



Particle identification and analysis in the SciCRT using machine learning tools

R. Garcia^{a,*}, M. Anzorena^a, J.F. Valdés-Galicia^a, Y. Matsubara^b, T. Sako^c, E. Ortiz^d,
A. Hurtado^a, R. Taylor^a, O. Musalem^a, L.X. González^l, Y. Itow^b, T. Kawabata^b, K. Munakata^e,
C. Kato^e, W. Kihara^e, Y. Ko^e, S. Shibata^f, H. Takamaru^f, A. Oshima^f, T. Koi^f, H. Kojima^g,
H. Tsuchiya^h, K. Watanabeⁱ, M. Kozai^j, Y. Nakamura^k

^a Instituto de Geofísica, Universidad Nacional Autónoma de México, Ciudad de México, 04510, Mexico

^b Institute for Space–Earth Environmental Research, Nagoya University, Furo-cho, Chikusa-ku, Nagoya 464-8601, Japan

^c Institute for Cosmic Ray Research, University of Tokyo, Kashiwanoha, Kashiwa, Chiba, 277-8582, Japan

^d Escuela Nacional de Ciencias de la Tierra, Universidad Nacional Autónoma de México, Ciudad de México, 04510, Mexico

^e Department of Physics, Shinshu University, Asahi, Matsumoto, 390-8621, Japan

^f College of Engineering, Chubu University, Kasugai 487-8501, Japan

^g Faculty of Engineering, Aichi Institute of Technology, Toyota 470-0392, Japan

^h Japan Atomic Energy Agency, 2-4 Shirakata Shirane, Tokai-mura, Naka-gun, Ibaraki 319-1195, Japan

ⁱ National Defense Academy of Japan, 1-10-20 Hashirimizu, Yokosuka, Kanagawa 239-8686, Japan

^j Institute of Space and Astronautical Science, Japan Aerospace Exploration Agency, Sagami-hara, Kanagawa 252-5210, Japan

^k Institute of High Energy Physics, Chinese Academy of Sciences Yuquan Road, Shijingshan District, Beijing, 100049, China

^l SCiESMEX, Instituto de Geofísica, Unidad Michoacán, Universidad Nacional Autónoma de México, Michoacán Morelia, 58190, Mexico

ARTICLE INFO

Keywords:

Neural networks
Machine learning
Neutron spectrum unfolding
Pattern recognition
Cosmic rays

ABSTRACT

Machine learning is a powerful tool used in many different areas, from image processing to space navigation and high-energy physics. In this paper we present a configuration of different artificial intelligent tools aimed at the extraction of features from data registered in the SciBar Cosmic Ray Telescope (SciCRT).

The SciCRT is an array of plastic scintillator bars that work nearly independently as particle detectors. When a particle crosses inside the telescope, scintillation photons are emitted by the plastics. The intensity of photons is directly proportional to the energy deposited in each bar. Taking advantage of the construction of the telescope, the small transverse area of the scintillator bars, it is possible to do particle tracking and analysis.

The main purpose of SciCRT is the detection of solar neutrons originated in the violent phenomena taking place at the surface of the Sun. Nonetheless, the SciCRT is capable of detecting different kinds of secondary particles produced by the interactions of primary cosmic rays with the atmospheric nuclei. For this reason, the task of signal classification is essential.

Our final goal will be the classification of detected cosmic ray particles, as well as, the unfolding of the neutron energy spectrum and the estimation of the angular distribution. To achieve this our methodology relies on the design of radiation detectors with different characteristics. The design of these instruments depends entirely on the information that we want to obtain from the study of the radiation.

1. Introduction

Driven by the desire to understand the origin of the cosmic radiation, acceleration mechanisms, and their relationship with phenomena occurring at Earth and space environment, scientific research requires the design of radiation detectors with different characteristics. The design of these instruments depends entirely on the information that we want to obtain from the study of the radiation.

The Sun is the nearest cosmic ray accelerator. In the solar surface, vast explosions of bursting material known as Solar flares are capable of accelerating particles to high energies. Two types of particles are accelerated during solar flares: electrons and ions. The source of their energy is believed to be the complex geometry of the magnetic field at the site of the flare. The magnetic reconnection model is one of the candidates to explain this phenomenon [1]. Observation of the

* Corresponding author.

E-mail address: rocio@geofisica.unam.mx (R. Garcia).

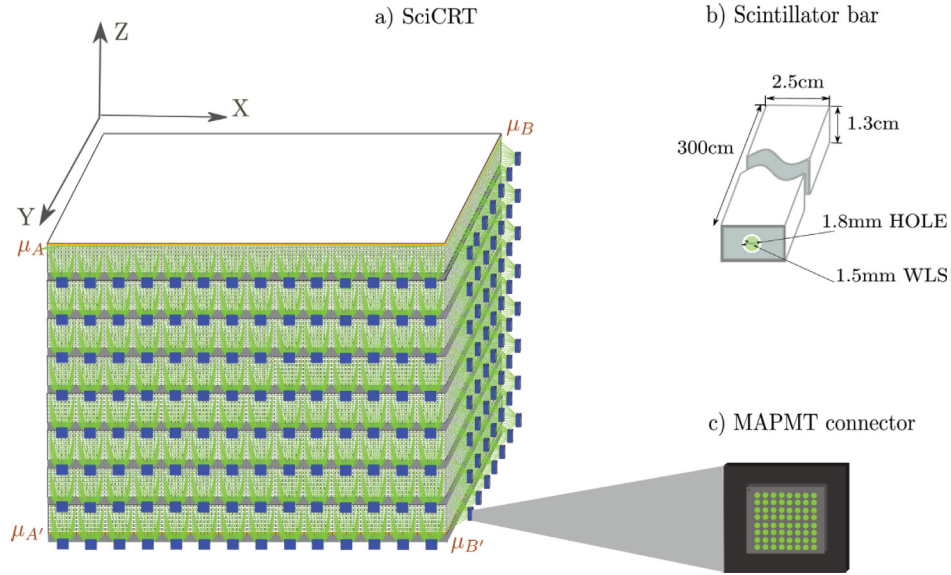


Fig. 1. Schematic diagram of the SciBar Cosmic Ray Telescope — SciCRT. a) The SciCRT has eight arrays of 1792 scintillator bars, called Super Blocks – SB – dedicated to neutron detection. The top and bottom layers are an exception because these bars are used in four fold trigger to detect muons (μ_A , μ_B). b) Scintillation bars have a hole in the center where a WLS fiber is inserted. c) MAPMT connector.

particles and radiation products is needed to test this hypothesis. The acceleration of ions and electrons have been studied by measuring X-rays, radio waves and γ -rays.

Solar neutrons are also part of the secondary energetic particles produced by nuclear reactions taking place in the solar atmosphere. They are of interest, since they carry information on the ion acceleration. The advantage of observing neutrons is that they arrive at the Earth free from the effects of the interplanetary magnetic field. Since neutrons have mass we need to measure their energy to study the acceleration of ions at the solar surface.

Neutrons are detected by instruments both at ground level and in space. Detectors for solar neutrons require the conversion from neutron to proton, and the ability to measure the energy and direction of the recoiling protons. It is also necessary to veto charged particles. Ground-based detectors are affected by the attenuation of neutrons in the atmosphere of the Earth. Free from this limitation, the first detection of neutrons from the Sun was done by satellite [2]. On the Earth two kinds of particle detectors have thus far observed neutron events associated with solar flares: Neutron Monitors and Solar Neutron Telescopes (SNT). A summary of observations is given in [3]. SNTs were designed and installed on high mountains to study solar neutron phenomena [4]. Of these, the SciBar cosmic ray telescope (SciCRT) is an improved solar neutron telescope installed on the top of the Sierra Negra volcano, Mexico [5].

The SciCRT aims to detect solar neutrons, however other particles coming from the interaction of primary cosmic rays with the atmospheric nuclei enter the detector. These particles are composed of both charged and neutral particles. On the other hand, during solar neutron events, the production of high energy γ rays at the sun is very likely. We may expect the detection of these at telescopes installed on high mountains [6].

In addition, the propagation of solar neutrons through Earth's atmosphere is characterized by absorption and scattering processes. Previous studies show [7,8] the atmospheric attenuation effect and angular spread during its trajectory to the detector, which is needed to be considered in order to deduce information of the intensity at the top of the atmosphere.

This work presents a method to classify between neutrons and γ -rays registered in the SciCRT using machine learning tools. Furthermore, we developed techniques to estimate the arrival direction and the energy spectrum of the detected neutrons. To achieve our goal we use the

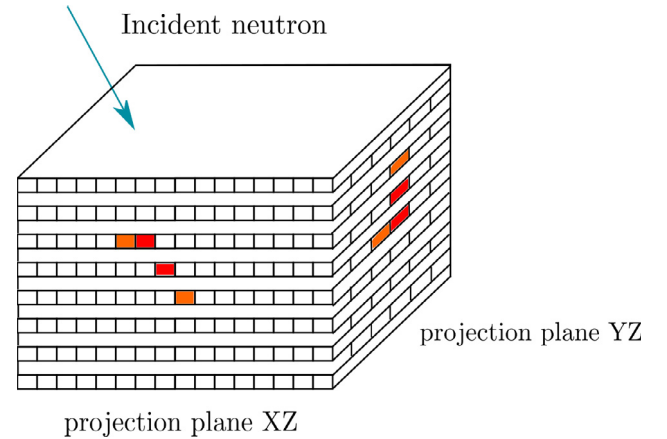


Fig. 2. Scheme of an incident neutron in the structure of the SciCRT. The track of the recoil proton (red line) is registered in two images, corresponding to the perpendicular projection on the plane YZ and XZ. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

information of energy deposition and track's morphology from data of the telescope. The machine learning algorithms allow the identification of particles and the unfolding of the neutron spectrum. The estimation of the angular distribution is done with a Fourier based track finding method.

The paper is organized as follows. Section 2 describes the operation of the SciCRT while Section 3 presents the development of a Monte Carlo simulation of the telescope, a key element in our work, since we use the output from the simulation to train and validate the machine learning algorithms. Methods for classification, the unfolding of the energy spectrum and the estimation of the angular distribution are described in Section 4. Section 5 shows the results of the algorithms applied to real data of the SciCRT. Finally, our conclusions are summarized in Section 6.

2. The SciBar cosmic ray telescope

The SciBar Cosmic Ray Telescope (SciCRT) uses the SciBar detector [9] as an improved SNT with better energy and angular resolution.

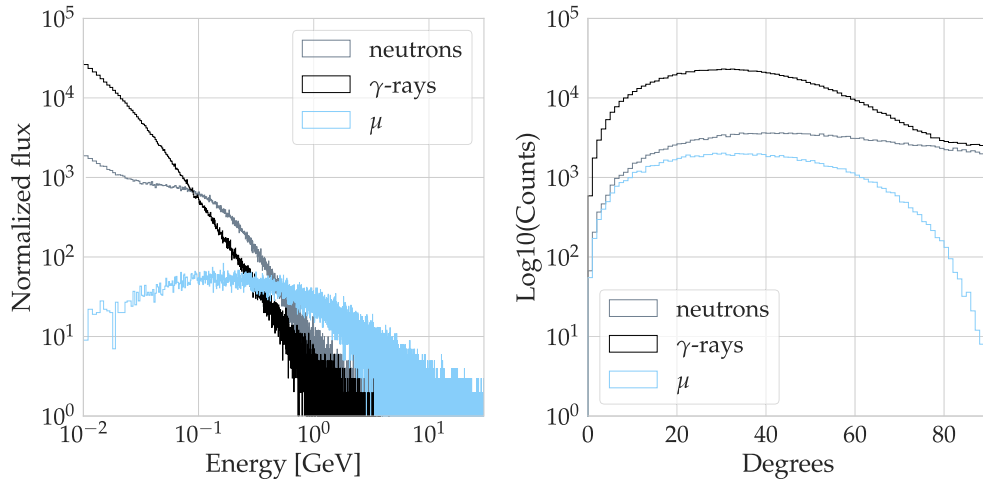
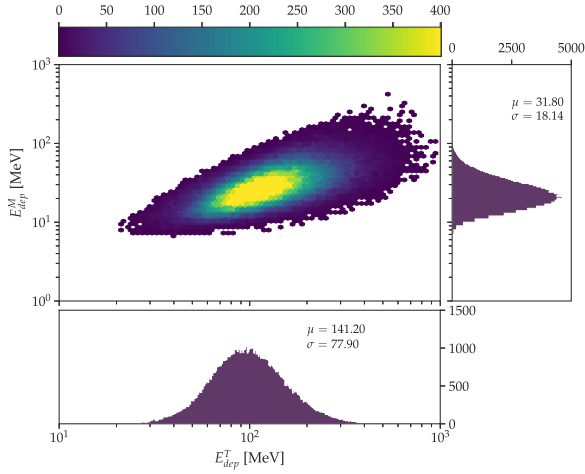
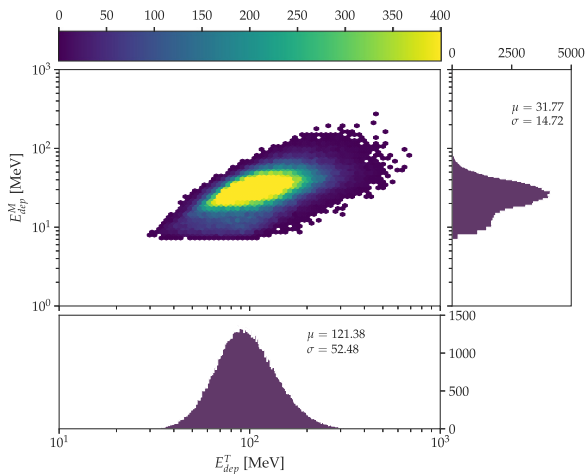


Fig. 3. The left panel shows the energy distribution of neutrons, γ -rays and μ^\pm calculated with the PARMA model. The right panel shows the zenith angle distributions for the aforementioned particles.



(a) Experimental data.



(b) Simulation

Fig. 4. Distribution of maximum energy deposition E_{dep}^M against the total energy deposition E_{dep}^T . The colorbar represents the total number of counts on each bin. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The telescope is made from plastic scintillator bars with cross section of rectangular shape and dimensions $1.3 \times 2.5 \times 300 \text{ cm}^3$. The scintillator material is Polystyrene with PPO (1 %) and POPOP (0.03 %). Each bar has a reflective coating of TiO_2 (15 %) infused with polystyrene of thickness 0.25 mm.

There are 14 848 plastic scintillator bars, organized in pairs of layers stacked up perpendicularly. One layer is composed of 112 scintillator plastic bars. An array of 16 layers of scintillators receives the name of Super Block (SB). The SciCRT has in total eight SBs, each one supported by a steel structure. A schematic description is shown in Fig. 1.

In the telescope we can discriminate cosmic ray muons using the anti-coincidence signal generated with the top and bottom layers of the SciCRT, but we need a more robust method in order to classify between γ -rays and neutrons.

When an incident particle interacts with the atomic structure of the bars it transfers kinetic energy. As result, the excited atoms emit photons until returning to the ground state. The produced photons are absorbed and then reemitted by wavelength shifting (WLS) fibers. The fibers transport the photons to a photon detector installed at one end of the fiber. The opposite side of it is painted with white paint to increase collection efficiency.

At the output of the Multi Anode Photomultiplier (MAPMT) we have an electric pulse proportional to the energy deposited. A fast custom made data acquisition system [10] and the use of trigger logic allows the reading from the MAPMTs.

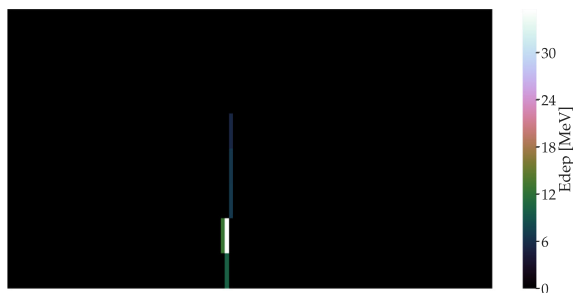
For each particle registered we assemble two images, with information from the energy deposited along the path. The images correspond to the perpendicular projection on XZ and YZ planes. Each image has a size of 62×112 pixels, since there are 112 scintillators on one layer (per each side) and 62 layers per side.¹ The intensity of each pixel represents the ADC value registered in it. Fig. 2 shows a schematic of a particle track recorded by the telescope. The generated recoil proton deposits energy along its trajectory in the layers of the detector.

3. Monte Carlo Simulation

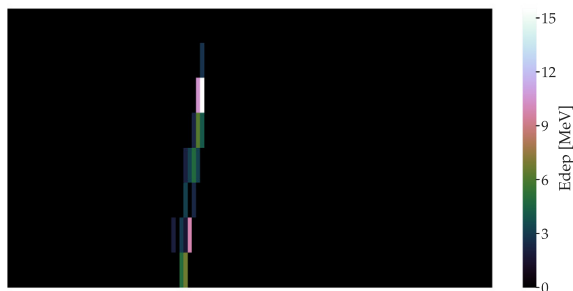
We developed a Monte Carlo simulation of the SciCRT using the Geant4 toolkit [11,12]. The architecture of the detector in the simulation consists of three main parts: plastic scintillator bars, WLS fibers and MAPMTs.

The sensitive area of the MAPMT is the photocathode. We define this surface in the simulation according to the characteristics of the

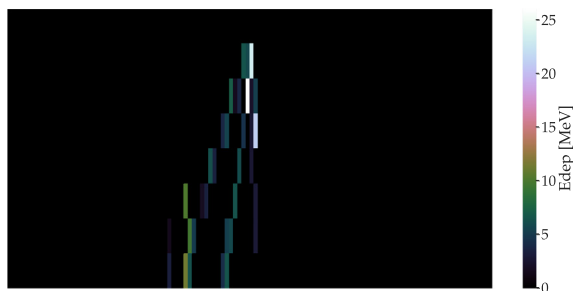
¹ 64 layers minus 2 layers used for muon detection.



(a) The main feature of clean tracks is that they have one pixel width.



(b) Spread tracks have a central path but with pixels around the main axis.



(c) The morphology of this type of track is created by multiple particle production coming from a hadronic shower.

Fig. 5. Some examples of particle tracks from the Monte Carlo simulation of the SciCRT. Each panel presents templates produced by different types of interaction.

MAPMT H8804 from Hamamatsu. The quantum efficiency is set in the wavelength range from 200 nm to 800 nm, with a maximum efficiency of 25 percent at 420 nm.

The WLS fiber includes a double cladding, with refractive indexes of 1.49 and 1.42, corresponding to the inner clad and the outer clad of the fiber. Additionally, we take into account the absorption (430 nm peak) and emission spectra (476 nm peak). Other important factor is the attenuation length of the WLS (>350 cm) because it directly impacts the number of photons arriving at the photocathode surface.

In relation to the plastic scintillator, we specify the emission spectrum reported in [13] with peak emission at 420 nm. The calibration of the remaining parameters from the simulation (time constants, attenuation length, light yield, among others) was made using experimental data. For this purpose we designed and performed an experiment using cosmic ray muons. Further details of this procedure are described in another paper [14].

After the adjustment to match real data, we set the photon yield equal to 8000 photons/MeV. Values for other parameters include: reflectance of the TiO₂ coating (0.9), fast time constant of plastic scintillator (3.6 ns) and bulk attenuation length of the scintillator (250 cm).

The output from the simulation is the energy deposited per bar and the number of photons detected by the sensitive area of the

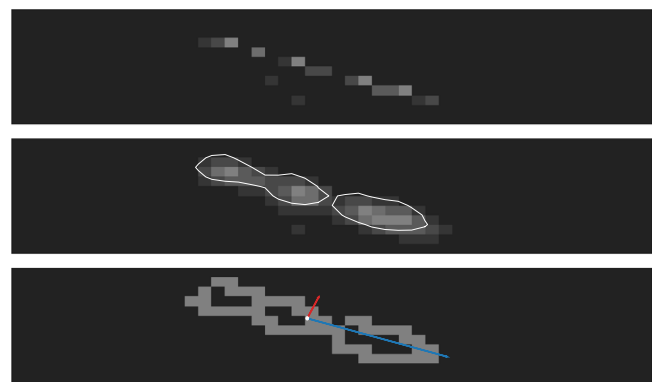


Fig. 6. Methodology for the morphology analysis of the tracks. The top image shows a particle track with discontinuities. The middle image shows the smoothing of the track after applying a Gaussian kernel and edge detection (white line). The bottom image is found connecting the nearest contours and finally fitting an ellipse. The red line is the semi minor axis and the blue line is the semi major axis of the ellipse that contains all pixels of the track. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

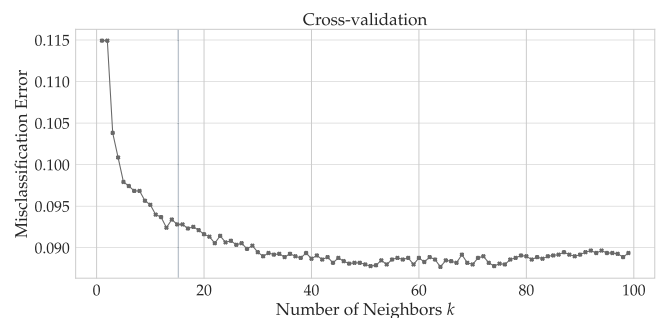


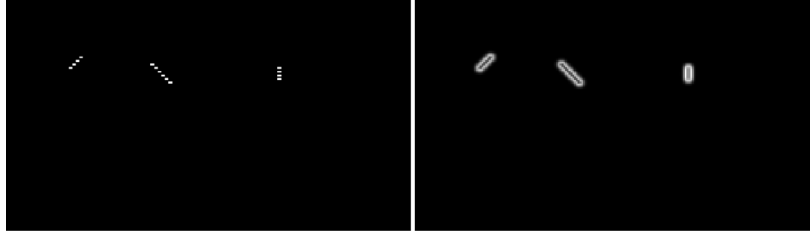
Fig. 7. Selection criteria to the voter number of the KNN method. We choose the value with the small misclassification error, keeping k to a minimum to avoid computational costs.

MAPMT. Employing these data we generate images corresponding to the projections of the tracks.

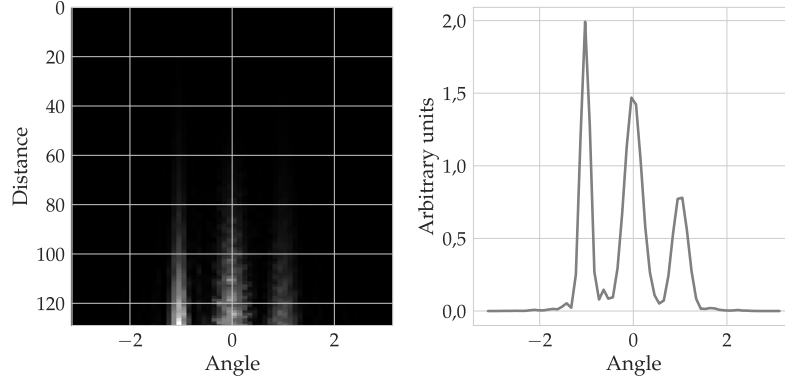
Regarding the physical processes, the simulation takes into account classic electromagnetic processes, ionization, Bremsstrahlung, multiple scattering, pair production, Compton scattering and photoelectric effect. The hadron interaction model is QGSP BERT [11,12]. Given that the simulation considers the propagation of optical photons in the WLS, the following optical processes are included to correctly simulate their behavior: scintillation and Cherenkov emission, bulk absorption, Rayleigh scattering and boundary processes (reflection, refraction and absorption).

We use PARMA model 4.0 (PHITS-based Analytical Radiation Model in the Atmosphere [15]) in conjunction with the detector architecture in Geant4 to get training data for the machine learning methods. The PARMA model is capable of estimating cosmic ray fluxes for different kinds of particles in the Earth's atmosphere at nearly anytime and anywhere.

Since 2013 the SciCRT has been working at the summit of the Sierra Negra volcano, Puebla (at an atmospheric depth of 575 g cm⁻² [16]) in Mexico. Using the PARMA model we estimated the secondary cosmic ray flux considering the altitude and the geomagnetic location of Sierra Negra. The calculation includes contribution from photons, neutrons and μ^\pm . Fig. 3 show the normalized flux and zenith angle distributions. We did not include other charged particles in the simulation given that we use an anti-coincidence system to discard them. Although muons are also charged particles they are useful for the calibration of the simulation and evaluation of the performance of the algorithms.



(a) The left panel shows the artificial tracks with different angles. The right panel shows the edge detection.



(b) The left panel shows the discrete Fourier transform mapped into polar coordinates. The right panel shows the angular energy density. The magnitude of the maxima is proportional to the length.

Fig. 8. Methodology for the calculation of the angular energy density.

The accurate Monte Carlo simulation of the interactions in the detector is an important part of the training phase of the learning algorithms. Fig. 4 shows a comparison between simulation and real data of the neutron energy deposition. For this we calculated the total energy deposition of the track (E_{dep}^T) and maximum energy deposition in a scintillator bar out of the track (E_{dep}^M). As shown in the figure, both distributions are centered around the same area (for the case of maximum energy deposition: $\mu_{sim} = 31.77$, $\mu_{exp} = 31.80$; in the case of total energy deposition: $\mu_{sim} = 121.38$, $\mu_{exp} = 141.20$) with large density of events in both cases (yellow color). Differences between the experimental data and the simulation arise from the finite resolution of the detector, the gain of MAPMT (both not included in the MC) and some contamination from gamma rays.

One remarkable feature of the distribution of maximum energy in the simulation is the ridge around 10 MeV. This characteristic is produced by protons that do not completely stop in the volume of the detector and therefore only deposit a fraction of their energy [17]. This feature is not present in the distribution from the experimental data and we suspect this is hindered by gamma ray tracks. Nonetheless we think that a detailed study is needed to better address this issue, which would be beyond the scope of this paper.

The neutron like events were selected from the data registered by the telescope employing the method described in the next section.

4. Description of the method

The first step in our analysis is data cleaning and event selection. We discard all the tracks that have energy deposited in the first layer of the detector, as these are certainly produced by charged particles. We use this anti-coincidence trigger to classify between charged and neutral particles.

The detection of neutrons in the SciCRT is based on elastic scattering between neutrons and the hydrogen nuclei in the scintillator material (see Fig. 5a). On the other hand, there is a small probability that

neutrons interact with the target nucleus via spallation. Fig. 5c shows an example of a hadronic shower produced by this process.

Furthermore, γ -rays are also detected at the SciCRT as neutral particles. The mechanism of interaction between the scintillator material and γ -rays is mainly pair production (e^+ , e^-) and Compton scattering. Fig. 5b shows a spread track that was produced by a γ -ray via pair production.

Considering the characteristics of the interactions, in order to classify the type of interactions between particles we made an analysis of the morphology of the registered tracks. For each event we measure the width and length of the tracks. To this purpose we use a segmentation method [18] and after we fit an ellipse to the tracks.

Following the segmentation, we apply a Gaussian blurring kernel to discard phantom edges. Edge detection and contour connection is performed afterwards on all the area of the tracks. As a result of the fitting we obtain the semi major and semi minor axes of the ellipse. This process is summarized in Fig. 6.

Therefore for each track we have two parameters to describe its shape, the width and length (w, l). For the classification we use k -means clustering as an unsupervised machine learning algorithm [19]. This method is appropriate for our goals since we need to retrieve information about the underlying structure in the data (inherent groupings) without prior information about the possible outcome or training data. The method makes the partition of N observations into k disjoint clusters, each described by the mean of the samples in the cluster. The mean of each cluster is commonly called the centroid.

As input to the algorithm we define three groups. One of them is the clean tracks produced by proton recoil, typically having one pixel width. The other two groups are tracks with broader shapes. We classify these as spread tracks and hadronic shower type.

Usually the seed centroids (w_c, l_c) that represent the groups are initialized randomly and the algorithm iterates until finding a suitable value. However, better results are obtained suggesting the initial centroids based on empirical knowledge of a small sample of the data set

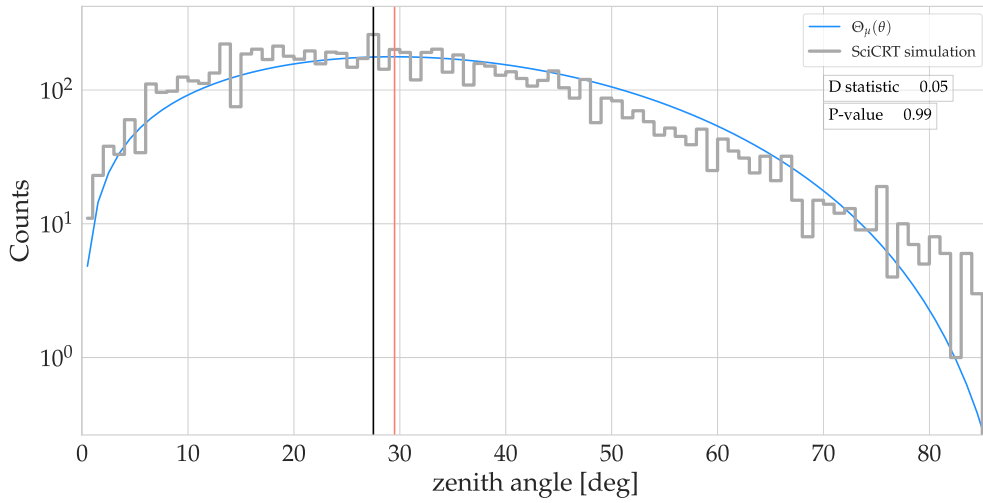


Fig. 9. The zenith angle distribution from muons after being detected in the simulation and the theoretical model $\Theta_\mu(\theta)$. The black vertical line indicates the angle 27.5° which gives the maximum counts of the histogram. On the other hand the red line indicates the angle 30° which gives the maximum of the model. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

from the simulation. Subsequently the k -means algorithm optimizes the position of the centroids.

The percentage of the events corresponding to either spread tracks or hadronic type tracks is $\sim 13\%$. After the classification, further analysis is done only with the group corresponding to clean tracks. This is because we need to develop a different method to estimate the arrival direction and identification of each particle generated in the hadronic type data.

Until now, only the geometric features have been of concern, however, in order to separate neutrons from γ -rays we need an analysis of the energy deposition. This is because γ -rays are capable of producing similar tracks to the proton recoil type.

Considering this, the next step is the extraction of the characteristics related to the energy deposition per track. In this step the parameters measured per track are: maximum energy deposition (E_{dep}^M), total energy deposition (E_{dep}^T) and energy density ($\rho_{Edep} = \text{total energy deposition/number of pixels}$). We analyze these features in a 3-dimensional space: $\vec{x}_i(\rho_{Edep}, E_{dep}^M, E_{dep}^T)$. From now on we will refer to this vector as the features vector. The method of classification that we use is a supervised learning algorithm: k -Nearest Neighbors (KNN) [20]. Taking advantage of previous information (simulation data) the KNN algorithm works on data which are labeled and therefore makes easier the classification.

The training data set is necessary to find a majority voting group that help us to identify the class of an unlabeled vector x_q (query point). In other words, for each classification group we have a pertinent training data set that has an exclusive label. When we want to classify new information, we check the label group of the k nearest neighbors and classify according to the majority.

As mentioned above, the training data set was obtained from the SciCRT simulation with the PARMA input. The simulation considers a large energy range from 0.1 GeV to 100 GeV, for both types of particles: neutrons and γ -rays. The data obtained from PARMA considers the solid angle of the detector. The injection of the particles in the Geant4 simulation is done from a hemisphere following the zenith angle distribution estimated by PARMA (see right panel in Fig. 3).

In order to select the number of neighbors k suitable for the problem we use the Cross-validation method. Cross-validation is applied in error testing associated with the learning method in order to evaluate its performance, or to select the appropriate level of flexibility [21]. The method involves randomly splitting the simulation data set into two groups, some used as testing data and others as training data. Afterwards, we apply the KNN method on the training set and evaluate it on the test set. Finally, we compute the average of the mean squared

Table 1

Result of classification between neutrons and γ -rays using KNN method.

		Real percentage	
		Neutron (55.55%)	γ -rays (44.45%)
Reconstructed	Neutron	90.92%	8.84%
	γ -rays	9.08%	91.16%

error and repeat this procedure for different values of k . Fig. 7 shows the result of the cross validation as a function of k .

The optimal number of neighbors is 64, the one with lowest misclassification error. However, a good compromise is reached at $k = 15$ with an error of ~ 0.09 . As k increases above 15 the rate of decrease in the misclassification error is small. Meanwhile, using larger values of k is computationally expensive.

To find the nearest neighbors we measure the distance including a weight parameter w_i in the measurement. This is defined as the inverse of the distance, to give closer neighbors a greater influence than the ones which are further away.

The evaluation of the performance of the KNN method classifying neutral particles was done with data from a simulation consisting of μ^\pm , γ -rays and neutrons. The first step was the classification between charged and neutral particles using the anti-coincidence trigger. After, we selected clean tracks employing the k -means clustering method. Finally the KNN method was used to separate between neutrons and γ -rays. Table 1 shows the percentage of separation in comparison with the real abundance in the data set. With our methodology we achieved a sensitivity of 90.92% in the neutron classification and an overall accuracy of 91.04%.

Our next goals are the unfolding of the neutron spectrum and estimation of the angular distribution.

The angular distribution was estimated using a Fourier based method. As a way to calculate the arrival direction of the incident particle in the detector we estimate the slope of the projected tracks and calculate the zenith angles. We tested several algorithms for line detection in our image set (such as Hough transform [22] and Radon method [23]) but decided upon the Fourier transform based method (FT) because we got better resolution and efficiency.

The slope detection is based on a method proposed in the study of astrophysical images [24]. The proposed methodology is as follows: 1. edge detection, 2. calculate two-dimensional discrete Fourier transform of the image, 3. mapping from Cartesian into Polar coordinates in the Fourier space, 4. integration of the Polar representation (Eq. (1)) to estimate the angular energy density $S(u)$.

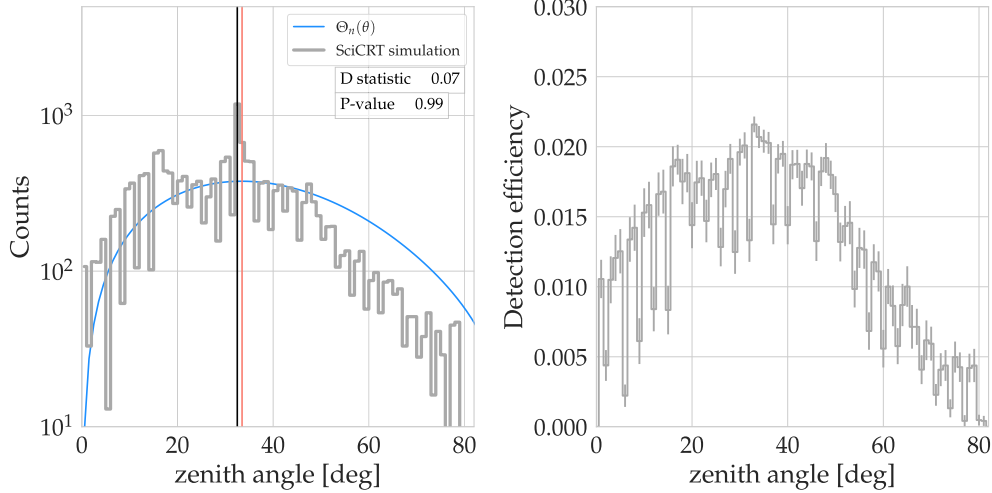


Fig. 10. The left panel: Zenith angle distribution from neutrons after being detected in the simulation and the corresponding theoretical model. The red vertical line points to the maximum value 33° of the model, while the black vertical line is the zenith angle 32° where the histogram is maximum. The right panel: Neutron detection efficiency as a function of the zenith angle with error bars representing $\pm 1\sigma$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The angular energy density represents the amount of energy in the image, with local maxima corresponding to the angles of the straight lines detected. In discrete polar coordinates the angular energy density takes the form

$$S(u) = \frac{2}{\pi N^2} \sum_{v=1}^N v \cdot F_{u,v} \quad (1)$$

where v, u represents the discrete polar coordinate radius and angle, respectively. $F_{u,v}$ is the Fourier transform in the polar system and N is the number of samples on this space.

An example of the procedure is shown in Fig. 8. The top panel 8(a) shows artificial data for test and the subsequent edge detection. The bottom panel 8(b) shows the result of mapping the Fourier transform into a polar system. Also, integrating the matrix of Fourier coefficients over the v axis gives the angular energy density.

In order to calculate the zenith angle of the tracks, we use a selection criteria, processing the tracks formed by four pixels or more in total. This is done to eliminate shorter tracks that do not contain enough information to accurately estimate the angular distribution.

We performed a test to evaluate the performance of our method. We simulate cosmic ray muons for this goal, since muons produce on average tracks of one pixel width, and only one track per incident particle.

The zenith angle distribution of the muons before interacting with the detector is shown in the right panel of Fig. 3 (PARMA data). We may model this distribution using Eq. (2). If we omit the temporal variability and assume a uniform azimuthal angular distribution, the number of muons (dN) that cross a defined area with zenith angle(θ) and azimuth angle (ϕ) is described as:

$$\frac{dN}{dA_\perp d\theta d\phi} = I_{(\theta=0)} \cos^n \mu(\theta) \sin(\theta) \quad (2)$$

where n_μ depends on the atmospheric depth X [g cm^{-2}] and the energy range. The $\sin(\theta)$ term comes from the solid angle element while dA_\perp is the area perpendicular to the incident direction [25]. A good fit to the PARMA data is achieved with $n_\mu = 2.26 \pm 0.03$.

In order to fit this model to the simulation corresponding to the interaction of muons with the detector we need to take into account the effective area of the SciCRT. This requires analyzing the case of particles entering at the horizontal and vertical faces of the detector.

Using Eq. (2) the total number of particles arriving on the horizontal face as a function of the zenith angle ($N_h(\theta)$) is obtained using $dA_\perp = \cos(\theta)dA$. The contribution from incident particles on the four vertical

faces ($N_v(\theta)$) is calculated using $dA_\perp = \sin(\theta)\cos(\phi)dA$. Hence, the distribution of the number of particles that enter the SciCRT is a linear combination (Eq. (3)) between the particles entering on the 5 faces of the telescope:

$$\Theta_\mu(\theta) = \alpha \cdot N_h(\theta) + \beta \cdot N_v(\theta) \quad (3)$$

The parameters α and β are proportionality constants that were estimated considering the area of the faces, the vertical intensity $I_{(\theta=0)}$ and the corresponding range of angles for each face (horizontal and vertical positions). Finally the model can be written using Eq. (4).

$$\Theta_\mu(\theta) = \sin(\theta) \cos^n \mu(\theta) (A \cdot \cos(\theta) + B \cdot \sin(\theta)) \quad (4)$$

where $A = 607.61 \pm 18.25$ and $B = 29.41 \pm 0.88$. Fig. 9 shows the model and the angular distribution of muons registered by the SciCRT simulation, with a good agreement between the two. A goodness of fit test (Kolmogorov–Smirnov Test) generated a p -value of 0.99 and therefore we do not reject the null hypothesis, that the data is sampled from the proposed distribution ($\Theta_\mu(\theta)$). One point to support this is that the maximum values of the distributions are very close to each other. The maximum of the model is located at 30° , meanwhile the maximum of the distribution is at 27.5° . The difference between the two may be regarded as a systematic error from the Fourier method and detector resolution.

Next, we evaluate the zenith angle distribution for the simulation with neutrons. This distribution is expected to be observed in the data from the telescope, consequently we apply the analysis previously made for the muon data.

We describe the angular distribution by PARMA (see right panel of Fig. 3) using Eq. (5), again including the solid angle of the detector [25]. In this case the parameter obtained by the fit is $n_{neut} = 2.14 \pm 0.03$ and the maximum value of the model is 41° .

$$\frac{dN}{dA_\perp d\theta d\phi} = a \cdot \sin(\theta)(1 + b \cdot \cos^n \mu(\theta)) \quad (5)$$

Taking into account the correction by the effective area of the SciCRT, we obtain the distribution $\Theta_n(\theta)$ using the same procedure previously described for muons. Eq. (6) shows the result of this calculation:

$$\Theta_n(\theta) = A \cdot (1.0 + B \cdot \cos(\theta)^{n_{neut}}) \sin(\theta)(C \cdot \cos(\theta) + \sin(\theta)) \quad (6)$$

where $A = 15.23 \pm 1.05$, $B = 2.83 \pm 0.04$ and $C = 41.94 \pm 2.84$. The left panel in Fig. 10 shows the zenith angle distribution from the simulation of neutrons interacting with the SciCRT, with maximum at 32° . The model (Eq. (6)) is shown in the same figure with a maximum value at

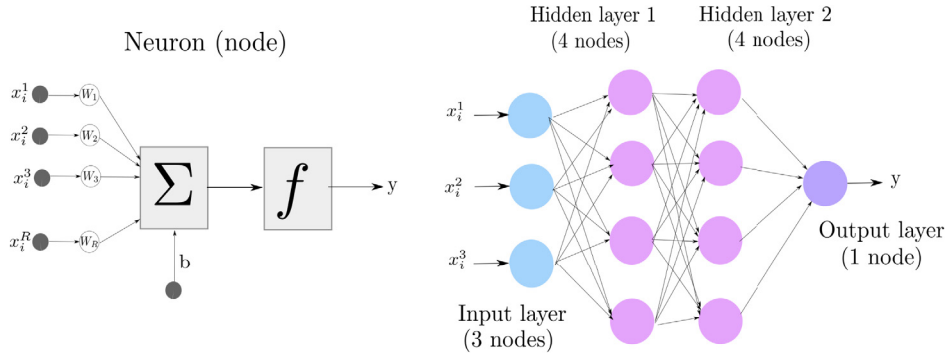


Fig. 11. The left panel shows the schematic structure of a simple neuron. The right panel shows a schematic example of an ANN. The architecture is formed by two hidden layers, each layer with four neurons, the input vector (\vec{x}) of three dimension and one node in the output layer.

33°. Although the difference between the peaks is not significant, we can observe that for angular values higher than 50° there is a lack of events in the histogram in comparison with the model. The $\Theta_n(\theta)$ model and the data distribution have a good agreement (p -value = 0.99) in angles less than 50°.

The possible origin of the lack of events for large angular values is related with the energy of recoiled protons and the length of their tracks. In principle the tracks with higher angular values (>50°) require larger track lengths, since shorter tracks do not have enough information to be measured accurately. The larger track lengths are produced by more energetic particles. On the other hand lower energy neutrons producing recoil protons with scattering angles in the range 50° to 80° will not deposit enough energy to generate tracks that pass the selection criteria (more than 4 pixels). Hence we have a deficit in the number of events at large zenith angles.

To further analyze this we calculate the neutron detection efficiency as a function of the zenith angle, measuring the relation between the reconstructed angle and the incident one. We use a simulation with a uniform angular distribution to this purpose. The efficiency is shown in the right panel of Fig. 10 with a bin width of 1°. The distribution describes fluctuations more or less periodic originated by the angular resolution of the SciCRT. In addition, the small decrease observed around 27° may be produced by the differential scattering cross section of the Hydrogen nuclei [26].

Lastly, we use an artificial neural network (ANN) to unfold the energy spectrum, estimating the neutron spectrum at the top of the detector. The ANN is a computational model inspired in the learning process of the human brain, recently gaining interest in many areas because it can identify the underlying patterns on the data and learn from them. We build a neural network to perform a regression. The use of the ANN for the reconstruction of the neutron spectrum is a novel method and it has been shown to produce promising results [27,28].

The mathematical model of an ANN is an ensemble of linear equations. Eq. (7) shows a description of a neuron, where the input vector $\vec{x}_i = [x_i^1, x_i^2, x_i^3, \dots, x_i^R]$ is the features vector of dimension R . The parameter weight (W_i) and bias (b) are the free parameters, which will be computed at the training stage. The task of the transfer function f is to map the result of the vector equation into the range 0 to 1.

$$y = f(\sum \vec{W}_i \vec{x}_i + b) \quad (7)$$

The left panel in Fig. 11 shows a scheme of a single neuron, while the right side panel shows a more complex topology where we have a neuron set conforming layers, connected in cascade. The number of neurons (nodes) in the input layer is equal to the size of features vector. These nodes are passive because they do not make any operation on the data.

The free parameters are adjusted using a particular input leading to a specific target output. To this goal we use the features vector $\vec{x}_i(\rho_{dep}, E_{dep}^M, E_{dep}^T, \theta)$ as input. The first three parameters were previously used in the KNN method and we add the zenith angle from the

tracks as input to ANN. The target output is the neutron spectrum from PARMA (observable events by the SciCRT).

The training process is based on an iterative comparison of the output with the expected data until a good agreement is reached: optimizing the mean squared error. The data used in training is in the range of energies from 0.1 GeV to 4 GeV. The lower energy limit was set considering the detection efficiency for neutrons (energies above 50 MeV [5]) and a criteria based on the reconstruction of the tracks. From simulation we know that only ~36% of neutrons with energy of 100 MeV are capable of producing tracks with 4 pixels (reconstruction criteria). Considering both conditions the number of detected events below 100 MeV becomes very low. On the other hand, the maximum energy deposition in one SB is ~4 GeV, therefore we established this as the upper limit for the training data.

For the development of the ANN we use Keras with TensorFlow [29, 30]. Both constitute a deep-learning framework for Python, convenient to define and train almost any kind of deep-learning model. The design of our neural network has five hidden layers with 64 neurons each. This architecture was selected by systematic experimentation [31,32]. For this we configured a workstation PC and tested a large number of different architectures, finally selecting the one with best performance. The metric we used to select the architecture was minimizing the mean square error. The transfer function applied is the Swish function [33].

The ANN model was tested using simulation data not used in the training step. The output from the model is shown in Fig. 12 where we observe a good agreement between the two distributions. To test this we used the two sample Kolmogorov-Smirnov test to compare the unfolded and true histograms. The result of the test (D -statistic= 0.05 and a p -value = 0.98) prove there are small discrepancies between both distributions and therefore we do not reject the null hypothesis.

An important point to mention before we finish this section is that the SciCRT is currently upgrading the data acquisition electronics so, one eighth (third SB) of the total detector operates with upgraded data acquisition at the moment. In the following section we use the data obtained from the third SB of the SciCRT and compare it with the simulation.

5. Assessment

In this section we present the results of our analysis using data from the SciCRT. The first step is to convert from ADC values to energy deposited and define a threshold in order to reject the pedestal data.

The typical ADC distribution of a scintillation bar has two main peaks (see Fig. 13). The left one is the pedestal, originated by noise and background radiation. The second peak is proportional to the energy deposited by high energy particles. In order to define a threshold, we use cosmic ray muons and calculate the ADC distribution for each MAPMT. The threshold level is determined beyond the muon distribution, around 90 ADC on average, which agrees to the value determined by the previous method reported in [34].

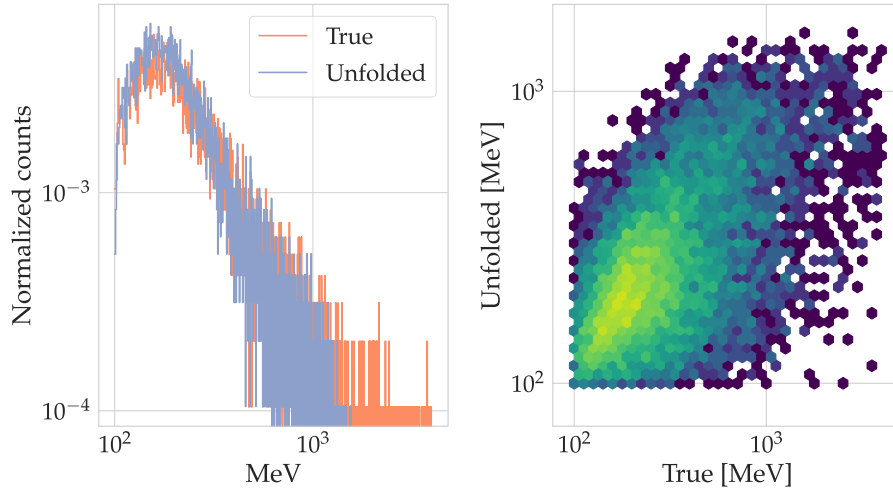


Fig. 12. Neutron spectrum unfolding using the ANN model. The left panel shows the histogram of data with a bin size of 1 MeV. The right panel shows the relationship between the input energy from the simulation and the estimated energy event by event.

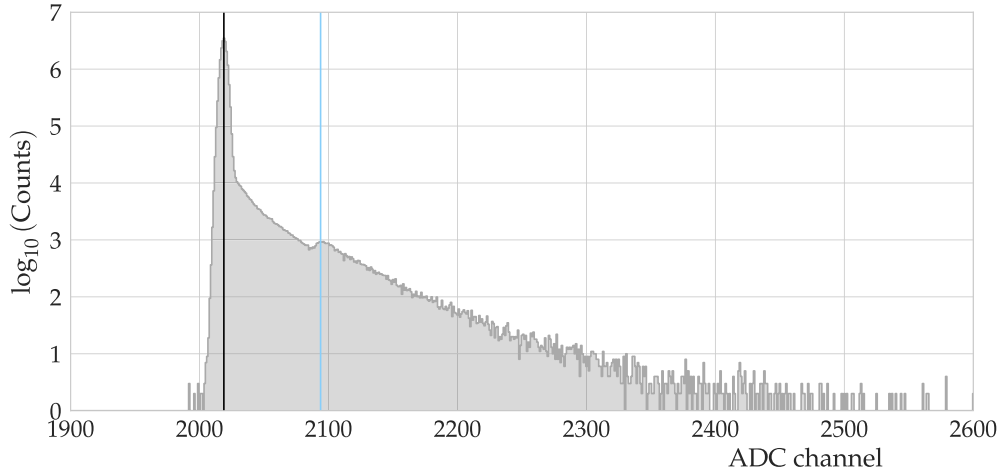


Fig. 13. The ADC distribution from neutron channels in the SciCRT. The black vertical line indicates the peak of pedestal and blue vertical line indicates the threshold. Peak signal from neutron is around 2100 ADC. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The conversion between ADC values and deposited energy is done with a gain value of 14 ADC/MIP. In this step we also require to correct the attenuation effect from the WLS fiber. For this, we calculate the distance x between the approximate point where the particle crosses the scintillator bar and the MAPMT. Using the distance, we correct the attenuation effect with [35]:

$$ADC_{real} = \frac{ADC_{obs}}{\left(\exp\left(-\frac{x}{\lambda}\right) + R \exp\left(-\frac{2.0x_{tot}-x}{\lambda}\right) \right)} \quad (8)$$

where $x_{tot} = 330$ cm is the total length of the scintillator bar, $\lambda = 408$ cm is the attenuation length of the WLS fiber and $R = 0.54$ is the reflectance of the white paint at the end of the bar.

Subsequently we made the selection of the candidate tracks for neutrons analyzing the geometric features and applying the clustering algorithm. The 85 % of all tracks is selected as clean tracks. After this, we classify the data between γ -rays and neutrons using the KNN model. As a result we have a neutron data set with a small percentage of contamination by γ -rays (≤ 10 %).

Next we perform the unfolding of the neutron spectrum using the ANN. Fig. 14 shows the obtained neutron spectrum using one day data from the SciCRT under normal conditions. The spectrum is compared with the simulation from PARMA (observable events by the SciCRT). As it may be seen, the unfolded spectrum is in reasonable agreement with

the spectrum from PARMA. Using again the two sample Kolmogorov-Smirnov test we obtain a D -statistic = 0.09 and p -value = 0.81 and conclude there is no evidence to reject the null hypothesis. The difference in the high energy region is the result of low statistics, since these events are less probable.

The last step in our analysis is the estimation of the angular distribution. We use the SciCRT data from the same date as before for our calculation (November 14 2019). Fig. 15 shows the resulting distribution, compared with the theoretical model (Eq. (6)). The maximum value in the SciCRT data is given at 35° whereas the model has a maximum at 33° . The goodness fit test was calculated for values of zenith angle lower than 50° (p -value of 0.96). The shape of the distribution is expected from the discussion in the previous section.

In general the real data set is affected by the contamination of γ -rays and other factors not included in the simulation. Some of these include the errors in the conversion gain from ADC values to energy deposition, the fluctuation of the gain of the MAPMT and variations in the pedestal level. Study of these effects will be subject of future research.

6. Conclusion

The SciCRT detector is a new and improved cosmic ray detector aimed at observing solar energetic particles. The analysis of its characteristics is indispensable to study the energy spectrum of solar neutrons.

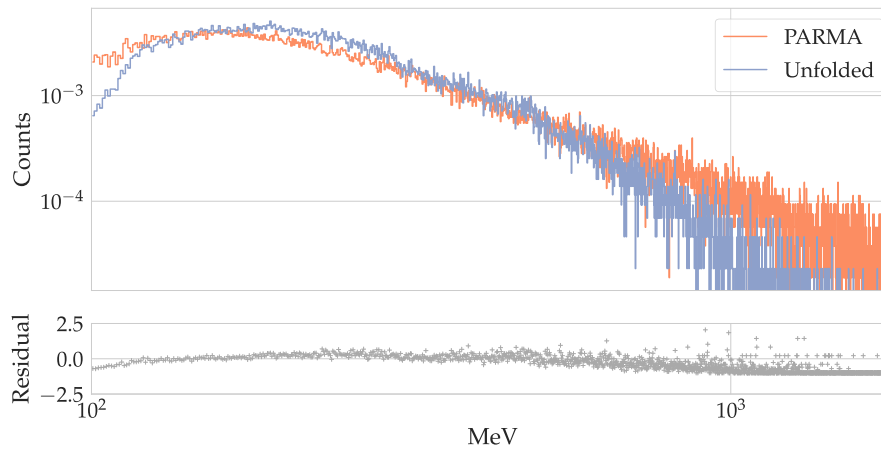


Fig. 14. The neutron energy spectrum estimated by ANN using SciCRT data from November 14 2019. We compared this with the neutron spectrum from the PARMA simulation. The histograms have a bin size of 1 MeV. The panel at the bottom shows the residuals.

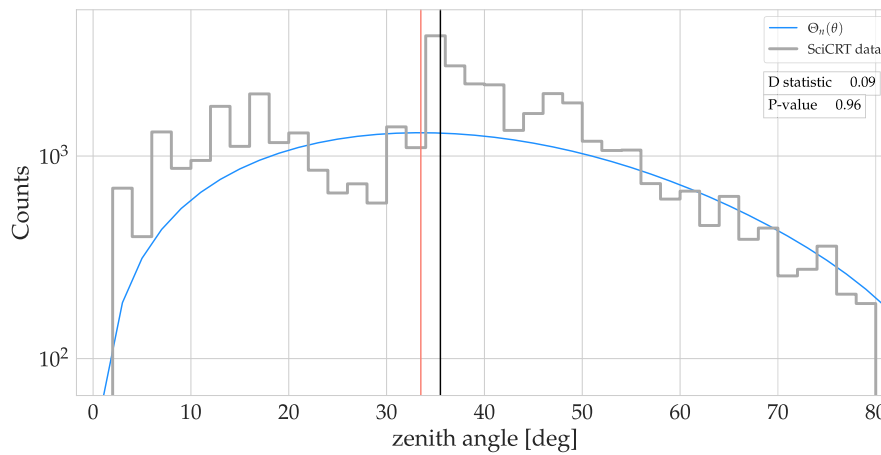


Fig. 15. Zenith angle distribution from neutron data registered by SciCRT on November 14 2019, with a bin width of 2°. The angle where the distribution is maximum (black line) differs 2° in relation to the model (red line). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

This provides a window of new opportunities to investigate particle acceleration mechanisms at the solar surface.

Regarding this, our present work shows the development of machine learning methods with capabilities to discriminate γ -rays from neutrons with high accuracy (91 %), which in turn increases the sensitivity to solar neutrons. Besides, the proposed neural network model will be useful for collecting information about the spectrum of solar neutrons at the top of the detector. The results presented in this work show that the ANN model is capable of reproducing the neutron spectrum. This was tested using a Kolmogorov–Smirnov test for goodness of fit. Concerning the zenith angular distribution the Fourier method allows to identify the direction of the registered radiation, making it possible to determine if the radiation is coming from the Sun or has another origin. In summary we have succeeded in developing an analysis method using machine learning tools for the SciCRT.

CRedit authorship contribution statement

R. Garcia: Methodology, Formal analysis, Investigation, Writing - original draft. **M. Anzorena:** Software, Methodology, Investigation. **J.F. Valdés-Galicia:** Writing - review & editing, Project administration, Funding acquisition, Supervision. **Y. Matsubara:** Writing - review & editing, Project administration, Funding acquisition, Supervision. **T. Sako:** Writing - review & editing, Conceptualization, Supervision. **E. Ortiz:** Conceptualization, Writing - review & editing. **A. Hurtado:** Resources, Conceptualization. **R. Taylor:** Resources, Conceptualization. **O. Musalem:** Conceptualization, Writing - review & editing. **L.X.**

González: Conceptualization, Writing - review & editing. **Y. Itow:** Conceptualization, Writing - review & editing. **T. Kawabata:** Resources, Conceptualization. **K. Munakata:** Funding acquisition, Conceptualization, Supervision, Writing - review & editing. **C. Kato:** Conceptualization, Writing - review & editing. **W. Kihara:** Conceptualization, Writing - review & editing. **Y. Ko:** Conceptualization, Writing - review & editing. **S. Shibata:** Conceptualization, Writing - review & editing. **H. Takamaru:** Conceptualization, Writing - review & editing. **A. Oshima:** Conceptualization, Writing - review & editing. **T. Koi:** Conceptualization, Writing - review & editing. **H. Kojima:** Conceptualization, Writing - review & editing. **H. Tsuchiya:** Conceptualization, Writing - review & editing. **K. Watanabe:** Conceptualization, Writing - review & editing. **M. Kozai:** Conceptualization, Writing - review & editing. **Y. Nakamura:** Conceptualization, Writing - review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors wish to thank the SciBar and SciBooNE experiments for allowing the use of SciBar detector in the SciCRT experiment. We also want to thank the National Institute for Astrophysics, Optics and

Electronics (INAOE) for allowing us the installation of the detector at the top of Sierra Negra volcano and giving us access to all the facilities to operate at the mountain.

The authors also acknowledge Dr. K. Caballero-Mora and staff from LARCAD for allowing us the use of their computer cluster to run the simulation.

This work was partially supported by UNAM-PAPIIT, Mexico (IN104115) and CONACyT, Mexico (180727T). This work was also partially supported by Grants-in-Aid for Scientific Research (B) 22340054 in Japan and the joint research program of the Institute for Space–Earth Environmental Research (ISEE) of Nagoya University, Japan.

References

- [1] K. Shibata, T. Magara, Solar flares: Magnetohydrodynamic processes, *Living Rev. Sol. Phys.* 8 (1) (2011) 6, <http://dx.doi.org/10.12942/lrsp-2011-6>.
- [2] E.L. Chupp, et al., A direct observation of solar neutrons following the 0118 UT flare on 1980 June 21, *Astrophys. J.* 2 263 (1982) L95–L99.
- [3] L. Dorman, Solar Neutrons and Related Phenomena, *Astrophysics and Space Science Library*, 2010.
- [4] Y. Muraki, et al., Observation of solar neutrons associated with the large flare on 1991 June 4, *Astrophys. J.* 400 (1992) L75–L78.
- [5] Y. Nagai, et al., First cosmic-ray measurements by the SciCRT solar neutron experiment in Mexico, *Astropart. Phys.* 59 (2014) 39–46.
- [6] Y. Muraki, et al., Possible detection of solar gamma-rays by ground-level detectors in solar flares on 2011 March 7, *Publ. Astron. Soc. Japan* 72 (2) (2020) <http://dx.doi.org/10.1093/pasj/psz141>.
- [7] S. Shibata, Propagation of solar neutrons through the atmosphere of the Earth, *J. Geophys. Res.* 99 (1994) 6651–6665.
- [8] L. Dorman, et al., Numerical simulation and analytical description of solar neutron transport in the Earth's atmosphere, *J. Geophys. Res.* 104 (1999) 22417–22426, <http://dx.doi.org/10.1029/1999JA900182>.
- [9] K. Nitta, et al., The K2K SciBar detector, *Nucl. Instrum. Methods Phys. Res. A* 535 (2004) 147–151.
- [10] Y. Sasai, et al., A faster and more reliable data acquisition system for the full performance of the SciCRT, *Nucl. Instrum. Methods Phys. Res. A* 857 (2017) 50–57.
- [11] J. Allison, et al., Geant4 developments and applications, *IEEE Trans. Nucl. Sci.* 53 (1) (2006) 270–278, <http://dx.doi.org/10.1109/TNS.2006.869826>.
- [12] S. Agostinelli, et al., Geant4—a simulation toolkit, *Nucl. Instrum. Methods Phys. Res. A* 506 (3) (2003) 250–303, [http://dx.doi.org/10.1016/S0168-9002\(03\)01368-8](http://dx.doi.org/10.1016/S0168-9002(03)01368-8).
- [13] M. Hasegawa, Measurement of Neutrino Oscillation Parameters with Neutrino-Nucleus Interaction Studies in the K2K Experiment (Ph.D. thesis), Kyoto University, 2006.
- [14] M. Anzorena, et al., Simulation and experimental validation of optimum read-out electronics design for scintillator bar cosmic ray telescope, *Nucl. Instrum. Methods Phys. Res. A* 991 (2021) <http://dx.doi.org/10.1016/j.nima.2021.165019>.
- [15] T. Sato, Analytical model for estimating terrestrial cosmic ray fluxes nearly anytime and anywhere in the world: Extension of PARMA/EXPACS, *PLOS ONE* 10 (12) (2015) 1–33, <http://dx.doi.org/10.1371/journal.pone.0144679>.
- [16] E. Ortiz, et al., Observation of cosmic ray hadrons at the top of the Sierra Negra volcano in Mexico with the SciCRT prototype, *Adv. Space Res.* 58 (10) (2016) 2018–2025, <http://dx.doi.org/10.1016/j.asr.2016.02.008>.
- [17] S. Agosteo, et al., A telescope detection system for direct and high resolution spectrometry of intense neutron fields, *Radiat. Meas.* 85 (2016) 1–17, <http://dx.doi.org/10.1016/j.radmeas.2015.12.005>.
- [18] A.M. Khan, S. Ravi, Image segmentation methods: A comparative study, *Int. J. Soft Comput. Eng.* 3 (4) (2013).
- [19] D. Arthur, S. Vassilvitskii, k -means++: the advantages of careful seeding, in: *Proc. of the 18th Annual ACM-SIAM Symposium on Discrete Algorithms*, 2007, pp. 1027–1035.
- [20] T. Cover, P. Hart, Nearest neighbor pattern classification, *IEEE Trans. Inform. Theory* 13 (1) (1967) 21–27.
- [21] J. Gareth, et al., *An Introduction to Statistical Learning: With Applications in R*, Springer, 2013.
- [22] P.V.C. Hough, Method and means for recognizing complex patterns, 1962.
- [23] B. Beylkin, Discrete radon transform, *IEEE Trans. Acoust. Speech Signal Process.* 35 (2) (1987) 162–172.
- [24] P. Tsitsipis, et al., Fast estimation of slopes of linear and quasi-linear structures in noisy background, using Fourier methods, *Pattern Recognit.* 40 (2) (2007) 563–577, <http://dx.doi.org/10.1016/j.patcog.2006.04.014>.
- [25] B. Olmos, A. Aguilar, A method to measure the integral vertical intensity and angular distribution of atmospheric muons with a stationary plastic scintillator bar detector, *Nucl. Instrum. Methods Phys. Res. A* 987 (2021) <http://dx.doi.org/10.1016/j.nima.2020.164870>.
- [26] M.B. Chadwick, et al., ENDF/B-VII.0: Next generation evaluated nuclear data library for nuclear science and technology, *Nucl. Data Sheets* 107 (12) (2006) 2931–3060, <http://dx.doi.org/10.1016/j.nds.2006.11.001>.
- [27] S. Avdic, et al., Detector response unfolding using artificial neural networks, *Nucl. Instrum. Methods Phys. Res. A* 565 (2) (2006) 742–752, <http://dx.doi.org/10.1016/j.nima.2006.06.023>.
- [28] S. Hosseini, Neutron spectrum unfolding using artificial neural network and modified least square method, *Radiat. Phys. Chem.* 126 (2016) 75–84, <http://dx.doi.org/10.1016/j.radphyschem.2016.05.010>.
- [29] F. Chollet, et al., Keras, 2015, <https://keras.io>.
- [30] M. Abadi, et al., TensorFlow: Large-scale machine learning on heterogeneous systems, 2015, software available from tensorflow.org URL <https://www.tensorflow.org/>.
- [31] I. Goodfellow, et al., *Deep Learning*, MIT Press, 2016, <http://www.deeplearningbook.org>.
- [32] Stathakis D., How many hidden layers and nodes? *Int. J. Remote Sens.* 30 (8) (2009) 2133–2147, <http://dx.doi.org/10.1080/01431160802549278>.
- [33] P. Ramachandran, et al., Searching for activation functions, in: *CoRR*, 2017, arXiv:abs/1710.05941.
- [34] Y. Sasai, et al., Performance of the SciBar cosmic ray telescope (SciCRT) toward the detection of high-energy solar neutrons in solar cycle 24, *Earth Planets Space* 66 (2014) 130, <http://dx.doi.org/10.1186/1880-5981-66-130>.
- [35] A. Tsuchiya, Solar Neutron Search using Energy Analysis Method in the SciCRT (Master's thesis), Nagoya University, 2017.