Studies on the unfolding of the atmospheric neutrino spectrum with IceCube 59 using the TRUEE algorithm

THE ICECUBE COLLABORATION[1]
[1]*See special section in these proceedings*

Abstract: The measurement of the atmospheric neutrino energy spectrum provides information about the diffuse neutrino flux from extragalactic sources. A relative increase of the spectrum toward higher energies could be evidence for neutrino producing hadronic processes in the cosmic high energy accelerators, such as active galactic nuclei or gamma ray bursts. IceCube is currently the largest neutrino detector on Earth and is placed in the antarctic ice at the geographic South Pole. IceCube permits the detection of neutrinos with energies beyond 10^6 GeV. Since the acceptance and the resolution of neutrino telescopes suffer from the finite resolution and limited acceptance, a regularized unfolding method is used to extract the energy distribution of neutrinos from the measured observables. For AMANDA, the unfolding was done with the $\mathcal{RUN}$ algorithm. Based on the basic concept of this program and for data analyses in the ROOT frame, a new deconvolution algorithm (TRUEE) has been written and tested. With this new algorithm, studies on the analysis of the atmospheric neutrino spectrum measured with the IceCube 59 string configuration will be presented.

Corresponding authors: Natalie Milke[2] (*natalie.milke@udo.edu*), Wolfgang Rhode[2] (*wolfgang.rhode@udo.edu*), Tim Ruhe[2] (*tim.ruhe@udo.edu*)
[2]*Department of physics, TU Dortmund University, D-44221 Dortmund, Germany*

Keywords: IceCube; TRUEE; RUN; regularized unfolding; atmospheric neutrino; energy spectrum

1 Introduction

IceCube is the largest neutrino detector ever built and is located at the geographic South Pole. It consists of 5160 digital optical modules (DOM) arranged along 86 strings forming a three-dimensional grid covering a cubic kilometer in the glacial ice [1]. While traveling through the ice the high energy neutrino-induced muons produce Cherenkov light which can be detected by the DOMs providing directional and energy information of the muon track. One of the main goals of IceCube is the detection of extragalactic neutrinos for understanding of cosmic ray production in cosmic accelerators. Neutrinos from interactions of cosmic rays with the Earth's atmosphere represent a background for the extragalactic neutrinos. Thus, a precise measurement of the atmospheric neutrino flux is important for understanding this background. Since the spectral index of the flux distribution depending on neutrino energy is lower for extragalactic neutrinos (following the spectral behavior of Fermi accelerated cosmic rays $\gamma \sim 2$ [2]) than for atmospheric neutrinos ($\gamma \sim 3.7$) [3], a contribution of extragalactic neutrinos would cause an enhancement of the flux in the high energy region of the spectrum.

The energy of the primary particles is convoluted with the interaction probability and the detector finite resolution and limited acceptance. Therefore the neutrino energy has to be estimated from energy-correlated, measured observables. For this purpose a regularized unfolding algorithm is developed and applied.

In this paper the atmospheric neutrino sample from the measurement with the IceCube 59 (IC 59) string configuration is used. The energy spectrum is unfolded with the new deconvolution algorithm TRUEE.

2 Regularized unfolding

The convolution of the neutrino energy with the interaction probability and detector response gives us the measured observables in the detector and this relation can be expressed as a Fredholm integral equation of the first kind if neglecting background. From discretization a linear matrix equation can be obtained where the measured distribution is a product of the detector response matrix and the neutrino energy distribution. The response matrix is obtained from Monte Carlo (MC) simulation. Thus, an unfolding algorithm needs as input MC-simulated assumed energy distribution with the resulting distributions of measured observables to determine the detector response and the measured observables distributions from data to estimate the neutrino energy flux.

Every unfolding requires an a-priori assumption about some properties of the result realized in a regularization to reduce strong negative correlations between unfolded data points. Here the Tikhonov regularization [5] is used achieving a smooth distribution by minimizing the curvature of the result during the unfolding fit.

3 Unfolding algorithm TRUEE

In AMANDA, the precursor of IceCube, the **R**egularized **UN**folding ($\mathcal{RUN}$) algorithm [6] was used to unfold lepton energy spectra [7]. $\mathcal{RUN}$ was developed in 1984 using the programming language FORTRAN 77. Therefore $\mathcal{RUN}$ is not easy to install and use in combination with modern software. TRUEE - **T**ime-dependent **R**egularized **U**nfolding for **E**conomics and **E**ngineering problems or just **TRUE E**nergy is a new software including the $\mathcal{RUN}$-based unfolding algorithm.

One special property of the $\mathcal{RUN}$ algorithm is the parametrization of unfolded distribution using a superposition of cubic basis splines. The spline coefficients are determined from the unfolding and the superposed function is transformed to the final histogram. At the spline overlaping points (knots) the function is continuously differentiable up to the second derivative, so that a Tikhonov regularization with the second derivatives in the smoothing operator can be performed. The user determines the number of splines by defining the number of knots. Regularization is controlled by the number of degrees of freedom, which can be given by the user but can also be suggested by the software. A small number leads to strong regularization.

TRUEE has been developed within the Collaborative Research Centre SFB 823 in Dortmund. Besides the core regularized unfolding from $\mathcal{RUN}$ the new software contains user friendly functions, which make the procedure of an unfolding analysis more comfortable. The functions used in this analysis are explained in the next section using the IC 59 neutrino sample.

4 Unfolding analysis

For this analysis 10 % of the measured data were used. After event selection a sample of 3160 neutrino events measured within the zenith angle range of 88° to 180° with IC 59 is obtained. Thus, most of the events in the sample were caused by neutrinos having traveled through the Earth before undergoing an interaction inside or in the vicinity of IceCube. For 100 % of the data we expect more than 30000 neutrino-induced events and thus higher statistics in the high energy region. Based on a Monte Carlo study the purity of the sample is estimated to be higher than 99.3 %, therefore the background formed by atmospheric muons is neglected. The sample is obtained using straight precuts followed by an event selection using the multivariate method Random Forest [8] within the framework Rapid Miner [9].

4.1 Selection of observables

As a first step the selection of energy-dependent observables is made. TRUEE automatically produces scatter plots of the sought-after variable and observables and the related profile histograms to check if a correlation is present. The inspection of scatter plots with different observables showed the correlation between primary neutrino energy and the following observables:

- Number of DOMs having a signal
- Number of photoelectrons in all DOMs (number of hits)
- Total charge caused in DOMs
- Energy loss per unit track length

Figure 1 shows the correlation between neutrino energy and number of hits.

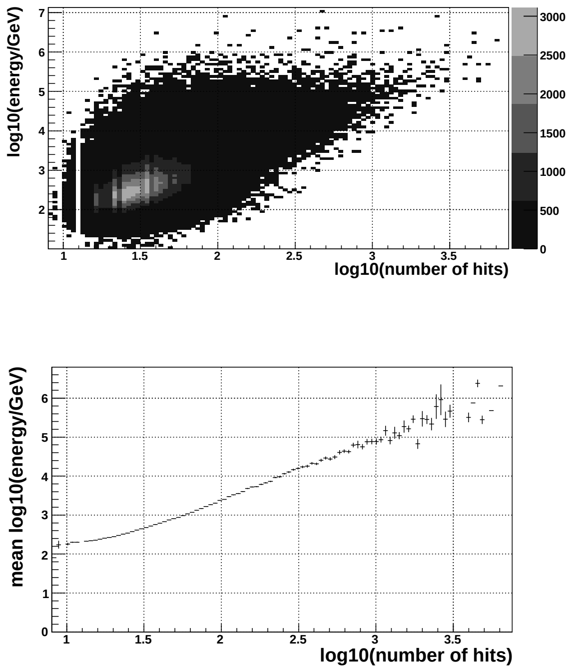

Figure 1: An example of scatter plot and related profile histogram to check the correlation between the energy and the observable (here number of hits). An optimal correlation is present in a monotonically changing profile function with small uncertainties.

Since TRUEE is able to use up to three observables at the same time for unfolding, different combinations of the four selected observables have been checked. The best results are obtained in the test mode (see Sec. 4.2) with the combination of number of hits and energy loss.

4.2 Test mode

An unfolding algorithm expects input from the user concerning some parameters such as the number of bins for histograms or the degree of regularization. To check which user-defined parameter settings give the optimal unfolding result a test mode is included in TRUEE. In this mode only simulations are used. Since we neglect the atmospheric muon background, the MC sample contains only neutrino events after application of event selection techniques. The energy distribution of simulated neutrino events has been reweighted, so that the generated flux follows the atmospheric neutrino flux predicted by Honda [3] containing a prompt component from Naumov Recombination Quark Parton Model (Naumov RQPM) [4]. The prompt component consists of neutrinos from decays of short-lived mesons containing charm quarks.

An MC sample that is statistically equivalent to the expected experimental data sample is used as a pseudo real data sample for unfolding. Since the real sought-after distribution is known in this case, it can be compared to the unfolded distribution. The optimal parameter settings are chosen with an L-curve approach [10] by examining the tradeoff between regularization strength and fit to the true distribution. For the IC 59 sample the following parameter settings are used for the final unfolding:

- Number of hits with 24 bins
- Energy loss with 4 bins
- Number of knots 15
- Number of degrees of freedom 4

The bin number for number of hits is obtained using the resolution of the observable (binsize $\sim 0.5*$standard deviation). The second observable serves as additional constraint and its bin number is low because of limited statistics. The unfolded test spectrum is shown in Fig. 2.

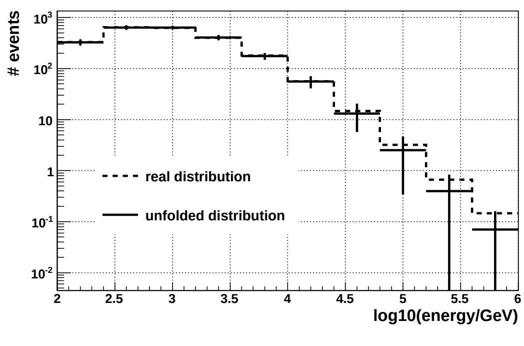

Figure 2: Unfolding of pseudo data compared to the real distribution using test mode. The unfolding does not consider the interaction probability, acceptance or systematic uncertainties.

4.3 Unfolding result

The unfolding procedure with the parameter settings determined in Sec. 4.2 can now be applied to the IC 59 neutrino sample. The generated MC neutrino sample for determination of the detector response contains only simulated events that undergo an interaction within or close to the detector. This procedure is necessary to reduce simulation time and memory. After passing all event selection steps the final sample contains only a fraction of neutrino events. Thus, the unfolded distribution represents only neutrinos which interacted, triggered the detector and passed the event selection (Fig. 3).

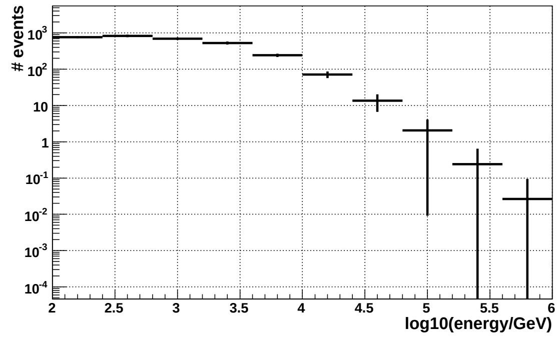

Figure 3: Unfolding of IC 59 data gives the distribution of selected neutrino events depending on energy. The unfolding does not consider the interaction probability, acceptance or systematic uncertainties, yet. Furthermore with the full data sample the number of events per bin will increase by a factor of ten.

To calculate the neutrino flux for all neutrinos within the zenith angle range, the unfolded spectrum has to be divided by the effective area. The effective area is the ratio of observed event rate and incoming flux and depends on the properties of the selected event sample. It includes the muon neutrino cross section, the probability for the muon to be detected and the detector efficiency for muon detection and event reconstruction. The effective area for the current sample is shown in Fig. 4.

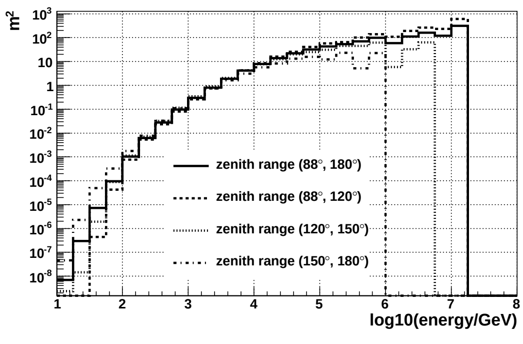

Figure 4: Effective area for the current neutrino sample dependent on neutrino energy. Illustrated are areas for different zenith angle ranges and for the average of the whole zenith range considered in the analysis.

The effective area is rising at higher energies due to the increasing cross section of neutrinos and to the higher length of the muon tracks. Therefore the probability to detect and reconstruct such a long track is rising. For the events with vertical upgoing tracks the effective area is decreasing because of the rising probability for absorption of neutrinos by the Earth.

We demonstrate the performance of the unfolding technique by showing an example in Fig. 5 of how an unfolded energy distribution (Fig. 3) can be translated into a neutrino flux spectrum when the effective area (Fig. 4) is known. Additional spectra have been unfolded with the same parameter settings but with different assumptions of the neutrino flux in MC to check the possible bias introduced by the assumption. Shown are results trained with MC weighted to atmospheric (Honda), to atmospheric with prompt (Honda-Naumov) and to atmospheric with prompt and $1.6 \cdot 10^{-8} E^{-2}$ neutrino flux. All three results have only small deviation in the low statistics region, thus the introduced bias is negligible.

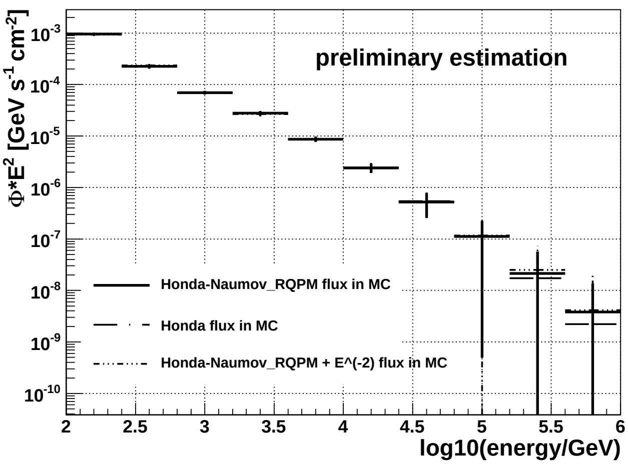

Figure 5: An example of the atmospheric neutrino energy spectrum from 10 % of IC 59 data unfolded with TRUEE. Shown are three unfolding results, using different MC distributions to determine the detector response. The uncertainties take into account statistics and bin-to-bin correlation, determined by the unfolding software. The spectrum is weighted by squared energy for a better illustration.

Since a continuous function is unfolded taking into account event migration between bins, the last bin is estimated to be non-zero even though the statistics of current sample is low.

4.4 Verification of simulation

A function to verify the result was developed in $\mathcal{RUN}$ and was transferred to TRUEE. The user has the possibility to check whether the simulation of all observables agrees with the experimental data and thus verify the unfolding result. After the unfolding, the MC events are reweighted by the unfolded distribution. The MC sample describes the real data now. In this case all observables, not only those which have been used for the unfolding, should match in their distribution the measured data. The observables whose distributions do not match are not correctly simulated. If none of the distributions match, probably the unfolding did not work properly. In this case the user should check if the simulation of the detector response was correct. For the unfolding example shown in this paper the verification showed an agreement between the experimental data and reweighted simulations.

5 Conclusion and outlook

The new unfolding algorithm TRUEE shows a good performance in estimation of an atmospheric neutrino spectrum. The algorithm is able to estimate a steep distribution covering several orders of magnitude and thus is a dedicated tool for astroparticle physics. The analysis is facilitated by additional functions and an easy installation and ease of use of the software.

A new energy region of the atmospheric neutrino flux can be explored with IC 59. The simulation predicts an extension to energies up to 10^6 GeV. The estimation of the neutrino energy spectrum with 10 % of the IC 59 data will be done by unfolding with TRUEE. The subsequent unfolding of the 100 % IC 59 data sample is expected to determine the energy spectrum with more precision in the high energy region due to higher statistics.

6 Acknowledgements

This work is supported by the German Research Foundation DFG (SFB 823/C4).

References

[1] F. Halzen, S. R. Klein, Rev. Sci. Instrum., 2010, **81**(081101)
[2] E. Fermi, Phys. Rev., 1949, **75**(8): 1169-1174
[3] M. Honda et. al., Phys. Rev. D, 2007, **75**(4): 043006
[4] E. V. Bugaev et. al., Il Nuovo Cimento C, 1989, **12**(1): 41-73
[5] A. N. Tikhonov: 1963, On the solution of improperly posed problems and the method of regularization, Sov. Math. 5
[6] V. Blobel: 1996, The RUN Manual - Regularized Unfolding for High-Energy Physics Experiments, Technical Note TN 361, OPAL
[7] R. Abbasi et. al., Astroparticle Physics, 2010, **34**: 48-58
[8] L. Breiman, Machine Learning, 2001, **45**(1): 5-32
[9] S. Fischer et. al., 2002, Technical Report, Collaborative Research Center 531, University of Dortmund
[10] C. L. Lawson, R. J. Hanson: 1974, Solving Least Squares Problems, Prentice-Hall, Englewood Cliffs, NJ