<u>Search for Astrophysical Neutrino Point Sources with Data from the IceCube Neutrino Observatory</u>

Word Count: 3655

Introduction

Neutrinos are nearly massless subatomic particles that are produced by the decay and interaction of other particles such as pions, muons, and tau leptons. There are three distinct flavors of neutrinos, known as the electron neutrino, muon neutrino, and tau neutrino, each of which are created through these slightly different particle interactions. Neutrinos are thought to be produced by the same high energy phenomena that produce cosmic rays. This means that much of the interesting neutrino flux that is naturally detectable on Earth is thought to be extraterrestrially in stars and other cosmic events like active galactic nuclei. These neutrino events, that come from astrophysical sources like quasars many light years away from the Earth, often have energy values much higher than those that come from sources much closer to the Earth, such as the Sun. These cosmic neutrinos then come to Earth, and can be detected by various different methods, which have been the subject of study for many years in the field of astroparticle physics. This is because neutrinos provide key information about processes inside astronomical objects many light years away from the Earth, and more knowledge about the neutrinos produced by these objects would allow physicists to study the makeup and processes within different astronomical objects.

However, neutrinos can also be produced in the Earth's atmosphere. This occurs through interactions in the atmosphere itself,which involve the same particles that produce astrophysical neutrinos, and often occur when cosmic rays impact with particles in the Earth's atmosphere. This creates a sort of background flux of only low energy neutrinos that are detected by any apparatus that also senses astrophysical neutrinos produced in high energy interactions in active galactic nuclei such as stars, quasars, or blazars[1]. These atmospheric neutrinos therefore influence sample purity in any large data sample produced by a neutrino detector, and can necessitate a background flux simulation in order to remove them from an analysis that focuses on neutrinos of only astrophysical origin.

This paper will focus on only one type of neutrino detector, known as a Cherenkov-radiation based detector. Cherenkov radiation is a physical phenomenon that occurs when a charged particle passes through a medium (such as water), at a speed faster than the speed of light in the same medium[2]. This causes a wavefront of visible electromagnetic radiation in the form of a flash of blue light, due to a charged particle's electromagnetic interaction within the medium. In practice, these charged particles are often muons or electrons created by the reaction of neutrinos with matter within the detector volume. When a high energy neutrino collides with a substance like water or ice, it interacts with the water or ice molecules creating charged particles, which in turn speed through the medium and create Cherenkov radiation.

Because particles from neutrino interactions often produce Cherenkov radiation, this phenomenon is extremely useful to those attempting to detect neutrinos, and is used in neutrino detectors around the Earth. These detectors consist of modules with photomultiplier tubes, a device for amplifying and detecting light waves (coupled with other electronics) that are spaced out over a large area of water or ice. These small modules are designed to accurately observe the Cherenkov light created by the neutrinos. Water or ice are chosen for their clarity, so that they do not heavily affect the direction or visibility of the light. The large arrays of detection modules that make up the detectors produce data on the time and direction of recorded events as they pass through the detector: such data includes values of declination, right ascension, time, and energy. These different

[1] Stettner, J. (2019)
[2] Richmond, M. (n.d.).

data can then be used to create a detailed picture of the amount of neutrino events that pass into the Earth from extraterrestrial sources at any given time.

The data used in this paper comes from the IceCube Neutrino Observatory, a Cherenkov radiation-based detector similar to those described previously. This detector uses a cubic kilometer of instrumented ice at the Earth's south pole. It is located at the Amundsen-Scott station at the South Pole due to the clarity of the deep ice present in that region, and additionally due to its position at a pole of the Earth, which allows for less change in detector response as a function of Right Ascension. It uses "strings" of Digital Optical Modules (DOMs), similar to the previously described detector modules of photomultiplier tubes, placed into boreholes 2450 meters deep, with spacing of 125 meters between each string. These modules are vertically spaced out along the strings at a distance of 17 meters, with 60 DOMs per string. A central part of the detector made up of 8 DOM strings, known as the DeepCore subdetector, uses the same setup with slightly different specifications: strings are spaced only 70 meters apart, and DOMs are separated by 7 meters along the strings [3] [4]. All DOMs within the detector are used in concert to observe Cherenkov radiation that is caused by potential cosmic neutrino events. Data produced by the detector takes the form of location, which is measured in terms of time and celestial coordinates, and energy. Time is measured in Modified Julian Date (MJD), a unit of time that measures the continuous passage of days since 17 November 1858[5]. Celestial coordinates are expressed in terms of right ascension (RA) and declination (Dec), which measure respectively the angle of a point on the sky at the March Equinox from zero degrees longitude East to West and the angle of a point on the sky from the equator North or South. Energy is measured in terms of electron volts. It is important to note that because of the detector's location at the South Pole, detector response varies by declination, and this phenomenon must be accounted for in any accurate inquiry.

Literature Review

Much scientific work has already been done in this area of inquiry. This existing research, however, focuses on single sources of cosmic neutrinos throughout the universe, and often includes some weighting based on energy values of neutrino events detected. For example, one recent study done by researchers working with the IceCube detector (known as the IceCube Collaboration) observed evidence of neutrino emission from blazar TXS 0506+056 with approximate coordinates of Dec +05.7° and RA 77.6° [6] [7]. Astronomical objects like TXS 0506+056 (known as blazars and quasars) emit massive amounts of matter and electromagnetic radiation, making them great potential candidates for neutrino emission, though none of these objects had been observed emitting neutrinos until the IceCube Collaboration study (IceCube Collaboration). Because of this, the 2018 paper on observed neutrino emission from the TXS 0506+056 blazar was a hallmark study on neutrino astroparticle physics, and added to the body of knowledge on cosmic neutrino production.

In addition to these findings, scientists have looked for correlations between individual cosmic ray sources and potential cosmic neutrino sources throughout the universe. A research

---

[3] IceCube Collaboration. (2017).
[4] *Detector*. (n.d.).
[5] Ray, R. (2000, March 17).
[6] IceCube Collaboration. (2018). Neutrino Emission...
[7] The IceCube, Fermi-LAT, MAGIC, AGILE, ASAS-SN, HAWC, H.E.S.S, INTEGRAL, Kanata, Kiso, Kapteyn, Liverpool telescope, Subaru, Swift/NuSTAR, VERITAS, VLA/17B-403 teams. (2018).

group using data from the IceCube detector coupled with data from the Pierre Auger Observatory and the University of Utah's Telescope Array project conducted a study in 2016 attempting to find such relationships[8]. Both the Pierre Auger and University of Utah projects aim to observe high energy cosmic rays which come from various sources throughout the universe, including neutrino producers like the aforementioned TXS 0506+056. By using these three different detectors in unison, researchers hoped to find similarities between neutrino production and cosmic ray production, leading to greater knowledge of the astronomical processes that produce neutrinos. However, the study concluded that there was very little correlation between the neutrino results from IceCube and the cosmic ray results from the other two detectors, making the research a slight dead end.

The study conducted in this paper attempts to characterize the clustering of the astrophysical neutrino flux, and examines this flux using a unique method which considers a spatial dimension only. The main goal of this research is to examine the entire sky for clusters of high energy neutrino production without considering energy weighting specifically, which will hopefully lead to the identification of new neutrino producing objects.

Methods

Python3 and the Matplotlib module

In this research, the python3 coding language was used because of its ease of implementation. Python3 is a language used primarily for data manipulation and analysis, and all scripts were written using Jupyter Notebook on the IceCube VM. These scripts often used the Matplotlib.pyplot 3.3 library, a module that includes powerful graphical analysis and mathematics functions. In addition to this module, researchers also used the numpy python3 module, a set of defined functions and elements that, similar to matplotlib, allows researchers to perform efficient mathematical operations on large data sets. Numpy also includes a powerful "array" element that can be used to organize data into matrices with rows and columns. These two modules allowed the researchers to create histograms, graphs, and plots of the data which were then analyzed for correlations between the axes of the different graphs. The researchers most frequently used two-dimensional histograms to analyze the data, a graphical representation that shows the frequency at which an intersection of two data points occurs. This tool is often used throughout the astroparticle physics field, and previous analyses of neutrino data from IceCube also used python3 based histograms to visually plot and dissect data[9].

The Data Sample

The data in this study was taken from a search for the diffuse neutrino flux over the whole sky by the IceCube Neutrino Observatory, and was collected from June 2010 to May 2011. This data includes the energy level for logged events, as well as the location of the events in the sky in the form of declination, and was available through public release. This data was downloaded from the IceCube website in the form of a .txt file, containing approximately 93,133 lines of data organized into 7 columns of Modified Julian Date (MJD), a log base 10 energy value in units of

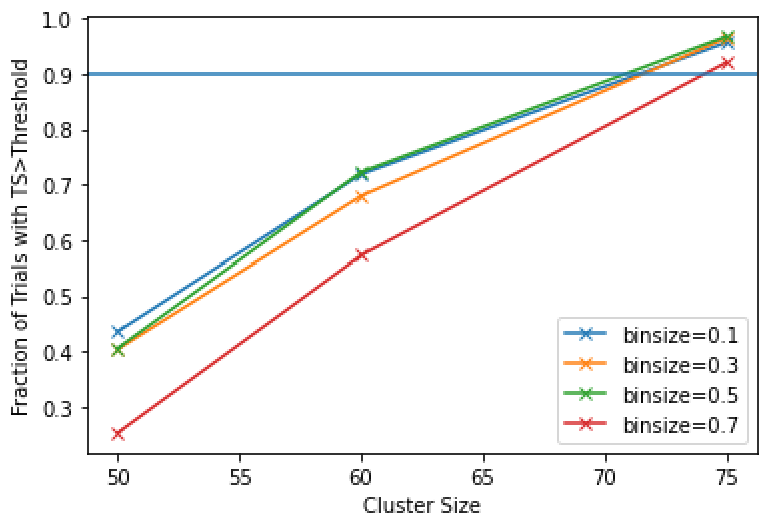
[8] The IceCube, Pierre Auger and Telescope Array Collaborations. (2016).
[9] Aartsen, M.G., et al. (2017).

giga-electron volts (GeV), angular error for each event, right ascension (RA), declination (Dec), azimuth, and zenith, respectively, and which each correspond to a single detected neutrino event. This data sample had a low sample purity level, including many atmospheric neutrino events, but did include many known astrophysical events as well, making it a viable set of data for analysis[10].

Statistical Methods and Hypothesis Testing

This research used a binned chi-squared test statistic in order to analyze the data sample for event clustering with a null hypothesis of little to no event localization, and an ideal alternative hypothesis of a significant number of events that are spatially clustered. Bins were organized to produce the most effective sensitivity rating across the whole sky. Each bin contained 0.25 square degrees in order to uniformly organize each histogram. This means that the RA sample had 720 bins, while the Dec samples had 360. This binning was chosen based on the efficiency curve below, which displays the fraction of trials in a simulated alternative hypothesis that have a test statistic value greater than a set threshold based on a certain single injected cluster size (fig 1). This analysis is done to determine at which binning the threshold includes 90% of a simulated alternative map, and this entire hypothesis testing procedure along with the binning was done in a blind manner, without using or looking at the observed data until the hypothesis simulation was complete.

Fig 1 - Efficiency Curves for Different Bin Size



In order to simulate the ideal null hypothesis and determine background event counts, 10000 randomized lists of RA values were created using the Numpy module containing the same number of values as the observed data array. This was done to remove any clustering present in the data. These lists were then used to create 10000 2D histograms for each energy bin to display the null case, with RA on the x-axis and Dec on the y-axis. A similar process was performed to simulate the alternative hypothesis, although instead of using a pre-defined numpy function to generate a randomized list, this research defined a new function to inject a specified number of values within a given range into the existing RA and Dec arrays. This function created clusters at a randomized RA

[10] Stettner, J. (2019)

and Dec coordinate pair, based on the np.random.random_sample function in the numpy library. These arrays - with the extra points added - were then used to create 2d histograms for each energy bin simulating the alternative hypothesis, showing definite event localization and using the same axes as the null case histogram. These histograms were used in conjunction with the observed data histograms to test the null and alternative hypotheses.

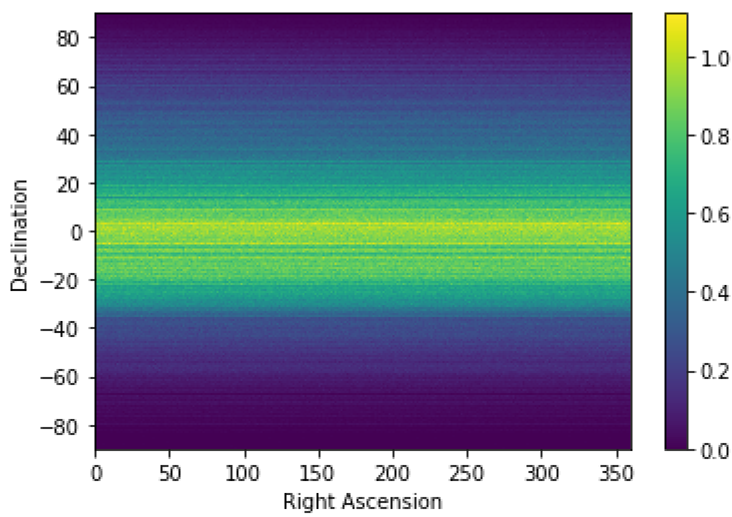
<u>Background and Hypothesis Simulation</u>

In order to utilize the chi-squared (χ^2) test statistic approach, this research created a function to calculate the test statistic given a set of data:

$$\chi^2 = \sum_{i=1}^{\#\ of\ bins} (observed\ value\ -\ average\ background\ value)^2$$

where the average background was a histogram calculated by creating 10000 random RA arrays with the same number of samples as the observed data with seed 0 to 9999, and then creating 2D histograms for each array and the observed declination values for each energy range and taking the mean of the histogram values (fig 2).

Fig 2 - Average Background Histogram



This test statistic was used to numerically summarize the data sample for each energy bin. χ^2 quantities were calculated using the histograms simulating the null hypothesis, and observed data. Many χ^2 values were calculated to simulate the ideal null and alternative hypotheses and to create a distribution of chi squared test statistics, using many different seeds to create the simulations of data with no significant clustering. For the null hypothesis, an array of chi squared values was calculated by repeatedly creating a single random array with seed 0 to 10000 and calculating a chi squared value for each array. These chi-squared values were then used to create a histogram showing the distribution of values. These histograms were then combined to show the difference between chi-squared values in the null and alternative hypothesis cases.

P-Values and Observed Chi Square Values

The p-value significance threshold in this analysis was a 3 sigma, or 0.0013, value due to the high amount of localization necessary for a trial to be statistically significant. The chi-squared function was then used to calculate a single test statistic value using the true RA and Dec values displayed in the data. This chi squared value was then converted into a p-value, with statistically significant results rejecting the null hypothesis having p-value less than 0.0013.

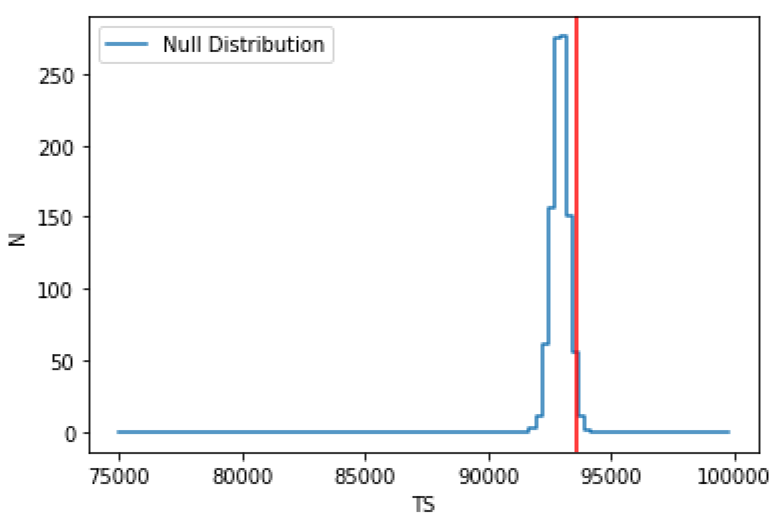
Comparison with Other Sources

Once a p-value had been calculated, the sample was analyzed to see where the greatest amount of signal was coming from, in the form of an equatorial coordinate pair. This location, once located, was compared to the Fermi-LAT 3FGL Catalog. This is a database of many known gamma ray emitters observed by NASA instrumentation, and provides a natural way to examine the multi-messenger behavior of neutrino point source search results, such as this study. The analysis based on this data did not include further hypothesis testing, but only involved an a-posteriori check. This cross-checking was done by scanning over the full data of the Fermi-LAT database, and finding sources, if any, that were within a one degree radius of the high density neutrino event coordinates.

Results

In this inquiry, a p-value of 0.031 was obtained, which is well above the significance threshold of 0.0013. This suggests that results in this analysis were similar to those represented in the average background distribution, which shows little clustering of events in the sky. However, this result does not confirm that no clustering is present, due to the hypothesis testing methodology used in this study. This means that the test statistic observed in the raw data sample was consistent the null hypothesis chi-squared distribution, as shown in the figure below:

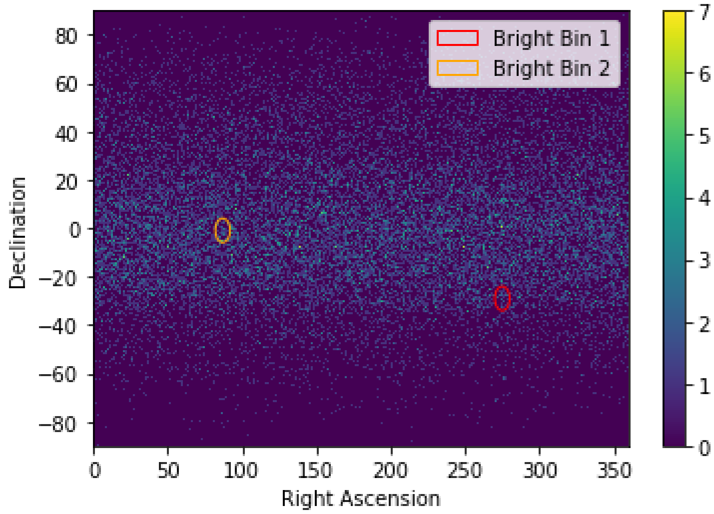
Fig 3 - Test Statistic Distribution with Observed Value



This figure shows the observed test statistic relative to the distribution under the null hypothesis, and displays how close the observed data is to the null distribution, showing how significant clustering was not identified.

The cross checking with the Fermi-LAT catalog yielded interesting results. The coordinates with the largest number of events in the sample were (275, -29) and (87, -1).

Fig 4 - Observed Sky Histogram with Largest Clusters Marked



These bins, however, had no parallel within the Fermi-LAT catalog, suggesting that no energetic astronomical objects have been found at the location that said neutrinos are being produced.

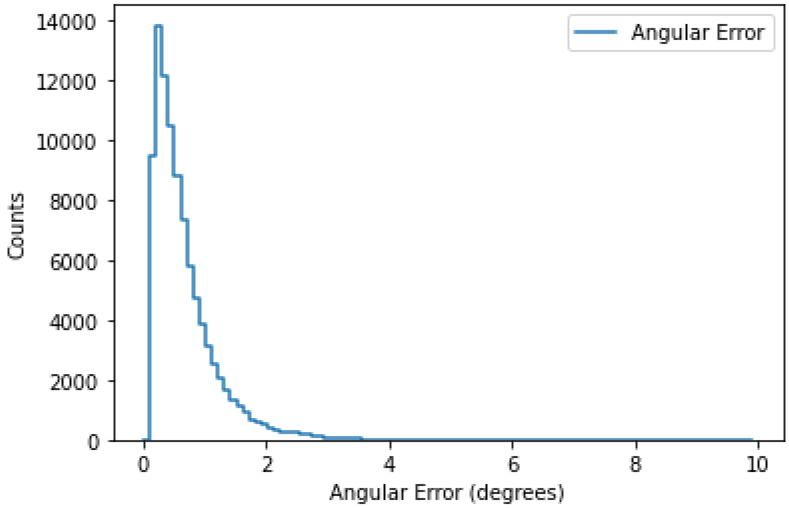
<u>Discussion</u>

No correlation between event clusters and actual sources was found in this analysis. However, it is possible that previously undiscovered astronomical objects exist at (275, -29), and (87, -1). Though the largest cluster in this analysis was not statistically significant, the lack of overlap between the Fermi-LAT database and the findings of this analysis could suggest some need for further understanding, and may additionally show some lack of information in the database. On the other hand, the odd disagreement between the database and detector could be because of a problem with this analysis, which had a few scientific limitations. This means that a more refined method of pinpointing astrophysical neutrino sources may be able to identify some Fermi-LAT sources as neutrino emitters, and could be able to pinpoint point-like sources of astrophysical neutrinos with greater accuracy.

<u>Limitations</u>

This analysis had a number of scientific limitations. Each event logged by the IceCube Neutrino Observatory detector had a certain angular error associated with it, meaning that the Right Ascension and Declination coordinate pairs associated with each event could have been slightly inaccurate.

Fig 5 - Histogram of Angular Error Values in Data



These angular error values were often rather significant, as displayed in the histogram above (fig 5). This analysis did not take into account this potential inaccuracy due to time constraints, and would have done so given more time to complete a full point source analysis, which often may take a year in the particle astrophysics field. This could be a possible area of future improvement and inquiry when doing a point source analysis with a similar data sample.

This analysis also did not consider the energy values of the given data set. Neutrino events are often more likely to be astrophysical if associated with a higher energy value, and therefore adding more weight to higher energy events could be an effective tactic to create a more accurate point source analysis. This consideration was not included in the spatial only analysis done in this paper, and has the potential to be a fruitful future area of inquiry.

In addition to the exclusion of energy considerations, this analysis was sensitive to a minimum cluster size of approximately 70 events (see fig 1). This means that this method would be unable to identify a cluster of less than 70 events as statistically significant, and would therefore not reject the null hypothesis of little clustering even if a data sample included a cluster of less than 70 neutrino events at a similar RA and Dec coordinate pair.

This analysis also used only 1 year of data from the IceCube detector. This means that any fluctuation due to time was unaccounted for, and therefore if a newly-formed point source began emitting neutrinos at a time later than May 2011, this analysis would not be sensitive to it. For this reason, further analysis could be done at a later time, and may potentially yield different results if there is a significant change in neutrino flux with respect to time, or if a newly born neutrino emitter is present.

The majority of the data used in this analysis is also largely known to be atmospheric in origin[11]. This means that though the study done in this paper attempted to characterize the clustering or lack thereof of the astrophysical neutrino flux, it included a large amount of atmospheric data within its bounds. This would result in some inaccuracies in the conclusions drawn in this study, and the inclusion of much atmospheric data, produced near the Earth as opposed to far away from it, could significantly skew the results produced by this analysis.

[11] Stettner, J. (2019)

Conclusion

This study attempted to characterize the clustering of the astrophysical neutrino flux observed by the IceCube Neutrino Observatory, and additionally searched for point sources of astrophysical neutrino production. This was done through a binned, chi-squared test statistic analysis using approximately one year of data from the IceCube detector that encompassed the entire sky. This analysis was done using the Python coding language and the matplotlib and numpy modules on a virtualbox virtual machine affiliated with the IceCube collaboration. The study was not able to identify any clustering in the data sample analyzed, failing to accept the alternative hypothesis of significant event clustering in the sky. Though no significant clustering was identified, the most significant locations in the data sample were then compared to the Fermi-LAT 3FGL catalog, and yielded no matches between any objects in the catalog and the locations in the IceCube data sample. This may have been because of the limitations that were associated with this analysis, which included a significant angular error on many values, lack of energy considerations to further remove atmospheric neutrinos from the analysis, a high threshold for significant cluster sensitivity, and a small time frame of data collection. These factors combined to fail to identify significant clustering in the data sample within this study.

Acknowledgements

This study benefited greatly from the assistance and support of many individuals other than solely the author. Of special help was William Luszczak, who assisted with the management of the entire analysis, from the outset of calculations to the completion of this paper. Additionally, Ms. Rachel Rauch assisted with the completion of this paper, and Dr. Michael Turner assisted greatly with the outset of this paper. This research was also greatly aided by the support of many friends, family, and colleagues, and the author would also like to thank them profusely for their help and support.

References

Aartsen, M.G., et al. (2017). All-sky Search for Time-integrated Neutrino Emission from Astrophysical Sources with 7 yr of IceCube Data. *The Astrophysical Journal*, *835*(151). https://doi.org/10.3847/1538-4357/835/2/151

Detector. (n.d.). IceCube Neutrino Observatory. Retrieved October 17, 2020, from https://icecube.wisc.edu/science/icecube/detector

Fermi-LAT 3FGL Catalog - Interactive Table. (2019). Fermi-LAT 3FGL Catalog. https://fermi.gsfc.nasa.gov/ssc/data/access/lat/4yr_catalog/3FGL-table/#ExportTableData

IceCube Collaboration. (2017). The IceCube Neutrino Observatory: Instrumentation and Online Systems. *Journal of Instrumentation*, *12*. https://doi.org/10.1088/1748-0221/12/03/P03012

IceCube Collaboration. (2018). Neutrino emission from the direction of the blazar TXS 0506+056 prior to the IceCube-170922A alert. *Science*, *361*(6398). https://doi.org/10.1126/science.aat2890

IceCube Collaboration (2018): All-sky point-source IceCube data: years 2010-2012. Dataset. DOI:10.21234/B4F04V

The IceCube, Fermi-LAT, MAGIC, AGILE, ASAS-SN, HAWC, H.E.S.S, INTEGRAL, Kanata, Kiso, Kapteyn, Liverpool telescope, Subaru, Swift/NuSTAR, VERITAS, VLA/17B-403 teams. (2018). Multi-messenger observations of a flaring blazar coincident with high-energy neutrino IceCube-170922A. *Science*, *361*(6398). https://doi.org/10.1126/science.aat1378

The IceCube, Pierre Auger and Telescope Array Collaborations. (2016). Search for correlations between the arrival directions of icecube neutrino events and ultrahigh-energy cosmic rays detected by the pierre auger observatory and the telescope array. *Journal of Cosmology and Astroparticle Physics*. https://doi.org/10.1088/1475-7516/2016/01/037

Ray, R. (2000, March 17). *Modified julian dates*. NASA. Retrieved October 28, 2020, from https://core2.gsfc.nasa.gov/time/

Richmond, M. (n.d.). *Neutrino telescopes based on "water."* Michael Richmond's Classes. Retrieved October 13, 2020, from http://spiff.rit.edu/classes/ast613/lectures/neutrino_water/neutrino_water.html#cerenkov

Stettner, J. (2019). Measurement of the Diffuse Astrophysical Muon-Neutrino Spectrum with Ten Years of IceCube Data. *36th International Cosmic Ray Conference*.