Solar Phys (2013) 282:303–318
DOI 10.1007/s11207-012-0125-3

Artificial Neural Network Approach of Cosmic Ray Primary Data Processing

P. Paschalis · C. Sarlanis · H. Mavromichalaki

Received: 23 November 2011 / Accepted: 10 September 2012 / Published online: 16 October 2012

Abstract One of the most critical points in the detection of cosmic rays by neutron monitors is the correction of the raw data. The data that a detector measures may be distorted by a variety of reasons and the subtraction of these distortions is a prerequisite for processing them further. The final aim of these corrections is to keep only the fluctuations related to the real cosmic-ray intensity. To achieve this, we analyze data from identical neutron monitor detectors which provide a configuration with the ability to exclude the distortions by comparing the counting rate of each detector. Based on this method, a number of effective algorithms have been developed: Median Editor, Median Editor Plus, and Super Editor are some of the algorithms that are being used in the neutron monitor data processing with satisfactory results. In this work, a new approach for the correction of the neutron monitor primary data with a completely different method, based on the use of artificial neural networks, is proposed. A comparison of this method with the algorithms mentioned previously is also presented.

Keywords Cosmic rays · Data processing · Neutron monitors · Neural networks

1. Introduction

Galactic cosmic rays, mostly protons and heavier fully stripped nuclei, are accelerated in our galaxy by shock waves originating from supernova explosions and from other energetic stellar sources. After traveling millions of years in our galaxy, they arrive in the solar system as highly isotropic and stable flux. On the other hand, solar cosmic rays are produced during high-energy events at the Sun. The solar cosmic-ray particles can travel away from the Sun

P. Paschalis · C. Sarlanis · H. Mavromichalaki (✉)
Nuclear and Particle Physics Section, Physics Department, National and Kapodistrian University of Athens, Zografos, 15784 Athens, Greece
e-mail: emavromi@phys.uoa.gr

P. Paschalis
e-mail: ppaschalis@phys.uoa.gr

along the open magnetic field lines and produce secondary cosmic-ray particles by their interaction with the atmosphere of the Earth.

The worldwide network of ground-based particle detectors measures time series of the secondary particles produced by the interactions of the primary cosmic-ray particles in the terrestrial atmosphere. The neutron monitors (NMs) located at different latitudes, longitudes, and altitudes have been monitoring the secondary cosmic-ray flux for more than 60 years to detect abrupt changes in intensity and/or long-term trends. The geomagnetic field and the atmospheric depth are very different among the detector locations and each detector has to be treated individually, in order for its data to be utilized (Carmichael, 1964; McDonald, 2000; Simpson, 2000).

The long-term operation of the proportional counters of the neutron monitors and the data acquisition electronics may suffer from several errors (spurious peaks due to electronics failures or lightning, drift of the count rate by degradation of detectors, gaps in the time series due to electricity failures, *etc.*) which should be detected and smoothed before presenting the data to the international databases (http://www.nmdb.eu). The remedy for these failures is comparing the data from different channels, taking advantage of the fact that a neutron monitor consists of many identical counters. If one channel of the NM is defective it can be localized and repaired or excluded. However, some neutron monitors are operated at remote sites, and sometimes the replacement of the failed equipment is not immediately possible. What is needed is an autonomous system for the correction of the raw data based on the abundance of the measuring channels.

For the analysis of the cosmic-ray measurements, it is necessary to "purify" (correct, filter, and smooth) the raw data. The quality of the data, the different types of instrumental variation, their possible causes, and the methods of their correction has been already analyzed (Belov *et al.*, 1988; Chilingarian, Hovhannisyan, and Mailyan, 2009; Hovhannisyan and Chilingarian, 2011). Filtering algorithms are usually based on the comparisons of data from identical measurement channels. Currently, for the online comparison of similar channels, two methods are proposed: the median method and the method of the ratio logarithms which has clear advantages in comparison with the ordinary method of ratios (Yanke *et al.*, 2011, ftp://cr0.izmiran.ru/HELP_Station/EDITORs).

The algorithms based on median filtering are currently widely used in pattern recognition, in multimedia technologies, and in scientific applications. For instance, to maximize data output from single-shot astronomical images, the rejection of the cosmic-ray background is performed by median algorithms (Farage and Pimbblet, 2005). Moreover, Van Dokkum (2001), using conventional algorithms noticed that the cosmic rays in single images or spectra can be removed by a variant of Laplacian edge detection. The procedure is robust, and requires very few user-defined parameters. The method rejects the cosmic-ray hits of arbitrary size and distinguishes under-sampled point sources from cosmic rays with a high confidence.

The Athens Neutron Monitor Station located at the Physics Department of the Athens University consists of six proportional counters of Super 6NM-64 type (Mavromichalaki *et al.*, 2001, http://cosray.phys.uoa.gr). Currently the correction of the data is performed by the application of the Median Editor (Yanke *et al.*, 2011). In this method, the ratio of the last counting rate compared to a previous one is evaluated for each counter. This counting rate then is reproduced by using the median value of all the ratios.

In this study a new approach using an artificial neural network method (referred to as ANN from now on) is proposed for the primary data processing of the neutron monitor measurements. The ANN is a well-known computational tool that can be used in many applications in a variety of fields. Even in the field of the cosmic-ray research, the ANN has already been used for different purposes, other than the primary data processing. For example, an event-by-event study of the mass identification in high-energy cosmic rays was carried out with simulated data and based on the neural network method (Riggi *et al.*, 2007). Extensive air showers were simulated with the CONEX code, using the hadronic model QGSJET-II-3. The effectiveness of the method in recognizing the mass of the primary particles was tested making use of the parameters extracted from the simulated longitudinal profiles. They showed that the designed neural network is able to discriminate, with high identification efficiency and purity, between proton- and iron-induced showers.

General information regarding the ANN can be found in several references over the Internet (http://en.wikipedia.org/wiki/Artificial_neural_network). In this work the ANN method is applied for the first time to the cosmic-ray data of the Athens Neutron Monitor Station in the phase of the primary data processing. The obtained results are compared with the ones achieved by the Median Editor algorithm which is currently used in the Athens Cosmic-ray Station. The most important general points of the ANN concept are described in Section 2, while the data processing of the neutron monitors is described in Section 3. The ANN approach on neutron monitor data and the comparison with the Median Editor results are given in Sections 4 and 5, respectively. The conclusions from all this analysis are presented in Section 6.

2. General Principles of Artificial Neural Networks

An ANN is composed of two or more layers and each one of them is composed of nodes named "neurons". The nodes of each layer are connected with the nodes of the next layer through connections named "synapses". Each synapse is related to a weight factor which acts as a multiplier factor when a value is transferred through it. At each neuron, the input values are summarized and the result is processed by an activation function (usually a sigmoid function, http://en.wikipedia.org/wiki/Sigmoid_function). The output of the neuron is transferred to the next layer and the process continues until the last (output) layer. A general structure of an ANN is given in Figure 1. It is obvious that an ANN consists of at least two layers, an input and an output layer. Apart from them, it is possible to have inner layers,

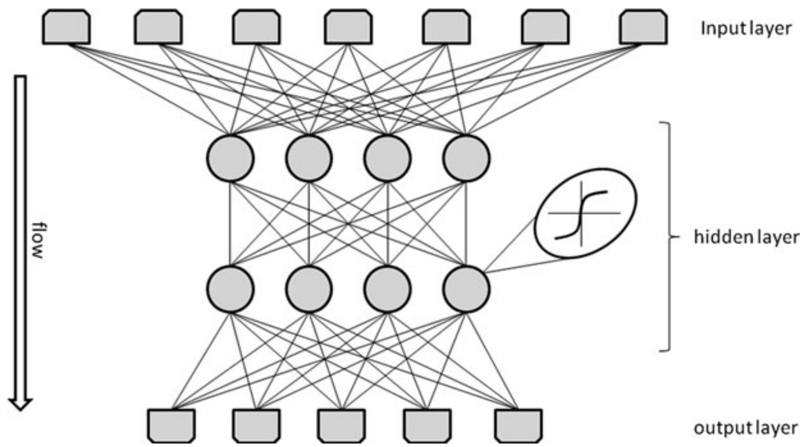


Figure 1 Artificial neural network structure.

which are called “hidden layers”. The number of layers and the number of nodes in each layer compose the architecture of the network.

When the ANN is firstly created, synapses are assigned with default or random values. In order to produce a correct output for a specific input of the network, synapses should have correct values. This is achieved through a learning procedure, during which the network is fed with training data (usually simulation data) and is forced to output the desired result. The network compares the actual output with the desired one and adapts the weights in individual small steps. The sample that the ANN is fed is called training sample and contains data of a variety of cases since the aim is to achieve a general behavior after training. To feed the network with all the sample of the training data is called an epoch of training. The learning procedure takes several epochs before it is considered completed. The progress of the training is supervised after one or more epochs, by a test sample. During this procedure, the network is fed with the test data and the actual output is compared to the output of the test sample. It should be emphasized that the training procedure takes place only once, at the beginning of the ANN set up. After that the ANN is ready to be used and gives a quick response for a defined input.

The critical points that should be taken into account when using an ANN are the following. The first point concerns the architecture that should have the appropriate complexity in order for free parameters to exist. A less complicated architecture will not solve the problem efficiently, while a more complicated one will obstruct the ANN from getting trained. In other words, when designing an ANN, a balance of the complexity of the layers and neurons has to be found. However, the most critical point when using an ANN is the training sample, since it represents the desired behavior of the network. Obviously, a wrong behavior of the network after its training should be firstly attributed to the training sample. The creation of the training sample should follow a thorough analysis of the problem and of the desired behavior of the network. Finally, one more thing that should be emphasized about the training is the required number of epochs. The epochs should be as many as needed and no more, otherwise the network will learn to react perfectly only with the training sample and will lose the ability to handle correctly the real data. In this case the network would be considered as “over-trained”.

Apart from these critical points, there are many points that should be taken into account when designing and training an ANN. These points concern the choice of the activation function, the momentum of weight adaptation, the cutoff of some synapses, and the training algorithm. The discussion of these points is not presented in this paper. They are just mentioned to highlight that the behavior of an ANN depends on and can be optimized by changing all these parameters.

3. Primary Data Processing of Neutron Monitors

In this section the problem of primary data processing is discussed. The detection of the nucleonic component of the cosmic rays is performed by using neutron monitors, as the one of the Athens Cosmic Ray Station. The counting rate that each counter measures mainly depends on four items:

i) the actual incoming intensity of the cosmic rays,
ii) static characteristics of the detector and the electronics that support it,
iii) statistical variations that exist because the procedure of neutron detection in each counter is of a statistical nature (Carmichael, 1964; Simpson, 2000), and

iv) undesired instrument variations such as voltage and amplifier variations that may lead to a problematic behavior (Belov *et al.*, 1988; Chilingarian, Hovhannisyan, and Mailyan, 2009; Hovhannisyan and Chilingarian, 2011).

The Athens Cosmic Ray Station consists of six NM64 counters. Let us assume that the incoming rate of neutrons is N. Ideally, the counting rate that each monitor measures should be N, as well. However, due to the four items mentioned above, the detector i measures

$$N_i = N \cdot A_i \pm \sigma_i \pm \delta_i, \tag{1}$$

where N corresponds to the actual cosmic-ray intensity (item i), A_i corresponds to the detector calibration (item ii), $\pm\sigma_i$ corresponds to statistical variations (item iii), and $\pm\delta_i$ corresponds to undesired instrument variations (item iv).

The A_i factor is related to static characteristics of the detector and has no dependency on time (or in the worst case scenario, a slight dependency on it), if we think of long periods of time (years). It should be emphasized that even if the detectors of the neutron monitor are considered as identical, slight differences in their characteristics exist and cause slight differences in factors A_i.

The parameters that appear in Equation (1) can be identified by comparing the time series of each counter for a specific period of time. The 1 min measurements of each detector of the Athens NM for the time period from 1 August 2011 to 15 September 2011 are presented in Figure 2. In these diagrams, it can be noticed that for a period of 45 days, the curve of each counter has exactly the same profile. The different counting rate of each counter is due to the factors A_i. The width of the curves is associated with the statistical variations, expressed by $+\sigma_i$. The source of some spikes of counters 4 and 6 is the factor $\pm\delta_i$. It is obvious that these peaks are sporadic, and correspond to a problematic behavior and distort the data.

If it is assumed that there is not any problematic behavior of the instruments ($\delta_i = 0$ and $\delta_j = 0$) and using Equation (1), the ratio N_i/N_j of counters i and j, $i \neq j$ becomes

$$\frac{N_i}{N_j} = \frac{N \cdot A_i \pm \sigma_i}{N \cdot A_j \pm \sigma_j} = \frac{N \cdot A_i \cdot (1 \pm \frac{\sigma_i}{N \cdot A_i})}{N \cdot A_j \cdot (1 \pm \frac{\sigma_j}{N \cdot A_j})} = \frac{A_i}{A_j} \cdot \frac{1 \pm \frac{\sigma_i}{N \cdot A_i}}{1 \pm \frac{\sigma_j}{N \cdot A_j}}. \tag{2}$$

But since the statistical variations are much smaller than the mean counting rate ($\sigma_j \ll N \cdot A_j$), we assume

$$\frac{1}{1 \pm \frac{\sigma_j}{N \cdot A_j}} \approx 1 \mp \frac{\sigma_j}{N \cdot A_j}. \tag{3}$$

Therefore, Equation (2) becomes

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

If we keep only the first-order terms, then

$$\frac{N_i}{N_j} = \frac{A_i}{A_j} \cdot \left(1 \pm \frac{\sigma_i}{N \cdot A_i} \mp \frac{\sigma_j}{N \cdot A_j}\right). \tag{5}$$

Equation (5) can be written in the form

$$\frac{N_i}{N_j} = \frac{A_i}{A_j} \pm \frac{A_i}{A_j} \cdot \Sigma_{i,j} \tag{6}$$

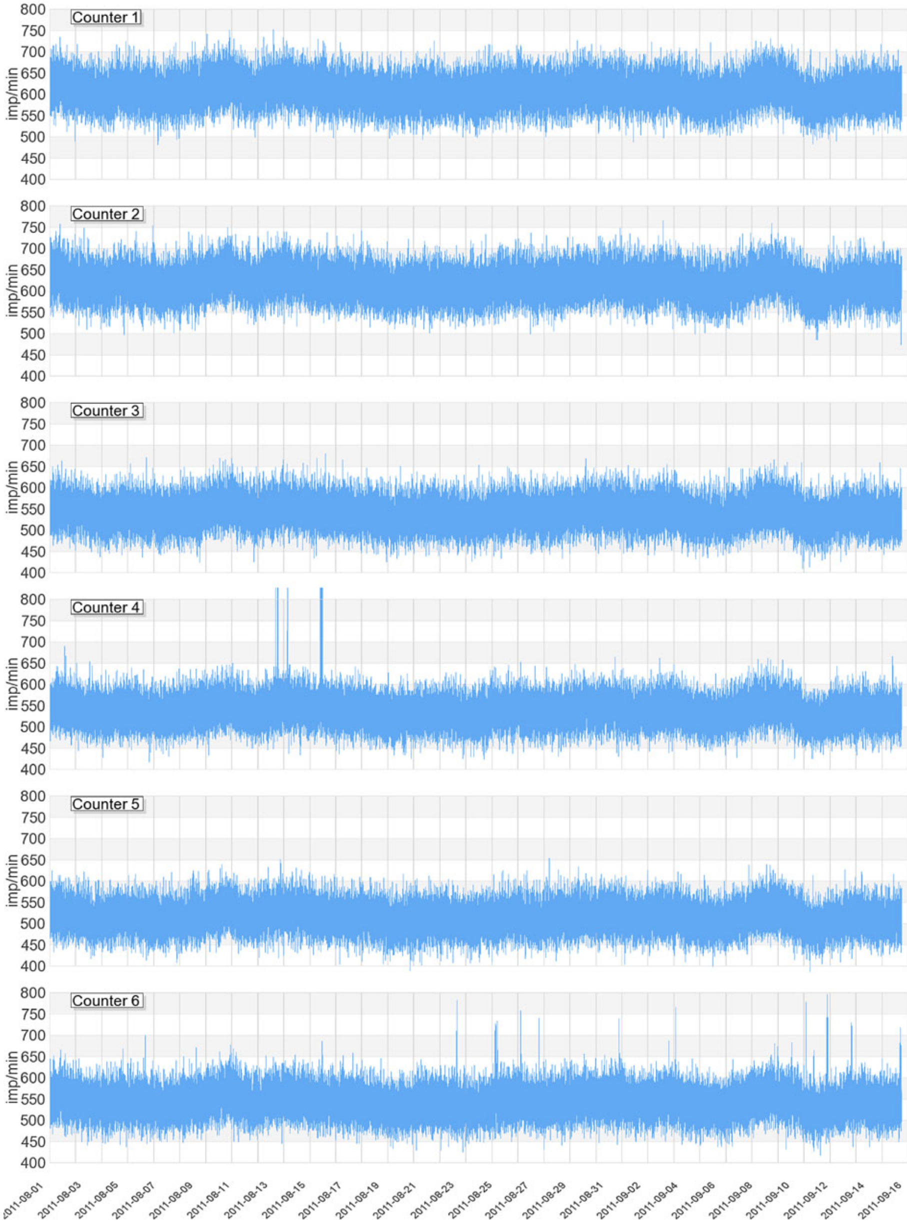


Figure 2 Uncorrected cosmic-ray data of Athens Neutron Monitor Station from 1 August 2011 to 15 September 2011.

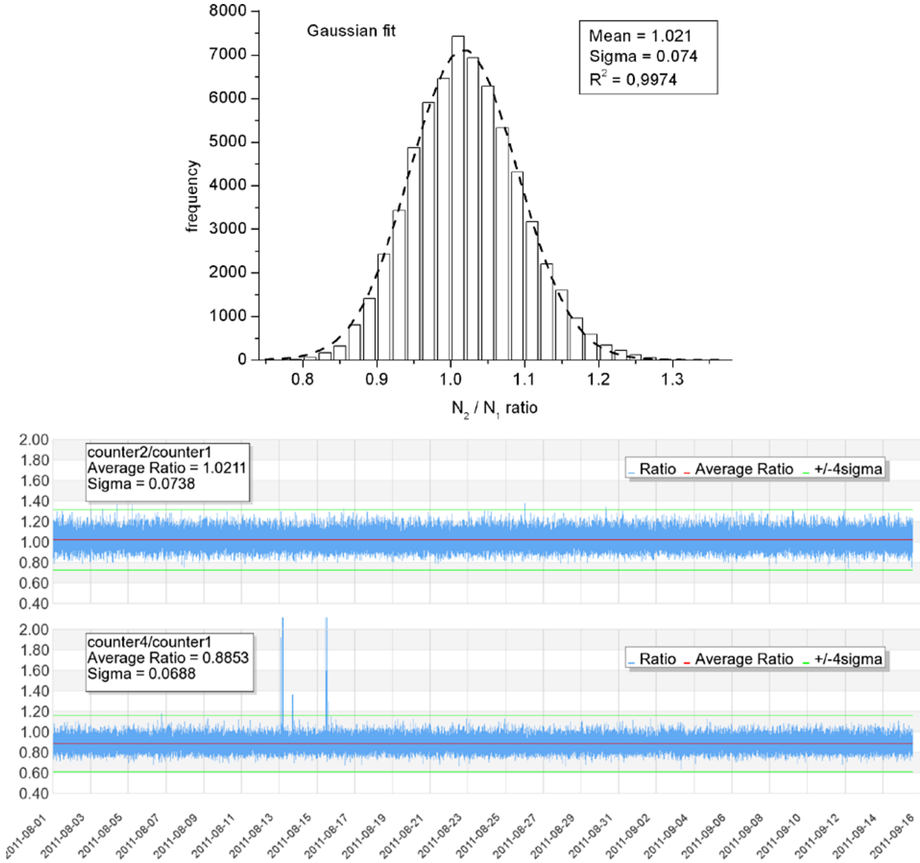


Figure 3 The histogram of the ratio N_2/N_1 (upper panel) and the time series of the ratios N_2/N_1 and N_4/N_1 (lower panel) based on the 1-min data of Athens station from 1 August 2011 to 15 September 2011.

where $\Sigma_{i,j}$ contains the sum $(\pm\frac{\sigma_i}{N\cdot A_i} \mp \frac{\sigma_j}{N\cdot A_j})$. Since $\frac{\sigma_i}{N\cdot A_i} \ll 1$ and $\frac{\sigma_j}{N\cdot A_j} \ll 1$, it becomes that $|\Sigma_{i,j}| \ll 1$. Also, since the parameters A_i and A_j are time independent, Equation (6) shows that the ratio N_i/N_j is fluctuating around the mean value A_i/A_j in a time series.

Using the 1-min raw data of the Athens station from 1 August 2011 to 15 September 2011, the histogram of the ratio N_2/N_1 is created as shown in Figure 3 (upper panel). In the same figure (bottom panel), the time series of N_2/N_1 and N_4/N_1 are also shown. Applying a Gaussian fit on the ratio N_2/N_1, it is noticed that a very high value of correlation coefficient R^2 is found, which means that the Gaussian approximation is sufficient (Hinkley, 1969; Roldugin and Vashenyuk, 1994). This conclusion is a very important result. If the counting rate of counter j is known, then the counting rate of counter i can be statistically predicted, as long as the distribution of N_i/N_j is known. The ratios N_i/N_j for all the combinations of six counters are calculated in Table 1. In order to exclude the spikes, such as the ones that are noticed in the plot of N_4/N_1, the calculation of the mean value and the sigma is performed in a $\pm 4\sigma$ trust interval. These spikes originate from the problematic behavior of counters 4 and 6 (parameter δ_i), as has already been mentioned.

4. ANN Approach and Implementation

Equations (2) to (6) assume that all the counters are working correctly, so parameters δ_i are equal to zero. Actually, in some cases, one or more detectors can measure a counting rate that is completely different from the counting rate that the other detectors measure. This measurement implies a problematic behavior of the counter such as the one noticed in counters 4 and 6. The problem that the primary data processing is called to solve, is to reject the variations $\pm\delta_i$ since they are not related to the variations of the cosmic rays or to statistical variations and they distort the raw data. The difficulty is that this rejection should be performed on a real-time basis, at a time when only the past measurements of the counters are known. So when a counter suddenly measures a value completely different from the previous one, it is not known if this change is related to a real difference of the cosmic-ray flux N, to the statistic $\pm\sigma_i$ of the measurements, or to the factor $\pm\delta_i$.

The rejection of the factor $\pm\delta_i$ could be made by comparing the counting rates of all counters. The general concept is that a sudden change of the counting rate of a detector is a real change in the cosmic-ray intensity only if similar changes are observed by the other detectors. This procedure is performed by algorithms that have already been designed. Median Editor, Median Editor Plus, and Super Editor are algorithms that are used to filter the primary data with good results (Yanke *et al.*, 2011). A different approach of the problem is discussed in this study. The filtering of the data is performed by using an ANN method where the uncorrected measurements are fed to the input layer of the network and the instrument variations filtered out are obtained in the output layer of the network. The training sample used for the training of the ANN should simulate the possible measurement states of the neutron monitor, taking into account the statistical variations of the measurements and the possible instrument variations. The training is performed by feeding the ANN with the simulation data containing instrument variations and by forcing it to output the corresponding simulation data without the instrument variations.

The analysis of the previous section is the base used to build up the architecture of the suitable ANN and the training sample. The generation of the training sample and the implementation of the ANN is made in C++. The ANN used is the MLP class from the ROOT data analysis framework, developed at CERN (http://root.cern.ch/drupal; http://root.cern.ch/root/html/TMultiLayerPerceptron.html). The multilayer perceptron, such as the one implemented in the ROOT framework, is a feed-forward artificial neural network that uses the back propagation technique for training and is consisted of at least three layers.

In order to generate the training and test samples, two kinds of random number generator are used. The first one gives numbers with uniform distribution in a defined range. The second one gives numbers with Gaussian distribution with defined mean value and sigma. The random generators are implemented in the TRandom3 class of the ROOT framework (http://root.cern.ch/root/htmldoc/TRandom.html).

The procedure of generating the training sample is shown in Figure 4 and can be described by the following three steps:

i) Using a circular procedure through the generation of the training set, a counter j is selected and is used as a reference one. For this counter, a counting rate N_j is generated that corresponds to its measurement. The range of the generated N_j was decided after browsing the Athens Cosmic Ray station data, from the beginning of its operation. By browsing the uncorrected data of the Athens station (http://cosray.phys.uoa.gr/Local_Data/form.html), it can be found that the counting rate

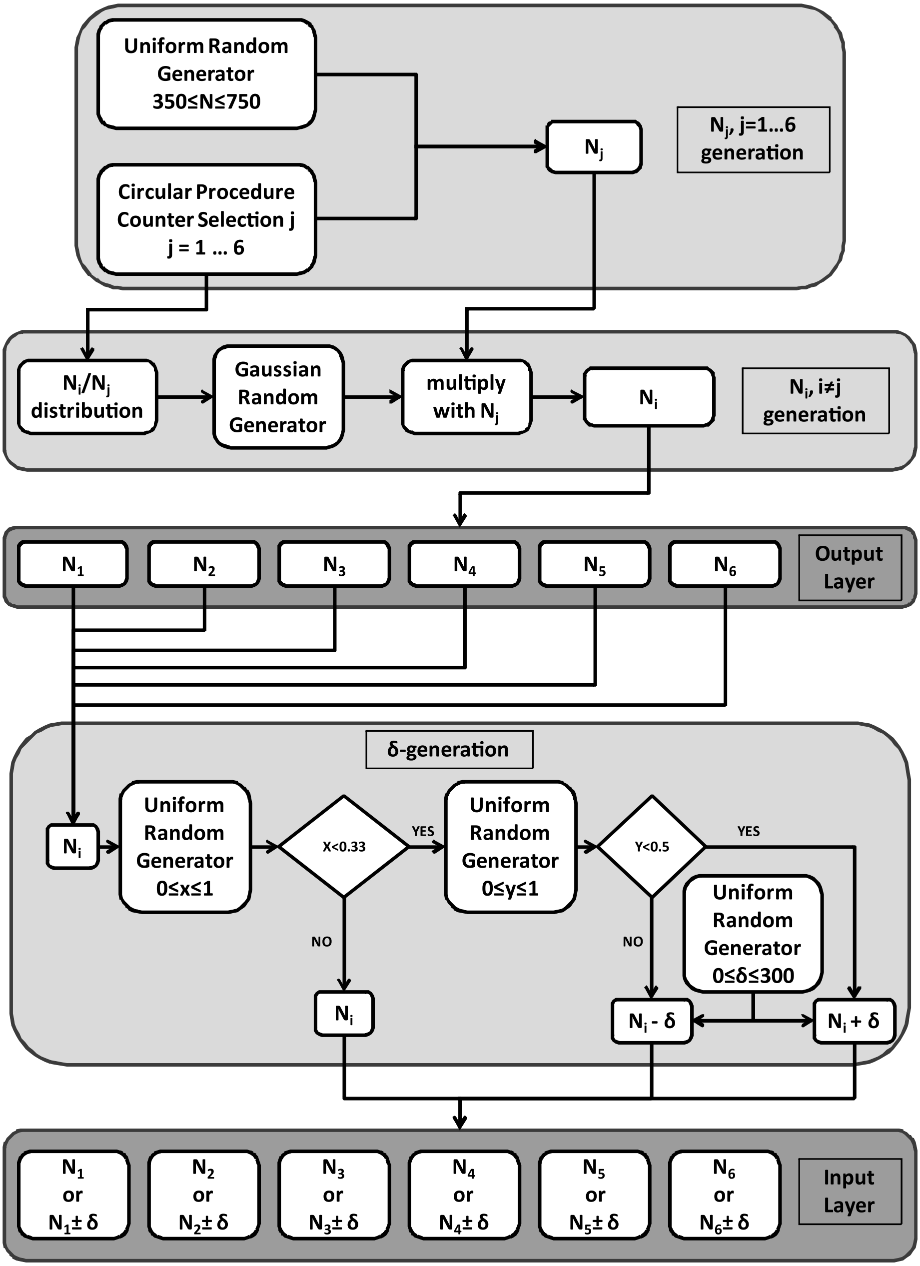


Figure 4 Generation of training and testing samples.

for the last 10 years is between 42 impulses s^{-1} and 65 impulses s^{-1} for the whole monitoring system, which means 420 – 650 impulses min^{-1} for each one of the six monitors. In order to teach the ANN to have a more general behavior, it was decided to define the counting rate of the reference counter in the range of 350 to 750 impulses min^{-1},

Table 1 Mean values and standard deviations (sigma) of the ratios for all the counter combinations.

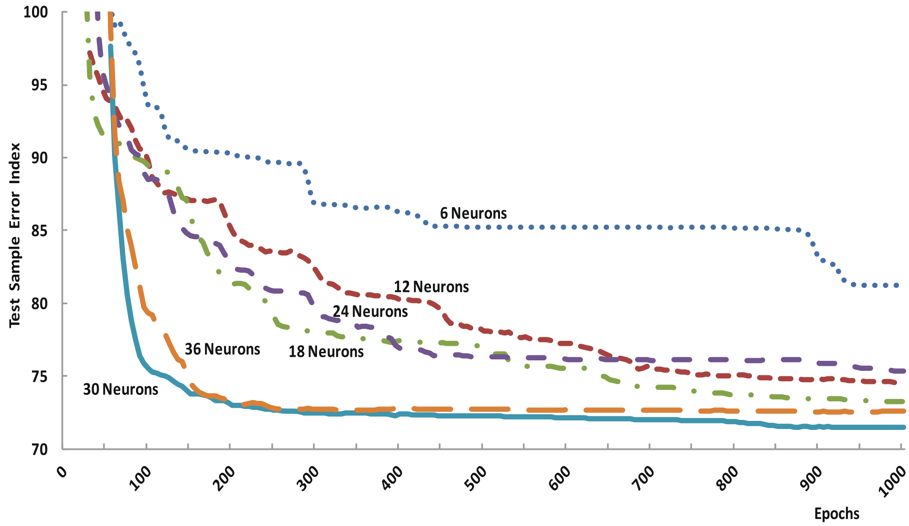
		Counter ratio distributions											
		Counter 1		Counter 2		Counter 3		Counter 4		Counter 5		Counter 6	
		Mean	Sigma	Mean	Sigma	Mean	Sigma	Mean	Sigma	Mean	Sigma	Mean	Sigma
Reference counter	1	1	0	1.021	0.074	0.893	0.068	0.885	0.069	0.849	0.065	0.895	0.060
	2	0.984	0.071	1	0	0.876	0.059	0.870	0.074	0.833	0.056	0.879	0.067
	3	1.126	0.085	1.146	0.077	1	0	0.995	0.087	0.954	0.074	1.006	0.080
	4	1.134	0.078	1.156	0.088	1.011	0.080	1	0	0.961	0.078	1.013	0.079
	5	1.184	0.090	1.205	0.081	1.054	0.082	1.046	0.093	1	0	1.057	0.084
	6	1.122	0.075	1.144	0.086	1.000	0.079	0.992	0.085	0.951	0.075	1	0

which covers the actual counting rate of the Athens NM station. Since for a long period of time a counter could measure all the values within the defined range with the same probability, the uniform random number generator is used for the generation of N_j.

ii) After N_j has been defined, the other five N_i $(i \neq j)$ should be generated as well. For this task a Gaussian random number generator is used. The generator produces five values that correspond to five ratios N_i/N_j, the distribution of which were calculated in the previous section and are shown in Table 1. Each ratio is multiplied by the value of N_j that was generated in the previous step. The result is the generation of the other five N_i. The steps i) and ii) produce a set of six values that simulates a measurement of the Athens neutron monitor in a condition where no erroneous behavior is present. More specifically, this set contains only statistical variations $(\pm\sigma_i)$ and no instrument variations $(\pm\delta_i)$. This set is assigned to the output of the ANN.

iii) In the third and last step, the generation of the input layer of the ANN is done by the following procedure. For each N_i produced in the previous step, a variation δ_i is assigned with a probability of 33 %. By selecting this probability, on average, two out of the six counters in each sample have an instrument variation. This is performed by using a uniform random generator, and by checking whether a generated number in the range between 0 and 1 is greater or less than 0.33. In the case the number is greater than 0.33, then the value N_i remains unchanged, otherwise the following procedure is performed. Using again a uniform random generator, a variation δ_i is generated in the range between 0 and 300. This range is selected for two reasons: a) since the instrument variation corresponds to an erroneous behavior, it is not possible to define a lower threshold, so it is set to zero and b) an upper threshold with a value more than 300 does not seem to improve the training of the ANN or the results, which means that 300 is an optimal value. This variation δ_i is added to or subtracted from N_i according to the algorithm shown in Figure 4. The result of this step is the distortion of some of N_i values. This new set of values is assigned to the input of the ANN.

By following the steps above, a training sample that exceeds both the characteristics of the real cosmic-ray measurements and the sporadic instrument variations was created. The counting rate exceeds the real counting rate of the last 10 years. In the training sample, two counters on average have an erroneous behavior, which is a rather rare case since the most common case is only one counter to present such a behavior. However, it is preferred to use the more general training sample, aiming to create an ANN that reacts correctly even in the most difficult cases. The generated training sample consists of 12 000 samples, 10 000 of which are used by MLP for training purposes and the rest for testing purposes. This number



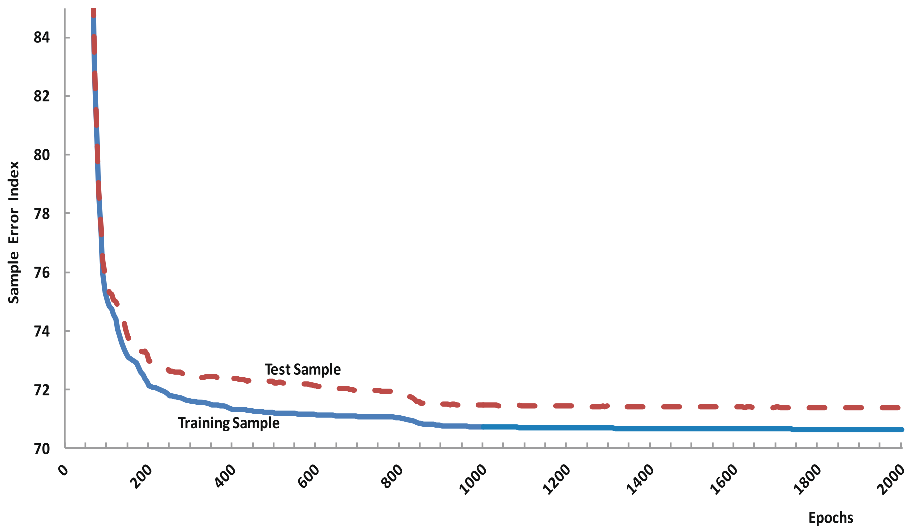


Figure 5 The error index of the test sample for various numbers of neurons in the hidden layer (upper panel) and the error index of the training and test samples for one hidden layer with 30 neurons (lower panel).

of training cases was found to be the optimal one for the multiplicity of the ANN that is trained.

After defining the input layer, the output layer, and the training sample, the best number of hidden layers is found with a trial process. It was found that the most optimal training and the best results were noticed when only one hidden layer was used. The results when using two hidden layers were worse, while the training for more than two hidden layers was not possible. In the case of the architecture with one hidden layer, the optimal number of neurons is determined again by a trial process. The error index of the testing sample during the training procedure, for various cases of the hidden layer, is shown in the upper panel of Figure 5. As can be concluded, the best results were produced when 30 neurons were used.

Table 2 ANN parameters used in this work.

ANN parameters	
Architecture	Input (6) : Hidden (30) : Output (6)
Activation function	Sigmoid
Learning method	Stochastic
Eta parameter	0.1
Training Sample	10000
Test Sample	2000
Weights Initialization	No randomization
Inputs	Normalized
Epochs	1000
Training time	≈ 10 minutes @ Intel i7

The use of more than 30 neurons gave similar results and was not selected since it is better to keep the network in the simplest architecture. The progress of the ANN training for the case of 30 neurons is shown in the bottom panel of Figure 5. Based on this plot, the optimal number of training epochs was found to be 1000. At this point the error index of the test sample has already converged to its final value, which means that the continuation of the training implies a danger of making the ANN over-trained. Finally, all the parameters of the ANN architecture and training procedure are given in Table 2.

5. Results

For comparison reasons and before announcing the results of the primary data processing using the ANN, the correction of data by using the Median Editor is presented. The profiles of the uncorrected versus corrected data for the period of 1 August 2011 to 15 September 2011 using Median Editor are presented in Figure 6. In this figure only counters 4 are 6 are presented since they are the ones with the problematic behavior. The data used were retrieved from the local data base of the Athens Cosmic Ray Station (http://cosray.phys.uoa.gr/Local_Data/form.html), where both the uncorrected data and the data corrected with Median Editor are stored. As can be seen, the Median Editor successfully filters the primary data. The problematic peaks in counter 4 and counter 6 are cut. The only problem with this method is that it also filters some of the real statistic of the counter measurements. This can be seen by the margin between the uncorrected and the corrected data, even in the cases where there are no problematic peaks.

The respective results using the ANN are shown in Figure 7. It is obvious that the selected architecture corrects efficiently the data. All the problematic peaks of counters 4 and 6 are filtered. Also, compared to the Median Editor algorithm, the ANN filters less the real statistic of the measurements, as can be seen by the smaller margin between the uncorrected and the corrected data. This improvement can be visually determined and will be statistically determined in the following paragraphs.

In order to verify the good performance of the ANN, the same ANN is applied to the Athens Cosmic Ray Station data of February 2011, which was a more cosmic-ray active period. The results are illustrated in Figure 8 where the corrected data of counter 6 are presented. The respective diagrams for the other counters have the same pattern. From this diagram, the good performance of the ANN is verified. The corrected data follow accurately

Figure 6 Correction on the data of Athens Cosmic Ray Station from 1 August 2011 to 15 September 2011 using Median Editor.

Figure 7 Correction on the data of Athens Cosmic Ray Station from 1 August 2011 to 15 September 2011 using ANN.

the fluctuations of the uncorrected ones and all the problematic peaks of counter 6 are filtered successfully.

As has been described above, the correction methods in general aim to filter all the problematic peaks while leaving unchanged the rest of the measurements. However, as can be seen in the presented results, there is a margin between the corrected and the uncorrected data in both Median Editor and ANN methods. This means that both methods, apart from the problematic peaks, compress also the statistic of the measurements even in the areas where there is no erroneous behavior. In order to determine the effect on the non-erroneous data, a quiet period of measurements where the counting rate is almost constant and with-

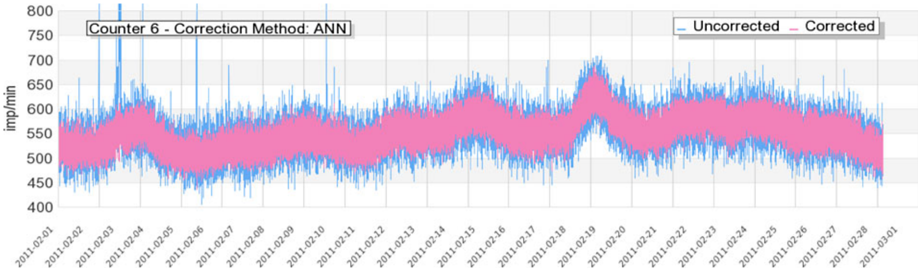


Figure 8 Correction on the data of Athens Cosmic Ray Station for the disturbed period of February 2011 using ANN.

Figure 9 Median Editor (upper plot) and ANN method (bottom plot) applied to data without errors (Athens Cosmic Ray Station for 17 – 21 August 2011).

out any problematic peaks is selected. The period 17 – 21 August 2011 is selected due to its low activity. For this period of time, the mean value and the standard deviation of the uncorrected data, the data corrected with the Median Editor and the data corrected with the ANN method data are calculated. The diagrams that correspond to the Median Editor, and to the ANN method are shown in the upper and bottom plots, respectively, of Figure 9. In this figure only counter 1 is shown since the rest of the counters have similar pattern. The calculation of the mean value and the standard deviation for the three series of data can be seen in Table 3. According to this table, the Median Editor, and the ANN method have similar performance. The mean value of their data is almost the same and very close to the mean value of the original uncorrected data. However, the ANN method reduces the standard deviation of original data less than the Median Editor, which is an improvement.

The compression of the standard deviation of the data is a known issue that the filtering algorithms present (Yanke *et al.*, 2011). The general principle of all the algorithms is that the subtraction of the undesired variations is performed by comparing the measurement of each counter with the measurements of the rest of the counters. Unfortunately, this procedure suppresses the data towards the mean value of the measurements, and as a result the real

Table 3 Statistics of Athens cosmic ray data for the quiet period of 17 – 21 August 2011.

	Uncorrected data		Corrected with Median Editor		Corrected with ANN	
	Mean value	Standard deviation	Mean value	Standard deviation	Mean value	Standard deviation
Counter 1	598.69	31.43	598.8	18.01	598.21	23.65
Counter 2	609.28	31.92	610.49	18.36	608.52	25
Counter 3	533.83	30.51	532.59	16.02	532.08	22.11
Counter 4	529.65	29.26	527.72	15.87	531.13	23.18
Counter 5	507.33	29.59	505.33	15.2	505.75	20.96
Counter 6	534.48	30.02	536.49	16.14	534.09	21.52

statistic is affected. In order to overcome this issue, the Median Editor has been improved to the Median Editor Plus version. The Median Editor Plus is the same algorithm but the correction of data is performed only to the counters that violate specific statistical criteria. As a result, only the problematic counters are corrected while the counting rate of the rest of the counters remains unchanged. The same procedure can be applied in the case of the ANN algorithm since the determination of the problematic counters through the violation of the statistical criteria does not depend on whether the method used for the correction is the Median Editor or the ANN algorithm. However, in the case of a problematic counter, the use of the ANN will give a better prediction of its correct value than the Median Editor does.

6. Conclusions

In this work, the application of the Artificial Neural Networks in the primary data processing of the cosmic-ray intensity registered by Athens Neutron Monitor Station was analyzed thoroughly. From this study it is concluded that the obtained results from the applied ANN method seem to have some advantages compared to the ones achieved by the Median Editor that is used currently for the filtering of the neutron monitors data. This is concluded since the ANN method effectively removes all the sporadic instrument variations, while it compresses the standard deviation of the rest of the data less than the Median Editor does. Moreover, the development of an ANN Plus method can be applied similarly to the Median Editor Plus, where the filtering of the data is performed only in the counters that present an erroneous behavior. Finally, in order to verify the accuracy and stability of the ANN method for a long time period, the trial application of the method can be performed by applying it to the Athens Cosmic Ray Station raw data in a real-time basis.

The implementation of new correction algorithms for the cosmic-ray monitoring data is of great importance nowadays. The technological growth and the expansion of the Internet, which makes easy the instant transfer of the information, allows the development of applications that gather the measurements of experiments and scientific programs worldwide, such as the neutron monitors, and use them in various physical models, producing physical results. For instance, the methods of space weather prediction are currently more advanced and the previous methods are now replaced by integrated knowledge-based neurocomputing models and other methods. Within the ESA Space Weather Program Study for example, a real-time forecast service has been developed for space weather and its effects. This prototype is now being implemented for specific users such as a power company system operator, who needs prediction of the local value of geomagnetically induced currents or a science tourist who needs to know whether or not aurora will occur. Soon it might even be able to predict the tropospheric climate and weather changes caused by the space weather (Lundstedt, 2005). Moreover, one more critical application that makes direct use of the neutron monitor data worldwide is the alert system that is operated in Athens Neutron Monitor Station in the framework of the High-resolution Neutron Monitor Database (http://www.nmdb.eu/) and predicts the onset of the ground level enhancements of cosmic-ray intensity (GLEs) (Mavromichalaki *et al.*, 2010). These kinds of application are highly dependent on the quality of data that are used. The use of erroneous data that are not related to the observed physical phenomena will definitely result in erroneous operation of the applications that make use of them. For all the above-mentioned cases, we believe that the use of the ANN method proposed in this work for the primary processing of neutron monitors raw data will improve all these applications since it could provide high-quality cosmic-ray data.

Acknowledgements The authors acknowledge the Special Research Account of the University of Athens for supporting the cosmic-ray research. Thanks are also due to Dr. Pantelis Papachristou for useful discussions and to the anonymous referees for the very useful comments improving this work significantly.

References

Belov, A.V., Blokh, Y.L., Klepach, E.G., Yanke, V.G.: 1988, *Kosmicheskie Luchi*, **25**, Nauka, Moscow, 113.
Carmichael, H.: 1964, *Cosmic Rays*, *IQSY Instruction Manual* **7**, IQSY Secretariat.
Chilingarian, A., Hovhannisyan, A., Mailyan, B.: 2009, In: *Proc. 31st Int. Cosmic Ray Conf.*, icrc0677. http://icrc2009.uni.lodz.pl/proc/pdf/icrc0677.pdf.
Farage, C., Pimbblet, K.: 2005, *Publ. Astron. Soc. Aust.* **22**, 249.
Hinkley, D.V.: 1969, *Biometrika* **56**, 635.
Hovhannisyan, A., Chilingarian, A.: 2011, *Adv. Space Res.* **47**, 1544.
Lundstedt, H.: 2005, *Adv. Space Res.* **36**, 2516.
Mavromichalaki, H., Sarlanis, C., Souvatzoglou, G., Tatsis, S., Belov, A., Eroshenko, E., Yanke, V., Pchelkin, A.: 2001, In: *Proc. 27th Int. Cosmic Rray Conf.* **10**, 4099.
Mavromichalaki, H., Souvatzoglou, G., Sarlanis, Ch., Mariatos, G., Papaioannou, A., Belov, A., Eroshenko, E., Yanke, V., for the NMDB team: 2010, *New Astron.* **15**, 744.
McDonald, F.B.: 2000, *Space Sci. Rev.* **93**, 239.
Riggi, S., Caruso, R., Insolia, A., Scuderi, M.: 2007, In: *XI International Workshop on Advanced Computing and Analysis Techniques in Physics Research, Proceedings of Science*. International School for Advanced Studies (SISSA), PoS(ACAT)035.
Roldugin, V.K., Vashenyuk, E.V.: 1994, *Geomagn. Aeron.* **34**, 176.
Simpson, J.A.: 2000, *Space Sci. Rev.* **93**, 1.
Van Dokkum, P.G.: 2001, *Publ. Astron. Soc. Pacific* **113**, 1420.
Yanke, V., Belov, A., Klepach, E., Eroshenko, E., Nikolaevsky, N., Kryakunova, O., Sarlanis, C., Mavromichalaki, H., Gerontidou, M.: 2011, In: *Proc. 32nd Int. Cosmic Ray Conf.* **11**, 450 (icrc0599).