Search for Neutrino Emission at the Galactic Center Region with IceCube

The IceCube Collaboration
(a complete list of authors can be found at the end of the proceedings)

E-mail: xk35@icecube.wisc.edu

The Galactic Center (GC) region has long been subject to intense interest within the astrophysics community. At the GC, a strong radio source has been detected and identified as the super-massive black hole, Sgr. A*, which significantly affects the dynamics of the GC region within 300 parsecs. Many interesting astrophysical objects, including supernova remnants, high-mass X-ray binaries, pulsar wind nebulae, among others, have been found in this region. These objects are expected to produce high energy neutrinos, possibly in a transient manner, which can potentially be observed with the IceCube Neutrino Observatory. Due to the geographical location of the IceCube Neutrino Observatory, a large amount of cosmic-ray muons reduces IceCube’s sensitivity in the southern sky, where the GC is located. Thus, a dedicated event selection is performed to reduce this background and optimize the sensitivity. In this talk, we present this event selection in the GC region with IceCube data using classical and machine learning methods. This new dataset will be used to perform a time-dependent search for single flares at the GC.

Corresponding author: Xinyue Kang*

Drexel University

* Presenter
1. Introduction

The IceCube Neutrino Observatory is a particle detector designed to detect neutrinos. It is built deep within the ice of the South Pole to observe the cosmos. The IceCube detector consists of 5160 digital optical modules (DOMs). The 86 cables holding the DOMs are arranged in the form of a hexagonal grid 1.5 to 2.5 kilometers deep in the ice. Neutrinos interact with the nuclei in the ice and rock. Such interactions create particle cascades from the hit nuclei as well as charged particles such as electrons, muons, and taus. Smaller particle showers produced by the leptons along their way emit Cherenkov radiation. The DOMs collect photons and turn them into electrical signals. The collected information is used to reconstruct properties of the incoming event, for example the energy, direction, length, and flavor. This analysis starts from selecting events using the reconstructed information to reduce background. The selection uses 12 years of data focusing on the GC region. Traditional and machine learning methods were used in this selection. A final event rate of $50.47 \text{ mHz}$ is obtained for this selection.

With the recent evidence for a signal of astrophysical neutrinos from the Galactic Plane [1], the emission from the GC becomes particularly interesting. The GC is a region that promises high activity based on the detection of a PeVatron in the GC [2] and the presence of a supermassive black hole (SMBH) at the position of Sgr A* [3]. SMBHs can be sources of flare-like emission of cosmic rays and their secondaries, neutrinos and gamma-rays. This motivates our search for a single flare from the GC with the created dataset. Previous analyses in IceCube used similar methods to find time-dependent clustering of neutrino events in the direction of the blazar TXS 0506+056 [4] [5]. This time-dependent analysis focuses on the GC as a point source to search for potential bursts of neutrino events in this dense region. In the following sections, we present the methods and results of the event selection and time-dependent point source analysis for the GC.

2. Motivation

The existing IceCube datasets are mostly looking at the whole sky or southern sky optimized for an energy spectrum of $E^{-2.0}$. These event selections are not sensitive at the GC region, and there are not enough lower energy signal events for an analysis with a soft energy spectrum. Hence, we want to start from the same point as previous event selections and develop a better event selection for the GC region analysis that improve upon the issues mentioned above. This selection includes events in the GC declination ($-28.936^\circ$) $\pm 10^\circ$ and the whole right ascension range. Because the dataset focuses on a smaller region of the sky, a higher event rate can be tolerated. Comparing to other IceCube datasets, the Gamma-Ray Follow-Up (GFU) dataset [6] has an event rate of 0.530 mHz per steradian and the Point Source Tracks (PS Tracks) dataset [7] has an event rate of 0.313 mHz per steradian. This selection can lead to a final event rate two orders of magnitude higher. The high event rate of the GC dataset offers an opportunity to search for flares with shorter time windows. The selections are made to optimize the time-independent point source sensitivity at the GC with an energy spectrum of $E^{-2.7}$. 
3. Event Selection

The principal challenge of any neutrino telescope is the large amount of atmospheric muons and neutrinos produced in cosmic ray interactions in the atmosphere. The amount of these background events detected by IceCube is about ten orders of magnitude higher than the astrophysical neutrinos. Therefore, the goal of event selection is to reduce the background in the dataset to increase its sensitivity to astrophysical sources. For the event selections applied in this dataset, the actual collected data are used as background in the first two steps because at this stage, data are largely dominated by background events. In the last step, we replace the background by the simulated Monte Carlo background events. Simulated Monte Carlo signals are used throughout the event selection. In each step, the same cuts are applied to both backgrounds and signals to test the dataset sensitivity.

3.1 Pre-cuts

The pre-cuts are applied to reduce the event rate and select high quality events. The cuts applied are: 1. remove events that fail direction reconstructions, 2. remove events that have angular resolutions higher than 4.5 degrees, 3. remove events that hit less than 18 DOMs in the detector, 4. remove events with a reconstructed length less than 250 meters. These cuts reduced the background by 53.5% and signal by 24%.

3.2 BDT Classifier

The BDT model is an ensemble learning method that combines weak learning tree models into a strong learning tree. The advantage of using a BDT is it trains fast and reduces bias. The model used is XGBoost with 150 estimators and a maximum depth of 6. The BDT classifier here classifies atmospheric muons and atmospheric neutrinos (background) versus astrophysical neutrinos (signal). The top features that contribute to the model are the agreement between two different direction reconstructions, the percentage of the event contained in the detector, and the number of hit DOMs 15 ns before the strong pulse is detected (background muon bundles are likely to have more early hit DOMs comparing to single neutrinos). The score ranges from 0 to 1, indicating the degree to which an event exhibits background-like properties (0) or signal-like properties (1). The distribution of BDT scores for signal and background events is shown in Figure 1. As expected, a larger ratio of background events receive low scores. Please note that this plot shows the probability density of the two categories independently. As a result, it does not provide a comparison of the relative ratio of signals and backgrounds under the same score. A cut is made at BDT score = 0.025 where all events scored below that are removed. This cut removes 85.2% of background events and 22.1% of signal events.

3.3 BDT Regressor

The background rate has been largely reduced after the previous BDT classifier cut. The last step using a BDT regressor is to reduce the inclusion of muon bundles. Muon bundles are generated in the first interactions of a primary cosmic ray in the Earth’s atmosphere. Muons lose energy and produce charged particles that generate Cherenkov light in the IceCube detector. Since the muon bundles comprise numerous low-energy muons that simultaneously enter the detector, the
Search for Neutrino Emission at the Galactic Center Region

Figure 1: These plots show the distribution of BDT scores for background and signal events. The right plot is a zoomed-in plot of the left. The score, ranging from 0 to 1, indicates the probability of an event being a background versus signal (0 being more background-like and 1 being more signal-like).

Cherenkov light produced by those muons in the detector is comparable to a single high-energy muon from an astrophysical neutrino interacting with the ice. Hence, muon bundles are often misinterpreted as high-energy events during reconstruction. Due to their higher likelihood of being a signal, high-energy events make a larger contribution to source searches. Thus, it is necessary to identify and minimize the presence of muon bundles within the dataset.

The method for distinguishing between muon bundles and single muons is based on the energy loss pattern observed in the detector. In the case of single muons, their energy loss is primarily characterized by significant stochastic fluctuations. On the other hand, muon bundles exhibit a relatively continuous energy loss pattern. The variable *Stochasticity* is introduced here to quantify the level of stochasticity in the energy loss process. There are several ways to calculate *Stochasticity*. The main idea is to divide the detector into several sections and compare the energy loss in each section to a linear fit of the loss. Figure 2 shows the distribution of *Stochasticity* calculated with the equation

$$\text{Stochasticity} = \frac{1}{\sqrt{2}} \cdot (f_{fit} - \frac{1}{2} \cdot \log_{10}(\text{PeakOverMedian})), \quad (1)$$

where $f_{fit}$ is the linear fit of the energy loss and PeakOverMedian is the maximum divided by median energy loss in each bin. The figure compares the distribution of this variable in different energy ranges. Muon bundles have high energy and lose energy more smoothly, meaning muon bundles are high energy events with a low *Stochasticity*. The shape of the *Stochasticity* distribution for signal events remain similar across different energy ranges, but the background events at high energy tend to have low *Stochasticity*, indicating they are likely to be muon bundles.

Due to the objective of reducing high-energy background events, it is essential to approach the cut in a cautious manner. A straight cut using the *Stochasticity* calculated in Equation 1 deteriorates the sensitivity. Here, we introduce a BDT regressor to optimize the cut. Besides Equation 1, there are six other ways to calculate *Stochasticity*. The total of seven *Stochasticity* calculations are used as training features of the BDT regressor, predicting the *singleness* of each event. A high-energy event with a low *singleness* score is more likely to be a muon bundle. Figure 3 shows the *singleness* predicted by the BDT regressor versus event energy. The peak at around 0.5 in *singleness* is where
Search for Neutrino Emission at the Galactic Center Region

Figure 2: Stochasticity (Equation 1) distribution of background and signal in different energy ranges.

Figure 3: These two plots are 2D histograms of the singleness predicted by the BDT regressor versus energy. The left plot shows background distribution and the right shows signal. Events in the upper left corner ($\log_{10} E > 4.5$ and $\text{singleness} < 0.4$) are removed from the dataset.

The initial guess was seeded at for the BDT model. Certain events do not proceed to subsequent branches of the decision tree and consequently result in a final prediction based on the initial guess. A cut is made to remove events with $\log_{10} E > 4.5$ and $\text{singleness} < 0.4$. This cut removes 34.1% of background and 12.4% of signal.

4. Neutrino Flare Search

The GC’s dynamics are largely determined by the presence of the central SMBH, present at the position of Sgr A* [3]. Centers of Galaxies have early-on been identified as possible accelerators for cosmic rays. The detection of diffuse gamma-ray emission in the Galaxy up to 100 TeV photon energy indicates the presence of a PeVatron in the GC region [2]. SMBHs are in general predicted to be a source of short-term flares of high-energy emission, for instance through a tidal disruption event. We therefore propose a flare search for the GC region. For IceCube, a 1 degree region (in diameter) covers around 140 parsec diameter region near the GC. Due to an angular resolution limit of around 1.7 degrees in this dataset, it is hard to study the details and specify candidates in the GC. Hence, this analysis focuses on a generalized untriggered flare search at one point (GC), searching for the most significant flare with the dataset created above. Untriggered searches look for all potential flares and are not associated with any specific source lightcurves from multi-wavelength observations.
4.1 Analysis Method

The analysis uses an unbinned maximum likelihood method. This method includes a signal probability distribution function (PDF) to describe the signal distribution and a background PDF to describe the background distribution.

Similar to a time-integrated analysis, the hypothesis tested here is that there is a source located at $x_s$ with a spectral index $\gamma$ and signal strength (number of neutrino events expected in the final data sample) $n_s$. A weighted sum of the signal and background PDFs can be obtained to describe the probability of the $i$th event coming from the GC:

$$P_i = \frac{n_s}{N} S_i + \frac{1 - n_s}{N} B_i, \quad \text{(2)}$$

where $S_i$ and $B_i$ are the signal and background PDFs respectively. The full signal PDF is

$$S_i = P_{i}^{\text{space}} (|\vec{x}_i - \vec{x}_s|, \sigma_i) \cdot P_{i}^{\text{energy}} (E_i, \theta_i, \gamma) \cdot P_{i}^{\text{time}}. \quad \text{(3)}$$

The spacial PDF $P_{i}^{\text{space}}$ is a Gaussian function of $|\vec{x}_i - \vec{x}_s|$ (the space angular difference between each event $i$ and the source location) and the angular error estimation $\sigma_i$. The energy PDF $P_{i}^{\text{energy}}$ is a function of the event energy $E_i$, declination $\theta_i$, and the assumed energy spectral index of the source $\gamma$. The Gaussian time PDF is given by

$$p_{i}^{\text{time}} = \frac{1}{\sqrt{2\pi} \sigma_t} \cdot \exp\left(-\frac{(t_i - t_0)^2}{2\sigma_t^2}\right), \quad \text{(4)}$$

where $t_0$ and $\sigma_t$ are the mean and width of the burst and $t_i$ is the event time of the $i$th particle.

Similarly, the background PDF is given by

$$B_i = P_{i}^{\text{space}} (\theta_i, \alpha_i) \cdot P_{i}^{\text{energy}} (E_i, \theta_i) \cdot P_{i}^{\text{time}}. \quad \text{(5)}$$

The space PDF $P_{i}^{\text{space}}$ describes the background event distribution. For a time-independent analysis, it can be simply described using declination $\theta_i$ because the exposure for all right ascension directions is averaged out from the rotation of the Earth. However, an azimuth correction needs to be applied to the space PDF for time scales shorter than one day, which is why $P_{i}^{\text{space}}$ also includes the local azimuth $\alpha_i$. The energy PDF $P_{i}^{\text{energy}}$ is determined from the energy estimator distribution, which is dependent on the declination $\theta$. A uniform efficiency over time is assumed for background, so the time PDF here $p_{i}^{\text{time}} = 1$. Since this analysis is time-dependent, in addition to the space and energy PDFs, a Gaussian time PDF is included to account for the time dependence in $S_i$ and $B_i$ [8]. The time PDF is used to identify clusters of signal like events that stand out from background fluctuations in the time regime. The likelihood is the product of the probabilities of each event $P_i$, which is described as Equation 2. This likelihood favors short flares because there are more independent ways to choose the burst time for short flares comparing to long flares. We introduce a marginalization factor to penalize short flares, which leads to a new likelihood function:

$$L(n_s, \gamma, \sigma_T, T_0) = \prod_i P_i \sim \frac{\sqrt{2\pi}\sigma_T}{T_{\text{max}} - T_{\text{min}}} \cdot L(n_s, \gamma, \sigma_T, T_0), \quad \text{(6)}$$

6
Search for Neutrino Emission at the Galactic Center Region

Figure 4: This is a time series plot near the best fit flare. The events are shown as blue lines. The heights of the lines indicate the ratio between the signal probability (Equation 3) and the background probability (Equation 5) of each event. The x-axis shows MJD in days. The dashed black curve is the best fit Gaussian flare with a mean at $T_0 = 58470.42$ and a width of $\sigma_T = 0.063$. The height of the curve is arbitrary.

with $T_{\text{max}}$ and $T_{\text{min}}$ being the maximum and minimum livetime of the dataset. Finally, a test statistic (TS) can be obtained from the likelihood ratio test [9],

$$TS = -2 \log \left( \frac{T_{\text{max}} - T_{\text{min}}}{\sigma_T} \cdot \frac{L(n_s = 0)}{L(n_{\text{ns}}, \gamma, \sigma_T, T_0)} \right),$$

(7)

where $L(n_s = 0)$ is the null hypothesis indicating there is no significant time-dependent clustering of events within the search window. The signal hypothesis is a time-dependent single-flare Gaussian time profile at all times within the livetime with a flare width up to half of the livetime. The best-fit parameters were found by searching around the parameter spaces to maximize the TS. A maximum TS is searched for during the 12-year time window.

4.2 Results

No significant clustering of events was observed in the time regime for this analysis. The final best-fit values are: 1. $\gamma = 2.75$ 2. $n_s = 2.44$ 3. $T_0 = 58470.42$ MJD 4. $\sigma_T = 0.063$ Days. The time series of events and the best fit Gaussian flare is shown in Figure 4. A Gaussian fit flare is shown at mean $= T_0$ with a width $= \sigma_T$. This best fit flare has a $TS = 0$. This analysis sets a threshold of $S/B > 1000$ to only include events that contribute more to the point source search at the GC.

Since no significant flare has been found here, we calculate the upper limit with a 90% confidence level with the following procedure: First, obtain a background TS distribution by randomizing the data in right ascension (simulate background fluctuations) 100,000 times and calculate the TS of the best-fit flare of each randomization. Then, inject a Gaussian flare by injecting signal events that are generated from a Monte Carlo simulator to a $T_0$ with a width of $\sigma_T$. We then perform 500 trials of randomization and obtain a TS distribution of the injected signal. Finally, we keep injecting signal events until 90% of the TS is higher than the median of the background TS. Figure 5 shows the upper limit for different time windows assuming spectral indices of $\gamma = 2.0, 2.5, 2.7$. 

7
Figure 5: Upper limits for single flare at the GC with different flare widths assuming energy spectral indices $\gamma = 2.0, 2.5, 2.7$.

5. Conclusion

This analysis uses a newly created GC dataset to search for a single flare at the GC. The final data rate of this event selection is 50.47 mHz, which is approximately two orders of magnitude higher compared to other similar datasets in IceCube. This provides an opportunity to search for short flares since short time windows can be considered as background-free regions. This single flare search is focused at the GC as a point-like source. No significant time-dependent clustering of events was observed in this analysis. We obtain the upper limits of various time windows with the assumption of energy spectrum $\gamma = 2.0, 2.5$, and 2.7. Further analyses like a point source scan near the GC and extended source scan at different extensions can be done using this dataset.

References

Acknowledgements

The authors gratefully acknowledge the support from the following agencies and institutions: USA – U.S. National Science Foundation-Office of Polar Programs, U.S. National Science Foundation-Physics Division, U.S. National Science Foundation-EPSCoR, Wisconsin Alumni Research Foundation, Center for High Throughput Computing (CHTC) at the University of Wisconsin–Madison, Open Science...
Search for Neutrino Emission at the Galactic Center Region

Grid (OSG), Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support (ACCESS), Frontera computing project at the Texas Advanced Computing Center, U.S. Department of Energy-National Energy Research Scientific Computing Center, Particle astrophysics research computing center at the University of Maryland, Institute for Cyber-Enabled Research at Michigan State University, and Astroparticle physics computational facility at Marquette University; Belgium – Funds for Scientific Research (FRS-FNRS and FWO), FWO Odysseus and Big Science programmes, and Belgian Federal Science Policy Office (BELSPO); Germany – Bundesministerium für Bildung und Forschung (BMBF), Deutsche Forschungsgemeinschaft (DFG), Helmholtz Alliance for Astroparticle Physics (HAP), Initiative and Networking Fund of the Helmholtz Association, Deutsches Elektronen Synchrotron (DESY), and High Performance Computing cluster of the RWTH Aachen; Sweden – Swedish Research Council, Swedish Polar Research Secretariat, Swedish National Infrastructure for Computing (SNIC), and Knut and Alice Wallenberg Foundation; European Union – EGI Advanced Computing for research; Australia – Australian Research Council; Canada – Natural Sciences and Engineering Research Council of Canada, Calcul Québec, Compute Ontario, Canada Foundation for Innovation, WestGrid, and Compute Canada; Denmark – Villum Fonden, Carlsberg Foundation, and European Commission; New Zealand – Marsden Fund; Japan – Japan Society for Promotion of Science (JSPS) and Institute for Global Prominent Research (IGPR) of Chiba University; Korea – National Research Foundation of Korea (NRF); Switzerland – Swiss National Science Foundation (SNSF); United Kingdom – Department of Physics, University of Oxford.