

Real time processing of neutron monitor data using the edge editor algorithm

Pavlos Paschalis and Helen Mavromichalaki*

Nuclear and Particle Physics Section, Physics Department, National and Kapodistrian University of Athens, Zografos 15784, Athens, Greece

*corresponding author: e-mail: emavromi@phys.uoa.gr

Received 14 June 2012 / Accepted 4 September 2012

ABSTRACT

The nucleonic component of the secondary cosmic rays is measured by the worldwide network of neutron monitors (NMs). In most cases, a NM station publishes the measured data in a real time basis in order to be available for instant use from the scientific community. The space weather centers and the online applications such as the ground level enhancement (GLE) alert make use of the online data and are highly dependent on their quality. However, the primary data in some cases are distorted due to unpredictable instrument variations. For this reason, the real time primary data processing of the measured data of a station is necessary. The general operational principle of the correction algorithms is the comparison between the different channels of a NM, taking advantage of the fact that a station hosts a number of identical detectors. Median editor, Median editor plus and Super editor are some of the correction algorithms that are being used with satisfactory results. In this work an alternative algorithm is proposed and analyzed. The new algorithm uses a statistical approach to define the distribution of the measurements and introduces an error index which is used for the correction of the measurements that deviate from this distribution.

Key words. Cosmic ray (galactic) - Cosmic ray (solar) - Algorithm - Metadata

1. Introduction

The cosmic rays which reach the Earth are of two different kinds. There are the Galactic cosmic rays that reach the solar system as highly isotropic and stable flux. They mainly consist of protons and heavier nuclei, which are accelerated to extremely high energies at stellar sources and by astrophysical phenomena. Moreover, during the high energetic periods of the solar activity, solar cosmic rays are produced. In addition, the solar activity and the solar wind modulate the galactic cosmic rays. The primary cosmic ray particles that enter the Earth's atmosphere interact with molecules, mainly oxygen and nitrogen, and produce a shower of lighter secondary particles.

The measurement and the observation of the secondary particles that reach the Earth's surface are of great scientific interest. They give us information about the solar activity and the structure of the universe in general. In addition, nowadays, through the Internet and the real time technology that allows the instant transportation of the information, the observation of the cosmic rays makes possible the monitoring and the prediction of the space weather (Lundstedt 2005) and of the related phenomena such as ground level enhancements (GLEs) (Souvatzoglou et al. 2009; Mavromichalaki et al. 2010). This is of high importance for the communications and the protection of the satellite systems.

The secondary particles, which are produced by the interactions of the primary particles in the terrestrial atmosphere, are measured by networks of ground-based particle detectors. The worldwide network of neutron monitors (NMs) has been monitoring the secondary cosmic ray flux at different locations, for more than 60 years, to detect short-term and long-term changes in intensity (Carmichael 1964; McDonald 2000; Simpson 2000). The last years, most of the NM stations have been sending their measured data in real time and in common format, to the neutron monitor database (NMDB; http://www.nmdb.eu). This is a very important project, since it allows instant access to the data of many NM stations worldwide in real time and permits the implementation of various online applications, such as space weather centers and GLE alert systems.

The success of the applications that use the real time measurements of the NMs highly depends on the quality of the data. It is obvious, that the announcement of erroneous data that do not correspond to the real intensity of the cosmic rays would produce false results for the applications mentioned above. For this reason, the good quality of the real time data is of great importance. However, problems sometimes arise. A NM, apart from the identical proportional counters, consists of a great number of electronic modules that are necessary for the data acquisition. The power supplies, the preamplifiers and amplifiers modules, the ADCs and the discriminating modules are some of the electronics that are necessary for the processing of the primary signal before it is converted to counting rate. Unfortunately, errors may occur during the modules' operation or their interaction with one another. The instrument variations of one or more modules can cause the distortion of the measured data for one or more counters of the NM. Moreover, the meteorological phenomena like the snow or the wind can cause, in a smaller degree, similar problems. In general, the instrument variations can be categorized into four categories, the abrupt spikes, the slow drift, the abrupt change of mean



Fig. 1. Uncorrected and corrected with Median editor data of the Athens NM channel 4 for February 2011.

with recovery and the abrupt change of mean without recovery (Belov et al. 1988; Chilingarian et al. 2009; Hovhannisyan & Chilingarian 2011).

Although the instrument variations happen rarely, they should be excluded before the data are sent to the NMDB. The real time transportation of the measured data requires that this procedure should be applied instantly. This task is handled by the primary data processing algorithms that are responsible for the purification of the measured data. The greatest difficulty that these algorithms have to overcome is that they should correct the data in real time, when only the past measurements of the NM are known. For this reason, the general operational principle of the correction algorithms is the comparison of the different channels of the NM, taking into account the past measurements. A change, abrupt or not, to the counting rate of a channel is considered as valid only if similar changes are noticed in the other channels of the NM. Based on this principle, a number of effective algorithms have been implemented. Median editor, Median editor plus and Super editor are the algorithms that are used for the filtering of primary data with successful results (Yanke et al. 2011; ftp://cr0.izmiran.ru/ HELP Station/EDITORs).

The Athens NM station consists of six NM-64 counters (Mavromichalaki et al. 2001; http://cosray.phys.uoa.gr). The 1-min data are sent every minute to the NMDB (http:// www.nmdb.eu) in a real time process. The algorithm used for the correction of these data is the Median editor. However, the Athens group aim at the implementation of new algorithms for the improvement of the data quality of the different NM stations around the world. Recently, a new method has been implemented based on an artificial neural network model (Paschalis et al. 2012). The results were very satisfactory and the method is currently being applied to the real time data in order to be evaluated in time. In this work, a second method based on a statistical model is proposed. Because of the methodology that the data filtering is performed with, the new method has been named "Edge Editor". A complete and thorough description of all the required steps is given in the next sections of the paper, while the results and the conclusion are given in the two last sections.

2. The primary data processing problem

In this section the problem that the primary data processing algorithms have to solve is discussed. The counting rate, that a NM channel measures, mainly depends on four parameters.

(i) The first parameter, N, is the actual incoming intensity of the cosmic rays. This parameter depends only on the

cosmic ray intensity that reaches the NM site and therefore, it is the same for all the channels.

- (ii) The second parameter, A_i , is related to the static characteristics of the detector and the electronics that support it. This parameter is different for each channel and has constant value. Even if the detectors of the NM are considered as identical, slight differences in their characteristics and/or in the supporting electronics exist and cause slight differences in factors A_i .
- (iii) The third parameter, σ_i , concerns the statistical variations of the cosmic ray intensity and of the detection procedure. Similarly to parameter A_i , this parameter should also be considered as different for each channel, since it depends on the static characteristics of the detector and the electronics that support it as well.
- (iv) The last parameter, δ_i , concerns any possible undesired instrument variations, such as voltage and amplifier variations that relate to a problematic behavior. This parameter is present rarely and distorts the measurement of the specific channel.

According to the above-mentioned four parameters, the counting rate that the channel "i" of the NM measures is

$$N_i = n_i \pm \sigma_i \pm \delta_i \tag{1}$$

where

$$n_i = N \times A_i. \tag{2}$$

For a constant cosmic ray flux, the channel "*i*" measures a counting rate that is equal to n_i in average and with statistical variations σ_i . The statistical variations are much smaller than the average counting rate ($\sigma_i \ll n_i$). On the other hand, the instrument variations, δ_i , are unpredictable and in the abrupt spikes cases they could be much greater than n_i .

The aim of the primary data processing is to subtract the instrument variations δ_i when they appear to one or more channels. In general, a correction algorithm should have three main characteristics: (a) it should be quick in order to be applied in the real time process, (b) it should filter effectively all the erroneous data and (c) it should leave the nonerroneous data unchanged. While the first characteristic is easily achieved, the other two appear to be more complicated. A drastic filtering of the erroneous data will also result in the distortion of the non-erroneous ones. The distortion appears as a suppression of the measured data statistics. This effect can be observed well in Figure 1, where the uncorrected and the corrected with Median editor data for the fourth channel of the Athens NM and for the time period of February 2011 are plotted. As it can be seen, the correction algorithm filters very effectively the abrupt spike but

it also reduces the standard deviation of the data even in the cases where no problematic behavior is present.

This effect is common in the correction algorithm's application and has to do with their common operational principle. As it has already been mentioned, since for a real time correction algorithm, at the time of its operation only the past measurements and the present ones are known, the correction can be performed only via the comparison between the different channels of the NM. Actually, a correction algorithm "forces" the real time measurement of each channel toward a value that the algorithm considers as most probable (mean value or median value for example) for the specific real time set of data. As a result, the standard deviation of the corrected data is reduced compared to the standard deviation of the uncorrected ones. In order to overcome this effect, the new algorithms are applied only to the data that exceed some preselected statistical criteria. For example, the Median editor plus, which is an improvement of the simple Median editor, is applied only when the efficiency of a channel exceeds the 3 sigma rule (Yanke et al. 2011).

In general, a primary data processing algorithm which operates effectively and meets all the desired characteristics should consist of the following parts:

- (i) An offline statistical analysis of the past measurements. The statistical analysis is required in order to determine the general characteristics of the measurements' pattern. The results of the offline analysis are static parameters that are used in the following part.
- (ii) A real time procedure that has the real time measurements of the channels as input and the respective corrected values as output. This procedure uses the conclusions and the parameters of the offline analysis in order to determine which of the channels is erroneous and needs correction. Following, the algorithm performs the correction of the erroneous channels.

In the next sections the necessary offline analysis and the real time procedure of the edge editor are discussed.

3. Comparison of the NM channels

A primary data processing algorithm is based on the comparison between the counting rates of the different NM channels. However, a direct comparison between the different channels is not possible. According to Eqs. (1) and (2) and due to the difference of A_i factors, each channel measures a different counting rate for the same cosmic ray incoming flux. In order to make possible the comparison between the different channels, the counting rate of each channel should be normalized by transforming its counting rate to a common measurement level.

In the case where the NM channels work correctly, the instrument variation δ_i in the Eq. (1) is set to zero, and the counting rate is

$$N_i = n_i \pm \sigma_i. \tag{3}$$

Considering two different channels "*i*" and "*j*", and using Eqs. (2) and (3), the ratio $\frac{N_i}{N_i}$ is

$$\frac{N_i}{N_j} = \frac{A_i}{A_j} \times \frac{\left(1 \pm \frac{\sigma_i}{N \times A_i}\right)}{\left(1 \pm \frac{\sigma_j}{N \times A_j}\right)}.$$
(4)

Since the statistical variations are much smaller than the mean counting rate ($\sigma_i \ll N \cdot A_i$), it becomes that

$$\frac{1}{\left(1\pm\frac{\sigma_j}{N\times A_j}\right)}\approx \left(1\mp\frac{\sigma_j}{N\times A_j}\right).$$
(5)

So, Eq. (4) takes the form

$$\frac{N_i}{N_j} = \frac{A_i}{A_j} \left(1 \pm \frac{\sigma_i}{N \times A_i} \right) \left(1 \mp \frac{\sigma_j}{N \times A_j} \right).$$
(6)

If only the first grade of orders is used, it becomes

$$\frac{N_i}{N_j} = \frac{A_i}{A_j} \left(1 \pm \frac{\sigma_i}{N \times A_i} \mp \frac{\sigma_j}{N \times A_j} \right).$$
(7)

Equation (7) can be written in the form:

$$\frac{N_i}{N_i} = R_{i,j} \pm \Sigma_{i,j} \tag{8}$$

where

$$R_{i,j} = \frac{A_i}{A_j} \text{ and } \Sigma_{i,j} = \frac{A_i}{A_j} \left(\pm \frac{\sigma_i}{N \times A_i} \mp \frac{\sigma_j}{N \times A_j} \right).$$
 (9)

Since the parameters A_i and A_j are constant, the $R_{i,j}$ is also constant. Also, since $\frac{\sigma_i}{N \times A_i} \ll 1$ and $\frac{\sigma_j}{N \times A_j} \ll 1$, it becomes that $|\Sigma_{i,j}| \ll R_{i,j}$. According to Eqs. (8) and (9), the value of the ratio $\frac{N_i}{N_j}$ has a mean value of $R_{i,j}$ and fluctuates around this value due to the statistical variation $\Sigma_{i,j}$.

This conclusion can be easily verified experimentally. Using the "1 min" Athens station raw data of February 2011, the time series of N_1 , N_2 , N_6 , and the corresponding time series of $\frac{N_2}{N_1}$, $\frac{N_6}{N_1}$ are shown in Figure 2. According to Eq. (3), the N_1 , N_2 and N_6 time series have the same pattern, however the three channels measure a different counting rate due to the differences in the A_1 , A_2 and the A_4 factors. On the other hand, the $\frac{N_2}{N_1}$ and the $\frac{N_6}{N_1}$ ratios fluctuate around a mean value, as expression (8) predicted. The abrupt spike in the $\frac{N_6}{N_1}$ plot reflects the abrupt spike in the N_4 measurements, where Eq. (3) cannot be used since the instrument variation is not equal to zero. The histogram of the ratio $\frac{N_2}{N_1}$ is shown in Figure 3. A Gaussian fit has been applied which gives a very high value of the coefficient of determination R^2 . This means that the ratios $\frac{N_i}{N_j}$ of the NM's channels can be approximated with a Gaussian distribution.

The determination of the $\frac{N_i}{N_j}$ distributions is of great importance for the edge editor algorithm. The counting rate N_i of the channel "*i*" can be transformed to the counting rate level of the channel "*j*" by dividing the N_i with the mean value $(R_{i,j})$ of the distribution $\frac{N_i}{N_j}$. The normalized counting rate of the channel "*i*" by using the channel "*j*" as a reference one is noted as N_i^j and by using Eqs. (2), (3) and (9) is equal to

$$N_i^j = \frac{n_i \pm \sigma_i}{R_{i,j}} \Rightarrow N_i^j = n_j \pm \sigma_i^j \tag{10}$$

where

$$n_j = A_j \times N$$
 and $\sigma_i^j = \frac{A_j}{A_i} \times \sigma_i$. (11)



Fig. 2. The time series N_1 , N_2 , N_6 (upper panel) and the $\frac{N_2}{N_1}$ and $\frac{N_6}{N_1}$ ratios (lower panel) of the Athens neutron monitor for February 2011.

Also, about the statistical variation σ_i^j , it comes out that:

$$\sigma_i \ll A_i \times N \Rightarrow \frac{A_j}{A_i} \times \sigma_i \ll A_j \times N \Rightarrow \sigma_i^j \ll n_j.$$
(12)

With this transformation, all the channels measure the same counting rate n_j and any difference is due to the statistical variations σ_i^j . Therefore, a comparison between the normalized counting rates of the channels is possible. Since the measured

quantity from the NM is the N_i^j , an estimation of the n_j can be made by calculating the average value of the normalized measurements N_i^j s over all of the channels:

$$\langle N_i^j \rangle_i = \frac{\sum_i N_i^j}{\text{number of channels}}$$

= $n_j + \frac{\sum_i (\pm \sigma_i^j)}{\text{number of channels}}$. (13)



Fig. 3. Histogram of $\frac{N_2}{N_1}$ of the Athens neutron monitor for February 2011.

Since the σ_i^j s are statistical variations, the $\frac{\sum_i (\pm \sigma_i^j)}{n_{\text{umber of channels}}}$ tends to zero as the number of channels is increasing and the $\langle N_i^j \rangle_i$ becomes equal to n_j . However, even for smaller number of channels, the $\langle N_i^j \rangle_i$ can be used as estimation for the n_j :

$$\langle N_i^{\prime} \rangle_i \approx n_j.$$
 (14)

4. An offline analysis for the edge editor

The offline analysis for the edge editor aims at the determination of the $\frac{N_i}{N_j}$ distributions $(R_{i,j} \text{ and } \Sigma_{i,j})$ and at the study of the statistical variations σ_i^j . These are the parameters that will be used in the real time part of the edge editor. This study is necessary to be performed for time periods long enough in order to have a good statistical sample. However, before the calculations, a filtering of the data that are used is required. The reason is that in the past data of the NM, many problems may have happened, such as cases where some or all of the channels have null values or cases where instrument variations distort the measurements. In order to perform a correct and accurate statistical analysis, these cases should be excluded from the past data and only the ones where all the counters work correctly should be used.

The rejection of the measurements in the cases where one or more channels are equal to zero is easy to be done, while the rejection of the measurements where instrument variations are present is more complicated. The procedure that should be followed takes advantage of the fact that an instrument variation to the channel "*i*" or "*j*" reflects the ratio $\frac{N_i}{N_j}$. This effect has already been noticed in Figure 2, where the abrupt spike of the channel 6 distorts the time series of the $\frac{N_6}{N_1}$. Thus, having selected a channel "j" as a reference one, the data that should be considered as valid are only the ones whose all $\frac{N_i}{N_i}$ values are within a trust interval taking into account the respective ratio distributions. If even one ratio is out of the defined trust interval, an instrument variation exists and the corresponding set of data should be rejected. For a proper filtering, the 3σ trust interval is used since it statistically contains the 99.7% of the $\frac{N_i}{N_i}$ values. The use of a 4σ trust interval is also possible but the hazard of using a measurement that is distorted from an instrument variation increases.

According to this filtering procedure, the calculation of the $\frac{N_i}{N_i}$ distributions is shown in the flowchart of Figure 4. The $R_{i,j}$

and the $\Sigma_{i,j}$ of each distribution are calculated by using a smoothing procedure. According to this method, the $R_{i,j}$ and the $\Sigma_{i,j}$ are initially calculated using all the available data and then they are updated iteratively in small steps by using only the data that are within the 3σ trust interval, based on the $R_{i,j}$ and the $\Sigma_{i,j}$ of the previous iteration. The procedure ends when both the mean value and the sigma converge to constant values. With this method, all the instrument variations are rejected and the $R_{i,j}$ and the $\Sigma_{i,j}$ are calculated accurately. On the other hand, the study of the statistical variations σ_i^{j} refers to the calculation of the statistical variations σ_i^{j} for each channel "*i*" in conjunction with the n_j (Eqs. (10) and (11)). The estimation of the n_j for each dataset is done by using Eq. (14). This study requires the use of valid N_i^j data which are generated by the procedure described in the flowchart of Figure 5.

As it has already been mentioned, the Athens NMs consist of six channels. All the data are stored in a mySQL database which makes easy the data processing. In the cases where the data processing is not possible only with SQL queries, for example in the iterative method of the $R_{i,j}$ and the $\Sigma_{i,j}$ calculation, PHP scripts are used. The data that are used are from the years 2007 to 2011. Finally, for the offline analysis, the first channel is used as a reference, which is a random choice.

According to the analysis above, the characteristics of the five ratio distributions are given in Table 1. About the statistical variations, Figure 6 shows the histogram of the N_2^{1} for the case the n_1 is equal to 600 (Eq. (10)). A Gaussian fit has been applied and the high value of the R^2 shows that the Gaussian distribution approximation can be used. The conclusion that the distribution of the N_i^j , for a specific n_j value, can be approximated by a Gaussian distribution, is the base for the operation of the edge editor as it will be described in the next section. In Figure 7, the σ_i^j vs. n_j plot is presented for the six channels of the Athens station, if the channel 1 is used as reference. It is noticed that there is a linear dependence between the two variables, so a linear regression can be applied in order to calculate their relation. The result of the linear regression and the corresponding correlation coefficient is given in Table 2.

5. Real time procedure of the edge editor

The online operation of the edge editor aims at the determination of the erroneous channels and the correction of them in real time. This is performed via comparison of the channels' measurements and the use of the offline analysis's conclusions. The online part of the algorithm can be separated in four main parts:

5.1. The normalization of the channels

As it has already been analyzed in the third section of the paper, each channel of the NM measures a counting rate noted as N_i . In order for the counting rates to be comparable, they should be normalized by using a channel "*j*" as reference and be transformed to N_i^j . This is performed by dividing each measurement with the $R_{i,j}$ as expressed in Eq. (10).

5.2. The determination of the n_i

After the generation of the N_i^{j} s, the determination of the n_j (Eq. (10)) is very critical for the correct operation of the algorithm. During the offline analysis, the n_j was estimated, according to Eq. (14), as the average value of the N_i^{j} s

J. Space Weather Space Clim. 2 (2012) A15



Fig. 4. Flowchart of the $R_{i,j}$ and the $\Sigma_{i,j}$ calculations.

 $(\langle N_i^j \rangle_i)$. This is possible since the N_i^j s had been filtered from possible instrument variations and were valid data. However, this is not a correct approach for the real time handling where instrument variations may exist.

In general, the determination of the mean value in a set of measurements is an easy procedure when the set is large. The averaging of the measurements after the outliers is rejected by statistical tests, as the *Q*-test, is sufficient (Rorabacher 1991). However, the determination of the mean value gets more difficult as the set of measurements gets smaller. This case is met in many medicine and chemistry applications. It is also met in the NMs due to the small number of channels. The determination of the mean value in a small set of numbers is an issue that is studied by the small number statistics (Dean & Dixon 1951). The *t*-distribution, for example, is used for the estimation of the mean value when the standard deviation of the population where the numbers belong to is unknown (Senn & Richardson 1994).

A secure decision for the estimation of the mean value, in a small set of measurements, is the median value. Although the use of the median value is secure, it has the disadvantage that only 1 or two values are used and all the others are rejected from the estimation of the mean value, even in the case they are statistically correct. In the edge editor a new approach is presented which uses the conclusion that for a specific n_j the distribution of the N_i^j can be approximated by a Gaussian distribution. According to this approach, the mean value is estimated iteratively by using a weight factor for each N_i^j . The value of each weight depends on the distance of the respective N_i^j from an estimated mean value that is calculated by using the weights from the previous iteration. The initial mean value that is used at the beginning of the procedure is the median value of the N_i^j s and it is adapted as the iterations proceed.

The weights are calculated by using a weight function that gives a value equal to one when the N_i^j is equal to the estimated mean value and a value equal to zero when the distance between them tends to infinity. Such a function is the exponential part of the Gaussian probability density functions:

$$w(x) = e^{-\frac{(x-\text{mean})^2}{2\times\sigma^2}}.$$
 (15)

As σ in the equation above, the σ_i^j can be used, calculated from the equations of Table 2 and for n_j equal to the estimated mean value. Thus, the weights have the following form:

$$w_i(N_i^j) = e^{-\frac{(N_i^j - \text{estimated} n_j)^2}{2 \times (\sigma_i^j)^2}}.$$
 (16)

P. Paschalis and H. Mavromichalaki: Real time processing of neutron monitor data



Fig. 5. Flowchart of the generation of valid N_i^j s.

Table 1. The mean value *R* and the standard deviation Σ of the ratio distributions, using the channel 1 as α reference, are given. Data for the period 2007–2011 are used.

N_2/N_1		N ₃ /N ₁		N_4/N_1		N_{5}/N_{1}		N ₆ /N ₁	
<i>R</i> _{2,1}	$\Sigma_{2,1}$	<i>R</i> _{3,1}	$\Sigma_{3,1}$	$R_{4,1}$	$\Sigma_{4,1}$	<i>R</i> _{5,1}	$\Sigma_{5,1}$	$R_{6,1}$	$\Sigma_{6,1}$
1.019	0.071	0.889	0.065	0.882	0.057	0.846	0.062	0.893	0.057

The procedure described above for the case of the six channels of Athens NM is shown in the flowchart of Figure 8. This method actually positions the mean value in the place where the density of the measurements is maximum. The values that are too far from the others (which mean they probably are instrument variations) have a small weight which reflects its small affection on the mean value, while the rest of the values have a greater weight and affect the mean value more significantly. The result of each iteration in the case where two out of the six channels have erroneous data is given in Table 3. It is noticed that the estimated mean value begins by using all the measurements and gradually converges to its final value.

5.3. Determination of the erroneous channels and correction

The name of the edge editor is due to the procedure that the algorithm follows in order to discriminate the erroneous channels and correct them. After estimating the mean value

Table 2. The $\sigma_i^1 = f(n_1)$ equations, using the channel 1 as a reference using the data from 2007 to 2011.

Channel	$\sigma_i^j = f(n_j)$	Correlation coefficient
1	$\sigma_1^{\ 1} = 0.0291 \times n_1 + 8.1129$	0.986
2	$\sigma_2^{\ 1} = 0.0260 \times n_1 + 10.6687$	0.974
3	$\sigma_3^{\ 1} = 0.0290 \times n_1 + 11.7736$	0.975
4	$\sigma_4{}^1 = 0.0302 \times n_1 + 10.5180$	0.981
5	$\sigma_5^{\ 1} = 0.0293 \times n_1 + 12.1078$	0.973
6	$\sigma_6^{-1} = 0.0300 \times n_1 + 10.8838$	0.982

 n_j , the algorithm determines which channels are erroneous according to a statistical criterion. According to Figure 6 and to the equations of Table 2, the measurements of each channel follow a Gaussian distribution with sigma equal to σ_i^j .

Using the 3 sigma rule, the algorithm checks whether the N_i^{j} is in the 3 sigma trust interval:

$$\left|N_{i}^{j}-n_{j}\right|<3\times\sigma_{i}^{j}\left(n_{j}\right) \tag{17}$$

If the criterion given above is fulfilled, the algorithm considers that the measurement of the channel is valid and no action is taken. Otherwise the measurement of the channel is imposed to correction.

For the erroneous channel, an error index is introduced for the quantification of the channel's error. The error index is zero when the measurement of the channels is on the limit of the validation criterion $(|N_i^j - n_j| = 3 \times \sigma_i^j)$ and tends to zero, when the term $|N_i^j - n_j|$ tends to infinity. The algorithm uses the following function for error index:

$$E_{i}(N_{i}^{j}) = \begin{cases} 0, & |N_{i}^{j} - n_{j}| < 3 \times \sigma_{i}^{j} \\ 1 - \frac{1}{N_{i}^{j} - (n_{j} + 3 \cdot \sigma_{i}^{j})}, & |N_{i}^{j} - n_{j}| \ge 3 \times \sigma_{i}^{j} \text{ and } N_{i}^{j} > n_{j} \\ \frac{1}{e^{-\sqrt{2 \cdot \sigma^{2}}}}, & |N_{i}^{j} - n_{j}| \ge 3 \times \sigma_{i}^{j} \text{ and } N_{i}^{j} < n_{j} \end{cases}$$

$$(18)$$

The correction of the erroneous channels is performed by using the following procedure. The measurements that are outside the



Fig. 6. Histogram of N_2^{1} for $n_1 = 600$, using the data of 2007–2011.

trust interval are positioned inside it. The greater the error index is, the closer to the n_j the corrected value is positioned. The logic is that a measurement with a great error index is more possible to be an instrument variation and therefore the corrected value should be closer to the n_j . On the contrary, a measurement with a small error index is more possible to be a statistical variation which just exceeds the 3 sigma rule, so the corrected value should not differ a lot from the uncorrected one and be



Fig. 7. The σ_i^{1} vs. n_1 values for the six channels of the Athens NM using data for the years 2007–2011.

P. Paschalis and H. Mavromichalaki: Real time processing of neutron monitor data



Fig. 8. Flowchart of the real time estimation of n_i using a Weighted Mean Method.

Table 3. An example of the n_i estimation using a Weighted Mean Method.

	$N_1^1 = 532$	$N_2^1 = 571$	$N_3^{1} = 604$	$N_4^{\ 1} = 800$	$N_5^{1} = 515$	$N_6^{1} = 1000$	
Iteration	w_1	<i>w</i> ₂	<i>w</i> ₃	w_4	<i>w</i> ₅	w ₆	n_1
0	_	_	_	_	_	_	587.5
1	0.088613	0.816896	0.84875	5.28E-13	0.047036	3.45E-46	583.1684
2	0.124842	0.894978	0.768212	1.25E-13	0.065479	1.42E-47	580.0722
3	0.157272	0.939826	0.704627	4.34E-14	0.082124	1.38E-48	577.6461
4	0.186888	0.967084	0.65261	1.86E-14	0.097493	2.17E-49	575.6153
5	0.214678	0.983923	0.608273	9.04E-15	0.11209	4.51E-50	573.8263
10	0.35321	0.990734	0.434166	4.56E-16	0.18809	7.28E-53	565.9238
15	0.563685	0.876575	0.259475	1.09E-17	0.31807	2.72E-56	555.5213
20	0.905913	0.519077	0.083023	1.35E-20	0.614771	2.7E-62	539.4292
25	0.999894	0.288305	0.031207	1.01E-22	0.822055	1.4E-66	531.7739
30	0.999184	0.263895	0.027225	5.33E-23	0.845172	3.87E-67	531.0092
35	0.999032	0.262338	0.026979	5.11E-23	0.84665	3.56E-67	530.9605

positioned near the edge of the Gaussian distribution. According to this logic, the corrected measurements, noted as C_i^j , are

$$C_{i}^{j} = \begin{cases} N_{i}^{j}, & |N_{i}^{j} - n_{j}| < 3 \times \sigma_{i}^{j} \\ n_{j} + 3 \times \sigma_{i}^{j} \cdot (1 - E_{i}), & |N_{i}^{j} - n_{j}| \ge 3 \times \sigma_{i}^{j} \text{ and } N_{i}^{j} > n_{j} \\ n_{j} + 3 \times \sigma_{i}^{j} \cdot (E_{i} - 1), & |N_{i}^{j} - n_{j}| \ge 3 \times \sigma_{i}^{j} \text{ and } N_{i}^{j} < n_{j} \end{cases}$$

$$(19)$$

The diagram of the correction procedure is shown in Figure 9.

5.4. The de-normalization of the channels

This is the inverse procedure of the first part. In order for the statistical comparison of the measurements to be possible, the measurements have been normalized to the counting rate of J. Space Weather Space Clim. 2 (2012) A15



Fig. 9. The edge editor algorithm in a real time operation.



Fig. 10. Uncorrected (black line) and corrected (red line) data of the Athens NM for March 2012.

the channels "*j*". The corrected values should be de-normalized in order for the measurements to return to the original counting rate of each channel. This is performed by multiplying the corrected values with the $R_{i,j}$ (Eq. (10)).

6. Results and conclusion

From all the above-mentioned analysis given in detail, it is concluded that the application of a primary data processing algorithm requires a statistical analysis of the past data which leads to the

P. Paschalis and H. Mavromichalaki: Real time processing of neutron monitor data



Fig. 11. Uncorrected (black line) and corrected (red line) data of the Athens NM for August 2011.

determination of a number of parameters that are used by the online part of the algorithm. In the case of the newly proposed edge editor after a channel "*j*" of the NM is selected as a reference one, the necessary parameters are the mean values of the $\frac{N_i}{N_i}$ ratios (Table 1) and the equations $\sigma_i^j = f(n_j)$ (Table 2).

The real time application of the edge editor on the data of the Athens cosmic ray station for the time period of March 2012 is shown in Figure 10. The black line presents the uncorrected data, while the red one presents the corrected data. It can be noticed that the sixth channel of the NM shows some abrupt spikes. The edge editor successfully filters all the spikes of this channel while the rest of the data remain almost unchanged. The measurements of the other channels, that work correctly, are not affected. The effectiveness of the algorithm can be verified by its application to the data of August 2011 that is presented in Figure 11. In this case, the fourth and the sixth channels both present some abrupt spikes which the algorithm removes as well. The rest of the data are not affected. The channels 2, 3 and 5 are not presented since they behave in the same pattern as channel 1. Finally, an interesting example can be found in the data of April 2010. In this case the sixth channel presents a very unstable behavior. Apart from some abrupt spikes, a quick abrupt change of mean with recovery is noticed. Also, there is a long time period between the 16th and the 24th of April where the channel measures 1 imp/min. As it can be seen, the correction algorithm successfully filters all these heavy erroneous cases while the rest of the data remain almost unchanged for all the counters. The measurements of the channel 1 are presented for comparison reasons. The channels 2-5 present the same pattern as the channel 1.

According to the results above, it is concluded that the edge editor is very effective in the real time correction of the NM's data. The three characteristics, mentioned in Section 2, which a



Fig. 12. Uncorrected (black line) and corrected (red line) data of the Athens NM for April 2010.

correction algorithm should have, seem to be fulfilled. On a powerful computer, the algorithm filters the 1 min data of more than 1 month in less than a minute. Also, it corrects the erroneous data caused by instrument variations and leaves the rest of the data almost unchanged. In order to verify the second conclusion, the correlation coefficient between the uncorrected and the corrected data is calculated, for the time period of February 2012, where all the channels work correctly without any erroneous behavior. The correlation coefficient for each channel is given in Table 4. For comparison reasons, the respective value for the Artificial Neural Network algorithm and for the Median editor (used currently by the station) is also given

Table 4. Correlation coefficient of the uncorrected vs. corrected data for three different correction algorithms for February 2012.

_		Correlation coefficient	
Channel	Edge editor	Artificial neural networks	Median editor
1	0.9893	0.9477	0.8028
2	0.9924	0.9486	0.8017
3	0.9924	0.9485	0.7659
4	0.9924	0.9612	0.7723
5	0.9930	0.9327	0.7590
6	0.9914	0.9367	0.7683

(Paschalis et al. 2012). The high values for the edge editor mean that the corrected data are very close to the uncorrected ones.

To sum up, the algorithm that has been presented in this work shows very satisfactory results for the correction of the NM data. Its application to the Athens NM station as well as to the other stations of the worldwide network will improve the data quality that is being sent to the High Resolution Neutron Monitor Database-NMDB. As a result, the online scientific services and applications that use the real time data would be operated more reliably and contribute to the Space Weather monitoring.

Acknowledgements. The authors acknowledge the Special Research Account of Athens University for supporting the cosmic ray research. The authors also acknowledge Dr. Pantelis Papachristou for useful discussions.

References

Belov, A.V., Ya.L. Blokh, E.G. Klepach, and V.G. Yanke, Processing of cosmic ray station data: algorithm, computer program and realization, *Kosmicheskie Luchi*, 25, 113–134, [in Russian], 1988. Carmichael, H., Cosmic rays. *IQSY Instruction Manual No.* 7, 1964.

- Chilingarian, A., A. Hovhannisyan, and B. Mailyan, Median filtering algorithms for multichannel detectors, *Proc. 31st ICRC (Lodz)*, icrc0677, http://icrc2009.uni.lodz.pl/proc/pdf/icrc0677.pdf, 2009.
- Dean, R.B., and W.J. Dixon, Simplified statistics for small numbers of observations, *Anal. Chem.*, 23, 636–638, 1951.
- Hovhannisyan, A., and A. Chilingarian, Median filtering algorithms for multichannel detectors, *Adv. Space Res.*, 47, 1544–1557, 2011.
- Lundstedt, H., Progress in space weather predictions and applications, *Adv. Space Res.*, **36**, 2516–2523, 2005.
- Mavromichalaki, H., C. Sarlanis, G. Souvatzoglou, S. Tatsis, A. Belov, E. Eroshenko, V. Yanke, and A. Pchelkin, Athens neutron monitor and its aspects in the cosmic-ray variations studies, *Proc. 27th ICRC 2001 (Hamburg)*, **10**, 4099–4103, 2001.
- Mavromichalaki, H., G. Souvatzoglou, Ch. Sarlanis, G. Mariatos, A. Papaioannou, A. Belov, E. Eroshenko, V. Yanke, and NMDB Team, Implementation of the ground level enhancement alert software at NMDB database, *New Astronomy*, **15**, 744–748, 2010.
- McDonald, F.B., Integration of neutron monitor data with spacecraft observations: a historical perspective, *Space Sci. Rev.*, 93, 239–258, 2000.
- Paschalis, P., C. Sarlanis, and H. Mavromichalaki, Artificial neural networks approach of cosmic ray primary data processing, *Sol. Phys.*, 2012, in press.
- Rorabacher, D.B., Statistical treatment for rejection of deviant values: critical values of Dixon's "Q" parameter and related subrange ratios at the 95% confidence level, *Anal. Chem.*, 63 (2), 139–146, 1991.
- Senn, S., and W. Richardson, The first t-test, Stat. Med., 13 (8), 785–803, 1994.
- Simpson, J.A., The cosmic ray nucleonic component: the invention and scientific uses of the neutron monitor, *Space Sci. Rev.*, 93, 1–20, 2000.
- Souvatzoglou, G., H. Mavromichalaki, C. Sarlanis, G. Mariatos, A. Belov, E. Eroshenko, and V. Yanke, Real time GLE ALERT in the ANMODAP center for December 13, 2006, *Adv. Space Res.*, 43, 728–734, 2009.
- Yanke, V., A. Belov, E. Klepach, E. Eroshenko, N. Nikolaevsky, O. Kryakunova, C. Sarlanis, H. Mavromichalaki, and M. Gerontidou, Primary processing of multichannel cosmic ray detectors, *Proc. 32nd ICRC (Beijing)*, **11**, 450–453, 2011.