologies and scientific applications. For instance, to maximize data output from single-shot astronomical images the rejection of the cosmic ray background is made by the median algorithms (Farage and Pimbblet, 2005).

This paper consists of the following sections.

1. Classification of particle detector failures.
2. The description of “horizontal” median algorithm.
3. The description of “vertical” median algorithm.
4. The verification and the implementation of the method to the data of the Nor-Amberd NM during solar cycle 23.
5. Monitoring the stability of the measuring channels.

The advantages of the proposed method are

1. The algorithms are simple and do not need much computer resources for their realization.
2. They correct abrupt changes of means of measuring channels as well as spurious spikes.

Fig. 1. Aragats NM raw data with errors (four channels from 18 are introduced for better view); totally eight days; the time series are artificially shifted from each other; time is measured in minutes.
Using the obtained coefficients described in Section 3 we can monitor the stability of measuring channels and reveal even slow drifts of the channel-means.

The algorithms can be used not only for offline, but also for online data filtering.

### 2. Classification of particle detector failures

Below we introduce four main kinds of errors in the time series of secondary cosmic rays.

Fig. 2a shows spikes – the most common type of error that happens in time series of NM damages scintillation detectors. The reason of such spikes can be connected with electronics or abrupt changing of atmospheric conditions (thunderstorm, lightning, etc.). In Fig. 2b the drift of the count rate of a detector is shown. This effect can be caused both by electronics malfunctioning and instrumental failure. The possible reasons of electronics malfunctioning may be the degradation of discrete components of DAQ systems. The last two kinds of errors are caused by the electricity break on the high altitude station due to severe snowstorms (Fig. 2c and d).

Mentioned errors in particle detectors' operation can lead to erroneous estimation of the FD (Forbush Decrease) magnitude, prevent detection of the GLE (usually not very large at middle latitudes – 1.5–2%), etc. We introduce algorithms based on the stabilizing properties of the median for correction of multichannel detector’s data. We heavily use the “overabundance” of ASEC data due to numerous identical channels measuring one and the same physical quantity (flux of particles of definite type).

### 3. The description of horizontal median algorithm

The first algorithm we introduce is Moving Median Filter. Window of size \( L \) is “moving” from beginning to the end of the time series with unit step, the median of \( L \) values falling within window are continuously calculated. In Fig. 3 the flowchart of MMF is presented and below the detailed description of the algorithm is introduced.

**Description of algorithm:**

**Notion:**

- Count rate of detector channel at moment \( i \) is denoted by small letter \( v_i \); median of \( L \) successive count rate values where time \( i \) is the middle point if window – \( M_{i,L} \).
- Width of moving window – \( L \).
- Minimal and maximal width of window – \( L_{\min} \) and \( L_{\max} \).
- Maximal and minimal limits of \( M_{i,L} \) – \( P_{\max} \), \( P_{\min} \).
- Limit of maximal deviation of current count rate from median value (after different tests 5–6 \( \sigma \) was obtained as best value, where \( \sigma \) is Root Mean Square (RMS) of time series) – \( D_{\max} \).

**Steps of algorithm:**

1. Select time series from database with \( N \) elements.
2. Start algorithm operation, assign \( L = L_{\min} \) and \( i = L/2 \).
3. If \( L < L_{\max} \), continue, otherwise report program failure at point \( i \), increment \( i \) by 1, assign \( L = L_{\min} \).
4. If \( i < (N - L/2) \), calculate median value \( M_{i,L} \) for interval \([i - L/2, i + L/2]\), otherwise write filtered time series into database, stop.
5. If \( M_{i,L} \in [P_{\min}, P_{\max}] \), assign \( L = L_{\min} \), continue, otherwise enlarge \( L \) by factor of 2 and go to step 3.
6. If absolute value of \( (v_i - M_{i,L}) > D_{\max} \), assign \( v_i = M_{i,L} \).
7. Increment \( i \) by 1, go to step 3.

This filter is optimal for time series containing spurious short spikes and abrupt changes of mean followed by recovery (Fig. 2a and c). It cannot correct smooth trends due to slowly changing parameters of particle detectors due to altering of Photomultiplier (PM) or electronic elements of Amplitude-to-Digital-Converter (ADC) and power supply (Fig. 2b). Also it cannot recover long lasting (greater than width of moving window) absence of data (Fig. 2d). So another algorithm should be used to correct such defects using the data of other channels of the same monitor.

### 4. The description of vertical median algorithm

Let’s suppose that the detector consists of \( K \) identical channels, however due to individual characteristics of used sensors (photomultipliers, proportional counters, etc.) the mean count rates of channels \( n_j, j = 1, \ldots, K \), are dispersed within definite limits (in case of Nor-Amberd and Aragats NM the maximum deviation of one channel from the average count rate per channel is 12%).
Notion:

- \( K \) – number of the channel (it should be >2, if \( K = 2 \) it will be difficult to distinguish automatically which one has operated properly and which one was defective);
- \( \bar{n}_j \) – mean count rate of \( j \)th channel;
- \( \bar{N}_{\text{total}} \) – sum of mean values of all channels (detector mean count rate);
- \( \text{med} = \text{med}(F_j \bar{n}^i)_{j=1, \ldots, K} \) – median value of \( K \) channels at \( i \)th minute;

\[^{1}\] This median is different from median used by MMF algorithm. MMF median “moves” along the time series therefore we name it “horizontal” median. Median used by RMF is applied to different time series at the same minute. Because usually for display purposes different time series are stacked vertically (see Figs. 1 and 4) we name this median – “vertical”.

\( F_j \) – the equalizing coefficient of \( j \)th channel;

\( V^i_j \) – \( i \)th data point of \( j \)th channel;

\( V^i_i \) – estimate of the total detector count rate at moment \( i \) according to Eq. (2).

At the start of detector operation it is possible to equalize the mean count rates by assigning to each channel the appropriate coefficient \( F_j j = 1, \ldots, M \):

\[
F_j = \left( \frac{\bar{n}_j \times K}{\bar{N}_{\text{total}}} \right)^{-1}, \quad j = 1, K
\]  

(1)

Having equalizing coefficients it is possible to create one time series which will be equivalent to the sum of all \( K \) time series by the following equation.

\[
V^i_i = K \times \text{med}_i.
\]  

(2)
Also it is possible to recover all time series one by one using equalizing coefficients. If \( j \)th channel of detector is continuously and incoherently changing (operating unstable according to reports of MMF) all data points with errors can be substituted by the following formula:

\[
v_j' = \frac{\text{med}_i}{F_j}.
\]

The possible scenario of implementation of both algorithms can be as follows:

1. The initial period of time series should be selected to contain no errors (spikes, drifts, jumps). For that period the mean count rates \( \bar{n}_j \) and coefficients \( F_j \) are calculated and stored.

2. The data of all channels of one detector are filtered with the MMF algorithm at the end of each day. All time periods of the different channels, for which the MMF method could not smooth the data, are stored in memory;

3. RMF turns on. The data that could not be corrected by MMF, are read out from the memory as well as the stored mean values and the coefficients. Then RMF corrects the data of the erroneous or missing channels according to Eqs. (1) and (3). Means and appropriate coefficients are renewed and stored.

4. If the second algorithm does not correct the data (which means that more than half channels have been corrupted or the detector was switched off due to some overall failure) the system sends an e-mail to the operator.

By automatically implementing both filtering algorithms each day and storing renewed mean values of channels and appropriate coefficients, it is possible to correct all mentioned failures (demonstrated in Fig. 2). The time history

![Fig. 4. Aragats NM data, May 2008 – after filtering.](image4)

![Fig. 5. Pressure measurements at Nor-Amberd Station, before filtering.](image5)
of the equalizing coefficients will help to outline non-stable channels and correct them. In next section we demonstrate some examples of filtering by these algorithms.

5. The verification and the implementation of the method to the data of the Nor-Amberd NM during solar cycle 23

In Fig. 4 we present the corrected data of Aragats NM in May 2008 (see Fig. 1 for raw data). Time series in the figure are artificially shifted from each other for better assessment. The combination of MMF and RMF corrects data, eliminating spikes and filling corrupted periods. The data in all graphs shown here has been smoothed by constant window \( L = 60 \) min. The data of Nor-Amberd NM during solar cycle 23 was smoothed by changing width of the window started from \( L = 60 \) to \( L_{\text{max}} = 600 \) min, which means that these data was smoothed by non-constant window, i.e. during MMF if median of data in current window is out of limits, the window will be enlarged as written in the description of MMF algorithm.

In Fig. 5 the atmospheric pressure measurements performed at Nor-Amberd are depicted. It is easy to remove the spikes by using the MMF algorithm (due to failures of the pressure sensor) from these data, see Fig. 6. Only the MMF algorithm was used to correct data. As it can be seen from the graph mostly all spikes where removed. Several not big spikes remained because their amplitude was smaller than \( D_{\text{max}} \).

In Figs. 7 and 8, we present the correction algorithm operating according to the Eq. (2). After the execution of this correction algorithm the data were smoothed by MMF algorithm. Although the variance of the filtered time...
series is larger comparing with initial ones, it corrects many mistakes apparent in the raw data (red – raw data, blue – corrected).

In Fig. 9 we present the data of the Nor-Amberd NM as measured during solar cycle 23 (1997–2009). Due to numerous failures of several detectors at the end of the cycle, the overall count rate of monitor was underestimated. The efficiency of several channels decreased due to failures of high voltage supplies and aging effects of counters itself, and overall count rate of monitor also decreases.

As we can see in Fig. 10 after applying filtering algorithms overall pattern of monitor count rate changes according the solar activity and becomes comparable with the data of Moscow NM at the same period.

To check the suggested filtering techniques we perform simulation study with artificially corrupted time series. 18 time series with 1 million points have been simulated according to mean values and variances of the channels of Nor-Amberd NM (Fig. 11). A trend has been introduced to time series to imitate solar modulation. Then all four described types of failures were randomly introduced in the time series (see Fig. 12). In Fig. 13 one can see that after correction with median algorithms all errors have been washed out and the solar “modulation” effect is apparently seen.

6. Monitoring the stability of the measuring channels

During multiyear operation mean count rates of particle monitors continuously alter not only by solar modulation or possible entering of regions where Galaxy arms are sending abundant GCR from supernovae explosions, but
also by such elementary effects as electronic components aging. Therefore, to identify instrumental failures and to avoid exploration of the artifacts instead of new physics we have to monitor the parameters of the detectors carefully and continuously.

In this paper, we present a simple method to do it by monitoring equalizing coefficients of monitor channels. The monthly (or decade) plots of coefficients will help to find unstable channels.

The channel mean count rates are changing due to solar modulation effects, in contrast, the equalizing coefficients (see Section 3) should be stable despite changing means. Therefore, it will be much easier to detect non-stable channels by monitoring the plots of coefficients, than changing channel means. Figs. 14 and 15 are an example of our approach.

Although from Fig. 14 we can notice that the variations of two channels are significantly larger than variations of

Fig. 10. Data of Moscow NM in comparison with the filtered data of Nor-Amberd NM during solar cycle 23.

Fig. 11. Simulated and artificially spoiled time series with 1 million data points.
the other 16, in Fig. 15 the behavior of the corresponding coefficients demonstrate failure of two channels much more pronounced. The same method can be implemented to check data from numerous detectors. Several Eurasian countries joint the efforts to establish a Neutron Monitor Data Base (NMDB, data available from www.nmdb.eu). Data from numerous NMs are gathered in NMDB and for the “housekeeping”, it is necessary to run periodically tasks for checking if all monitors are adequate and the data is correct and coherent. We select time series from several monitors participating in the NMDB project to run the “coherence” test (see Table 1).

We have taken pressure corrected data of these monitors for the time period 24.10.2008–25.12.2008 and calculated coefficients for these seven time series, according to Eq. (1). In Fig. 16 the raw data are posted; in Fig. 17 we present the same data corrected with MMF (note, that the spikes in Nor-Amberd and Izmiran monitors are filtered out, the gaps were filled by lines for better performance). In Fig. 18 we present the variation of the “equalizing” coefficients in percents for all seven monitors calculated for each day of selected period. As you can see, six coefficients from seven demonstrate very stable behavior, proving that all parameters of the NM chambers remain stable and constant. The calculated coefficients for the Athens NM are much more variable. In Fig. 19a and b are shown distributions of Athens and Nor-Amberd NMs’ time series correspondingly calculated during the period where the
variation of Athens NM’s coefficient was maximal. The histograms are fitted by Gaussian function. As it can be seen from these figures in case of Athens NM chi-square 

\[ \frac{\chi^2}{n \, df} = \frac{1765}{12}. \]

On the other hand the Aragats NM data exactly follows the Gaussian distribution. The distributions prove that Athens NM data contained errors during the observed period. The high variability (non-coherence) of the Athens monitor’s coefficient may be caused by malfunctioning of the electronics parameters (including pressure sensor) in the end of 2008. When dealing with multiple measuring channels it is of vital importance to develop a number of quality tests to check continuously the data coming from different remote destinations.

**Table 1**

Neutron Monitors used in the "coherence" test.

<table>
<thead>
<tr>
<th>Neutron Monitor</th>
<th>Altitude, m</th>
<th>R rigidity, GV</th>
<th>Type</th>
<th>Average count rate per second</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alma Ata</td>
<td>3340</td>
<td>6.69</td>
<td>18NM64</td>
<td>1361.63</td>
</tr>
<tr>
<td>Rome</td>
<td>60</td>
<td>6.32</td>
<td>17NM64</td>
<td>150.16</td>
</tr>
<tr>
<td>Aragats</td>
<td>3200</td>
<td>7.14</td>
<td>18NM64</td>
<td>721.77</td>
</tr>
<tr>
<td>Nor-Amberd</td>
<td>2000</td>
<td>7.14</td>
<td>18NM64</td>
<td>455.12</td>
</tr>
<tr>
<td>Moscow</td>
<td>200</td>
<td>2.46</td>
<td>24NM64</td>
<td>239.59</td>
</tr>
<tr>
<td>Oulu</td>
<td>0</td>
<td>0.81</td>
<td>9NM64</td>
<td>106.04</td>
</tr>
<tr>
<td>Athens</td>
<td>260</td>
<td>8.53</td>
<td>6NM64</td>
<td>54.49</td>
</tr>
</tbody>
</table>

**Fig. 14.** Day-to-day changes of the mean values of Aragats NM; at November 22 there were power supply cutoff at Aragats station.

**Fig. 15.** Day-to-day changes of the channel coefficients of Aragats NM.
Although the data of NMDB detectors are similar, different groups use different data acquisition electronics, pressure sensors and data transfer protocols. The time history of the “equalizing coefficients” is one of such tests to help keep NMDB data reliable and adequate for further physical inference.

7. Influence of filtering algorithms on thunderstorm triggered enhancements of count rate of secondary cosmic rays

The Aragats Space Environment Center facilities are routinely measuring fluxes of neutral and charged secondary cosmic ray particles that reach the Earth’s surface. In 2009 we detected simultaneously very large fluxes of electrons, gamma-rays and neutrons correlated with thunderstorm activity. During the period of the count rate enhancements lasting tens of minutes, millions of additional particles were detected (Chilingarian et al., 2008b). These enhancements last about 10–20 min, and the count rate of monitors can increase up to 100% during such periods. Therefore, we have to check how the filtering algorithms influence on these enhancements in the data. For that purpose event on May 21, 2009 has been selected to be tested with the algorithms described above. In Fig. 20 data of Aragats Multichannel Muon Monitor on May 21, 2009 is shown before and after filtering. Right peak on the picture is the enhancement associated with thunderstorm activity, and others are consequence of malfunctioning of some channels. As it can be seen Moving Median Filter has removed all errors from the data, but it has also
Fig. 18. Equalizing coefficients for seven NMs; data taken from NMDB.

Fig. 19. Distributions of Athens and Nor-Amberd NMs’ data calculated for time period with duration of five days.
removed the signal. In Fig. 21 is presented the effect of same algorithm with 20 min window. As it can be seen the filter hasn’t removed the signal, but it has distorted it. In Fig. 22 we present the same data filtered with Relational Median Filter. Not all errors have been removed successfully, but the signal has stayed as it was. So if one wants to investigate long time variations of cosmic rays during solar cycles, we recommend using a combination of the two algorithms to remove all spikes and also thunderstorm triggered enhancements, if we are interested in revealing thunderstorm events only Relational Median Filter should be used.

MMF would treat geomagnetic effects and GLEs in same way as it treats thunderstorm events, if duration of these effects doesn’t exceed the length of moving window, i.e. it will remove or distort them. So if one wants to ana-

Fig. 20. Data of AMMM (Aragats Multichannel Muon Monitor), filtered with MMF, Length of moving window is 60.

Fig. 21. Data of AMMM (Aragats Multichannel Muon Monitor), filtered with MMF 1, Length of moving window is 20.
lyze Solar Modulation Effects, it would be appropriate to use Relational Median Filter.

8. Conclusion

Filtering the data of multichannel cosmic ray particle detectors is crucial to investigate the effects of solar modulation over many years and of the diurnal variations on the cosmic ray intensity near Earth. During multiyear measurements, characteristics of detector undergo critical changes due to aging effects of sensors and of discrete elements of electronics. The big number of measuring channels allows introducing correction algorithms using stabilizing properties of the median of time series. Continuous storing and monitoring of the mean values of all channels along with their equalizing coefficient allows archiving of time-history of the behavior of all channels. Examining the relative behavior of channel means and coefficients during multiyear operation it became possible to distinguish the physical effects from instrumentation failures. See for example discussion in Bieber et al., 2007. Also our approach allows not only correction of mistakes due to hardware malfunction, but simple and efficient method of timely detection of non-stable channels or/and mistakes in data bases collecting time series from different remote detectors.

Acknowledgment

We thank colleagues from Cosmic Ray (CRD) division of Yerevan Physics Institute for useful discussions on CRD seminars and colleagues from NMDB for posting and correcting data in NMDB.

Authors are grateful to staff of Aragats and Nor Amberd research station for the uninterruptable operation of ASEC monitors. The work was supported by the grant RA610 and ISTC grant A-1554. The part of research leading to these results has received funding from the European Community’s Seventh Framework Programme (FP/2007-2013) under grant agreement no 213007.

References


