Improving reconstruction methods for radio measurements with Tunka-Rex

P.A. Bezyazeekov[a], N.M. Budnev[a], O. Fedorov[a], O.A. Gress[a], A. Haungs[b], R. Hiller[b], T. Huege[b],
Y. Kazarina[a], M. Kleifges[c], E.E. Korosteleva[d], D. Kostunin[b], O. Krömer[c], V. Kungel[b],
L.A. Kuzmichev[d], V. Lenok[a], N. Lubsandorzhiev[d], T. N. Marshalkina[a], R.R. Mirgazov[a], R. Monkhoev[a],
E.A. Osipova[d], A. Pakhorukov[a], L. Pankov[a], V.V. Prosin[d], F.G. Schröder[b], A. Zagorodnikov[a]

[a]*Institute of Applied Physics ISU, Irkutsk, Russia;*
[b]*Institut für Kernphysik, Karlsruhe Institute of Technology (KIT), Germany;*
[c]*Institut für Prozessdatenverarbeitung und Elektronik, KIT, Germany; and*
[d]*Skobeltsyn Institute of Nuclear Physics MSU, Moscow, Russia;*

Tunka-Rex is a detector for radio emission produced by cosmic-ray air showers. It is located in Siberia, and triggered by Tunka-133, a co-located air-Cherenkov detector during night, and by the scintillator array Tunka-Grande during day. Tunka-Rex demonstrates that the radio technique can provide a cost-effective extension of existing air-shower arrays. Operating in the frequency range of 30-80 MHz, Tunka-Rex is limited by the galactic background and suffers from the local radio interferences. We investigate the possibilities of improving measured data using different approaches, particularly, a multivariate treatment of background is considered, as well as an improved likelihood fit of the lateral distribution of amplitudes.

I. INTRODUCTION

Tunka-Rex [1, 2] is an antenna array for the detection of radio emission of extensive atmospheric showers (EAS) created by cosmic rays [3]. It works jointly with the non-imaging air-Cherenkov light detector Tunka-133 [4] and the scintillators of Tunka-Grande [5] and receives triggers from both of them. Comparing to classical optical methods of EAS detection, detection of radio emission is interesting because it is independent of the time of the day (measurements are possible during sunlight) and of weather conditions (except thunderstorms). Consequently, radio has higher duty-cycle than optical methods. Tunka-Rex antennas are stable and cost-effective and able to operate in a sparse large-scale configuration [6, 7].

At the present moment Tunka-Rex consists of 63 antennas covering an area of about 3 km². Each Tunka-Rex antenna station consists of two perpendicularly aligned SALLAs (Short Aperiodic Loaded Loop Antenna) [8, 9] with 120 cm diameter mounted on a wooden pole on a height of about 2.5 m. At the top they are connected to a low noise amplifier (LNA). This LNA is connected to a filter-amplifier via 30 m coaxial RG213 cables. THe filter-amplifier is connected to the Tunka-133 or Tunka-Grande FADCs with a short 1 m RG52 cable. The sampling rate of the FADCs is 200 MHz, and the length of the trace is 1024 samples with a bit depth of 12 bit.

Tunka-Rex operates in an environment with low signal-to-noise ratio (SNR) conditions, that requires a sophisticated post-processing of recorded traces. The standard approach is to use different filters (broadband or narrowband) to increase the SNR. Then, the signal peak is defined as maximum of the filtered trace. The aim of the present work is developing and testing new methods for the improvement of signal reconstruction and for fitting the lateral distribution of the radio amplitudes.

II. STANDARD PROCEDURE OF DATA RECONSTRUCTION

For signal reconstruction Tunka-Rex uses a modified radio extension of the Offline software framework developed by the Pierre Auger Collaboration [10]. Raw Tunka-Rex data consist of ADC traces from the detector containing measured amplitudes in two orthogonal polarizations. The signal reconstruction includes following steps:

1. Upsampling of recorded traces by a factor of 4, and applying the following filters: first, a band-stop filter suppresses narrowband interferences occurring each 5 MHz; second, a bandpass filter, restricts the band to 35–76 MHz;

2. The amplitude of the measured signal S is defined as maximum of Hilbert envelope in the signal window of the recorded trace;

3. The noise level N is defined as RMS of amplitudes in the noise window.

The SNR is calculated as square of ratio between measured signal and noise:

$$\text{SNR} = S^2/N^2 \,. \qquad (1)$$

Antenna stations with SNR < 10 in at least one of the channels are rejected.

The rest of the stations is used for the reconstruction of the arrival direction by fitting peak times with a plane wavefront model. Knowing the arrival direction (shower axis), the electrical field is reconstructed by applying the antenna pattern and assuming that the electrical field along the shower axis is zero. Then

the cut $\text{SNR} \geq 10$ is applied again for the vectorial traces of the electrical field on the antenna station.

Let us consider the influence of the noise on the measured signal. The signal measured at the frequency ν is defined as sum of the true signal and the noise contribution:

$$A_m^\nu e^{i\phi_m} = A_t^\nu e^{i\phi_t} + A_n^\nu e^{i\phi_n}, \qquad (2)$$

where A_m^ν, A_t^ν, A_n^ν are the amplitudes of the measured, true and noise signals, in the frequency domain, and ϕ_m, ϕ_t, ϕ_n are their phases. The unknown phase of the noise component can shift resulting amplitude by $\pm 30\%$ for $\text{SNR} = 10$.

The contribution of the noise to the total power of measured signal is corrected using the parameterization:

$$A_t = f(\text{SNR})A_m, \qquad (3)$$

where

$$f(\text{SNR}) = \sqrt{1 - k/\text{SNR}}, \qquad (4)$$

where parameter k is extracted from a fit to simulations (details in [6]), and then this parameterization is applied to the measured signals. Besides this effect, the measured amplitude is characterized by the uncertainty defined as standard deviation σ_n of $(A_m - A_t)/A_m$. This uncertainty is parameterized by LOPES [11] as a function of SNR (see Fig. 1):

$$\sigma_n = \frac{a}{\sqrt{\text{SNR} - a^2}} \left(b + c \exp(-\sqrt{\text{SNR}}/a) \right), \qquad (5)$$

where $a = 0.602$, $b = 0.616$ and $c = 0.213$ are dimensionless parameters. This uncertainty is included in the chi-square fit of the lateral distribution function (LDF) [12]:

$$\mathcal{E}(r) = \mathcal{E}_0 \exp \left(a_1(r - r_0) + a_2(r - r_0)^2 \right), \qquad (6)$$

where $\mathcal{E}(r)$ is the amplitude at the antenna station with distance r from the shower axis, $\mathcal{E}_0$ is a normalization factor proportional to the energy of the primary cosmic ray, r_0 is an arbitrary parameter chosen to obtain maximum correlation with the primary energy and the distance to the shower maximum, a_1 and a_2 are parameters proportional to the slope and width of the LDF, respectively.

III. SIGNAL RECONSTRUCTION WITH A NEURAL NETWORK

For the design of the neural network we use the PyBrain library [13] written in Python. At the first stage we have trained a test network which performs the classification of the transient background in a signal window. It is simple "perceptron" (neural network with single hidden layer) trained on a dataset
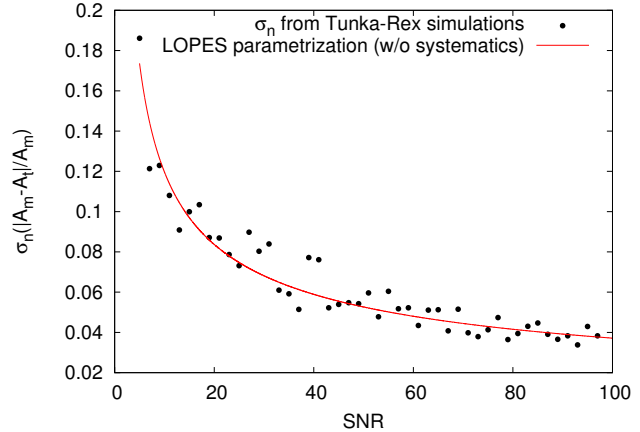

FIG. 1: Dependence of the uncertainty of the signal reconstruction on the SNR. The points are obtained as standard deviation of normalized difference between simulated and reconstructed amplitudes using the Tunka-Rex noise library and hardware response (the spread between neightboring bins is due to low statistics); the curve is obtained from a LOPES parameterization (see Eq. 5), and is in agreement with Tunka-Rex simulations.

containing separated samples containing pure noise and samples with background pulses The network returns two values: first the probability of pure noise and, second, the probability of background pulse (with high SNR). Resulting efficiency of that classification is 98%. For the amplitude reconstruction we designed another type of neural network, with the following structure:

- 200 input neurons

- 3 hidden layers, with each 500 neurons

- 1 output neuron

The activation function of the neurons is sigmoid, for training we use resilient propagation method. The input layer takes the Hilbert envelope of trace with 200 ADC counts of the trace centered around the assumed signal as input. The output neuron returns the reconstructed amplitude. As training dataset we simulated radio pulses with CoREAS [14] (for two Tunka-Rex channels) and then distort them using a background library collected by the Tunka-Rex experiment. The example of traces is given in Fig. 2. Training dataset contains about 7000 distorted traces with different SNR levels and known amplitude of true (simulated) signal. In each training epoch, the network takes 200 values of distorted traces as input and the value of the true amplitude as output.

For training and test of the neural network we selected only traces containing a true signal higher than the noise level (i.e. $\text{SNR} > 1$). Thus, network is not able to distinguish between pure noise and low-SNR signals, but is able to reconstruct amplitude for signals with $\text{SNR} > 1$.
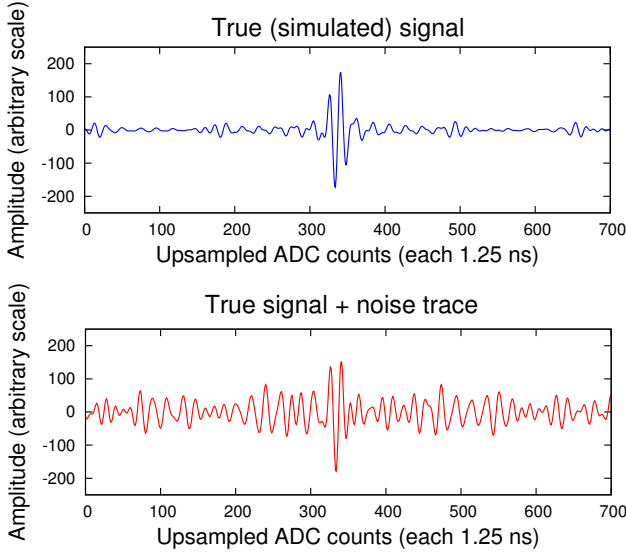
FIG. 2: Example of ADC trace with clear (top) and distorted (bottom) signal.
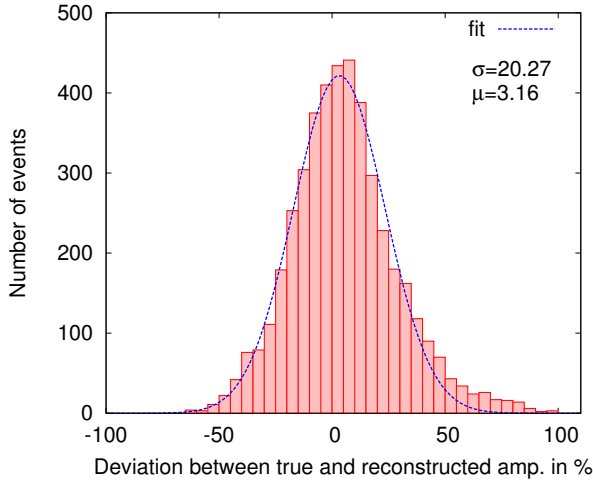

FIG. 3: Deviation between true and reconstructed amplitude given by the neural network for signals with SNR < 5.

Running the designed neural network on the testing dataset has shown that the uncertainty of the amplitude reconstruction is about 20%, which is worse, than given by our standard methods. However, the same uncertainty is obtained for very low SNRs, and the correction for the bias due to noise is obtained automatically. The distribution of $(A_m - A_t)/A_m$ for signals with SNR < 5 processed with the neural network is given in Fig. 3.
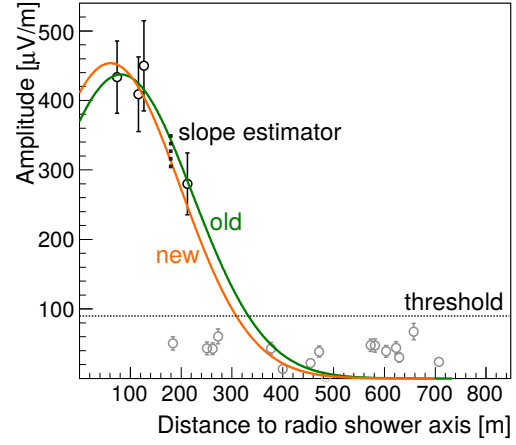
FIG. 4: Difference between best LDF fits using standard (old) and modified (new) likelihood functions. Although the visual change of the slope is not very significant, the reconstructed depth of shower maximum has changed by about 100 g/cm^2 in this case. Let us note, that such events are not passing standard X_{max} quality cuts of Tunka-Rex reconstruction.

IV. OPTIMIZATION OF LIKELIHOOD LDF FIT

For the fitting of lateral distribution function (LDF) the standard chi-square fit is used, where uncertainties are defined as a function of SNR (see Sec. 2). This way, the amplitudes from antennas stations with SNR > 10 are included in the fit and weighted with their uncertainties. The information from the antenna stations without signal was not used. However, they still contain information about the confidence levels of the signal amplitudes, i.e. these stations provide an upper limit for the amplitude. They are weighted according to the general threshold of signal detection if the fitted LDF is higher than this threshold, and the get zero weigth if the LDF is already below the threshold.

Thus, we modified the LDF fit and defined a likelihood function as follows:

$$
\mathcal{L} = \prod_{i=0}^{N} \exp\left\{ -\left[\frac{(f_i - y_i)^2}{2\sigma_i^2}\Theta(y_i - y_{th}) \right.\right.
$$
$$
\left.\left. + \frac{(f_i - y_{th})^2}{2y_{th}^2}\Theta(f_i - y_{th})\Theta(y_{th} - y_i) \right]\right\}, \quad (7)
$$

and the chi-square becomes

$$
\chi^2 = \sum_{i=0}^{N}\left[\frac{(f_i - y_i)^2}{\sigma_i^2}\Theta(y_i - y_{th}) \right.
$$
$$
\left. + \frac{(f_i - y_{th})^2}{y_{th}^2}\Theta(f_i - y_{th})\Theta(y_{th} - y_i) \right], \quad (8)
$$

where f_i and y_i are the fitted and measured values of the amplitudes, σ_i is the uncertainty of i-th station of N, $\Theta(x)$ is the Heaviside step function.

One can note, that antenna stations with amplitudes $y_i < y_{th}$ are weighted with y_{th}, where $y_{th} \approx 90$ μV/m is the threshold of signal detection [15]. Since $y_{th} \geq 10\sigma_i$ for SNR > 10 the weights of the antenna stations without signals are relatively small. The first tests of this modified LDF fitting have shown, that the reconstruction of air-shower parameters is slightly improved, however this simple parameter-free weighting can be further developed. An example of a modified fit is given in Fig. 4. It is worth noticing, that only reconstruction of low-quality events (e.g. not passing $X_{\max}$ quality cut, see Ref [6] for details) is significantly affected, the result of the reconstruction of the standard event set is almost the same. Which means, that such improvement of LDF fitting can increase the statistics near threshold.

V.   SUMMARY

We present our current progress on the improvements of reconstructing radio events from cosmic-ray air-showers. We have worked in two directions: reconstruction of the low-SNR radio pulses using neural networks, and improving the fit of the LDF by including stations with amplitudes below the SNR cut.

- *Low-SNR signal reconstruction.* This study shows that machine learning has a potential for solving tasks related to multi-variable conditions, like radio background. For high SNRs the resulting precision of the reconstructed amplitude is worse comparing to our standard methods, however the neural network shows better performance at low SNRs (< 5). We plan to continue this study by testing different configurations of neural networks and another machine learning methods, e.g decision trees.

- *Modification of LDF fit.* We added weights of antenna stations without signal into the standard chi-square fit. The simple parameterization of the weights shows a slight improvement of the air-shower reconstruction. We plan to tune parameters in order to optimize weighting of the antenna stations without signal. After this we include this stations also for shower core reconstruction with radio.

[1] P.A. Bezyazeekov, et al. - Tunka-Rex Coll., Nucl. Inst. Meth. A **802**, 89 (2015).
[2] D. Kostunin, et al. - Tunka-Rex Coll., EPJ Web of Conferences **proc. of ISVHECRI 2016, in press** (2016).
[3] Schröder, Frank G., Prog. Part. Nucl. Phys. **93**, 1 (2017), 1607.08781.
[4] V.V. Prosin et al. (Tunka-133 Collaboration), Nucl.Instrum.Meth. **A 756**, 94 (2014).
[5] N.M. Budnev et al., Bull. Russ. Acad. Sci. Phys. **79**, 395 (2015), [Izv. Ross. Akad. Nauk Ser. Fiz.79,no.3,430(2015)].
[6] P.A. Bezyazeekov, et al. - Tunka-Rex Coll., JCAP **01**, 052 (2016).
[7] W. D. Apel et al. (Tunka-Rex, LOPES), Phys. Lett. **B 763**, 179 (2016).
[8] O. Krömer et al. (LOPES Collaboration), Proc. of the 31st ICRC, Łódź, Poland (2009), http://icrc2009.uni.lodz.pl/proc/html/.
[9] P. Abreu et al. (Pierre Auger), JINST **7**, P10011 (2012), 1209.3840.
[10] P. Abreu et al. (Pierre Auger), Nucl.Instrum.Meth. **A 635**, 92 (2011), 1101.4473.
[11] F. Schröder et al., NIM A **662, Supplement 1**, S238 (2012), ISSN 0168-9002, proc. of ARENA 2012, URL http://www.sciencedirect.com/science/article/pii/S0168900210024514.
[12] D. Kostunin et al., Astropart. Phys. **74**, 79 (2016).
[13] T. e. a. Schaul, Journal of Machine Learning Research **11**, 743 (2010).
[14] T. e. a. Huege, AIP Conf.Proc. **1535**, 128 (2013), 1301.2132.
[15] R. Hiller, et al. - Tunka-Rex Coll., EPJ Web of Conferences **proc. of ARENA 2016, in press** (2016), arxiv:1611.09614.