Search for neutrino counterparts of Galactic PeVatrons with IceCube's 12.3-year multi-flavor sample
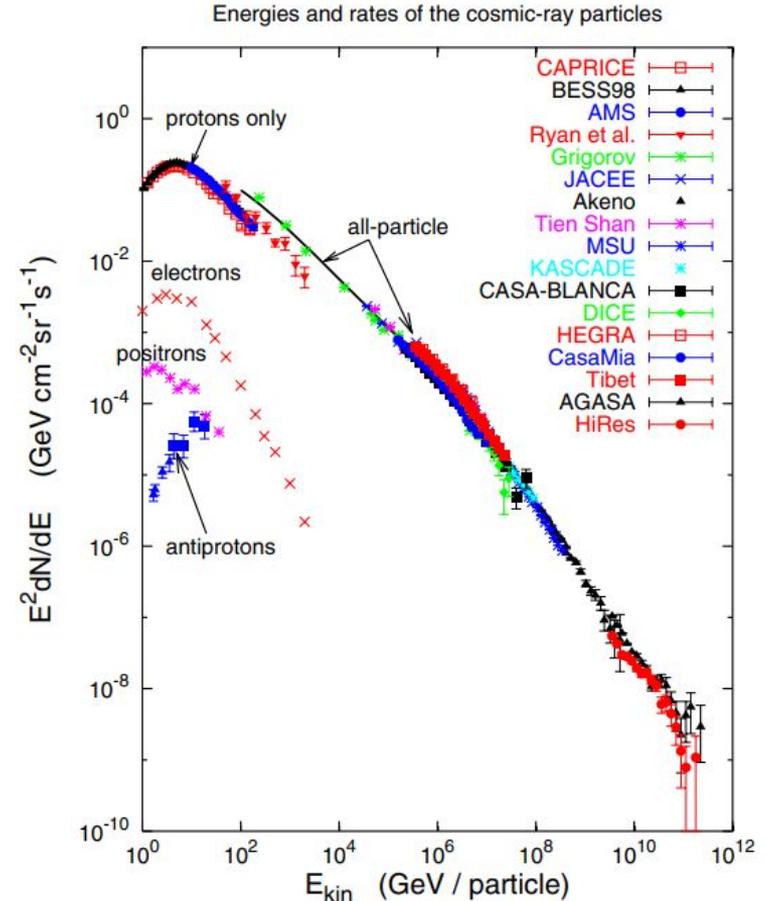
Leo Seen, Albrecht Karle, Lu Lu, Matthias Thiesmeyer, Yuhua Yao, Tianlu Yuan,
for the IceCube Collaboration

Lepton-Photon 2025

1

Galactic cosmic rays

- Comprised of charged particles, mostly protons and various nuclei
- Cosmic rays below the knee (~5 PeV) are thought to originate within the Milky Way
- Bend in magnetic fields which removes information about their source location
- The origin of PeV cosmic rays remains unknown

Energies and rates of the cosmic-ray particles
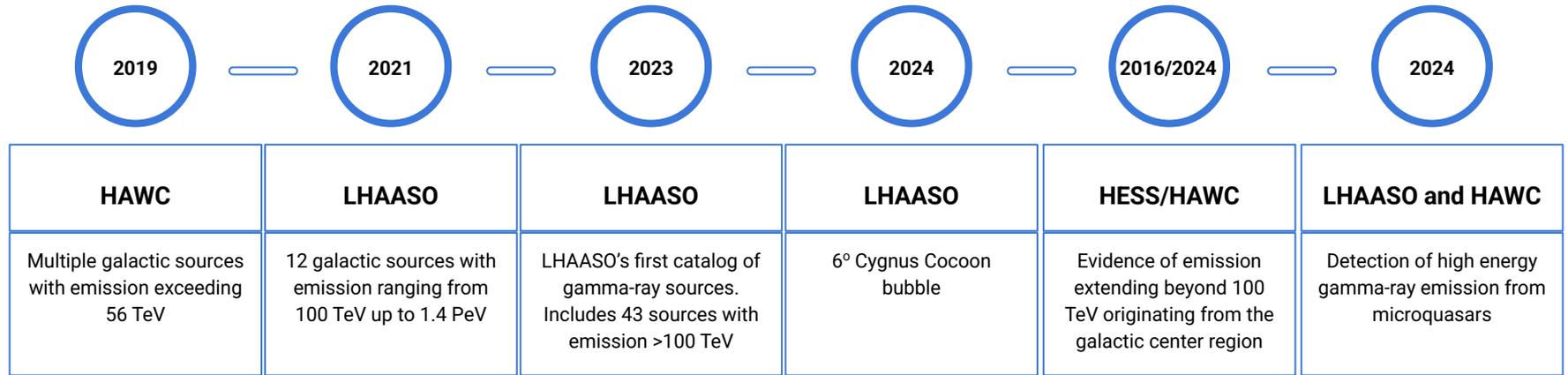
Galactic PeVatrons

- Hypothesized cosmic accelerators within the Milky Way capable of accelerating cosmic rays up to peta-electronvolt (PeV) energies
- Cosmic rays undergoing proton-proton collisions will produce both gamma-rays and neutrinos through charged/neutral pion decay
- Gamma-rays alone can't identify Galactic PeVatrons
  - Can be produced via inverse compton scatter, synchrotron, etc.
- Detection of both neutrinos and gamma-rays originating from the same source signify it as a cosmic ray accelerator

Image credit: Adam Block/Mount Lemmon SkyCenter/University of Arizona
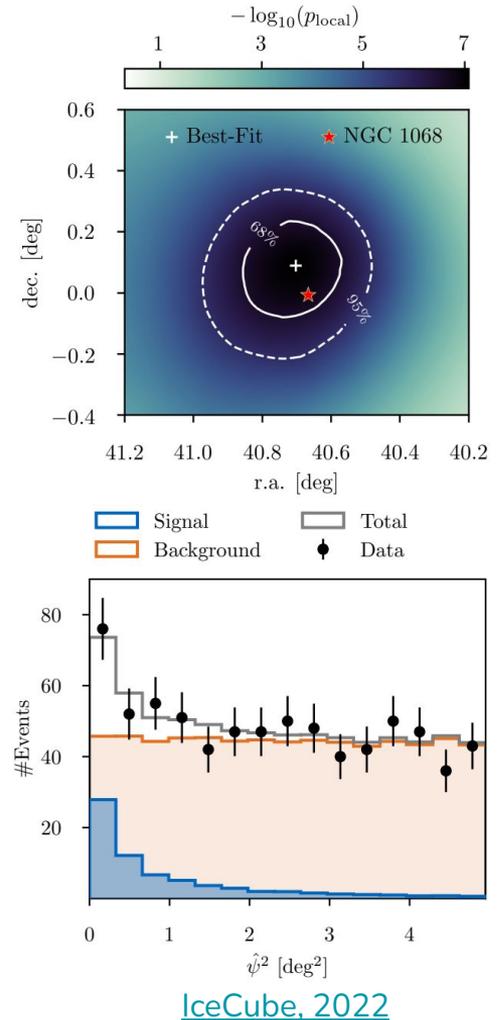

Cygnus Cocoon

Breakthroughs in γ-ray astronomy

2019	2021	2023	2024	2016/2024	2024
HAWC	**LHAASO**	**LHAASO**	**LHAASO**	**HESS/HAWC**	**LHAASO and HAWC**
Multiple galactic sources with emission exceeding 56 TeV	12 galactic sources with emission ranging from 100 TeV up to 1.4 PeV	LHAASO's first catalog of gamma-ray sources. Includes 43 sources with emission >100 TeV	6° Cygnus Cocoon bubble	Evidence of emission extending beyond 100 TeV originating from the galactic center region	Detection of high energy gamma-ray emission from microquasars

- Detection of TeV - PeV gamma-ray sources in the Milky Way suggests the existence of Galactic PeVatrons
- Inverse compton scattering is greatly suppressed at energies >100 TeV due to the Klein-Nishina suppression
  - Photon-electron cross section decreases

4

Breakthroughs in neutrino astronomy

- IceCube has reported evidence for neutrino emission from active galaxy NGC 1068 and blazar TXS 0506+056 at 4.2σ and 3.5σ respectively
- 5.7σ excess originating from the galactic plane with additional years of data, updated reconstruction and ice modeling, and an all-sky all flavor dataset
  - 4.5σ excess previously reported in IceCube Science, 2023

IceCube, 2022
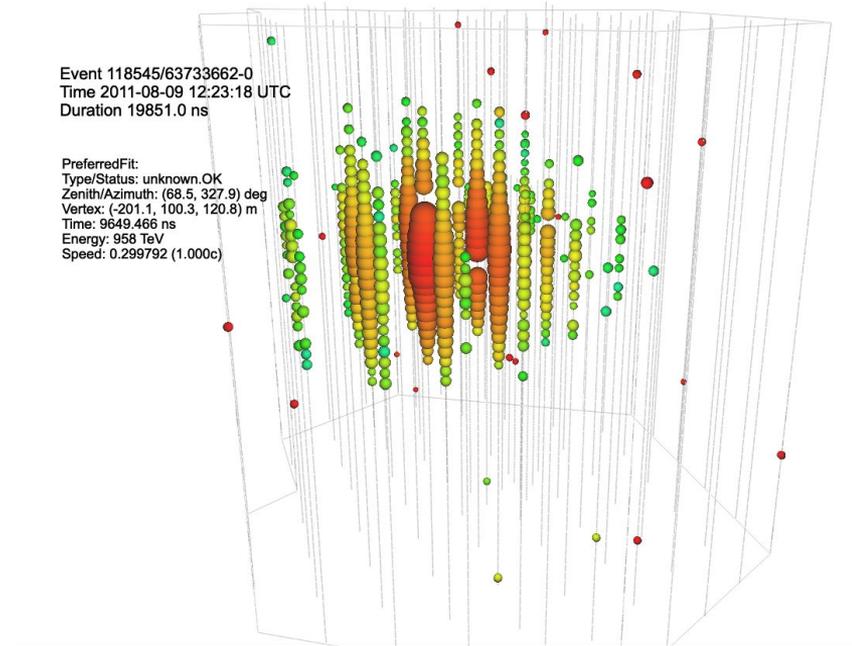
Dataset and Improvements

- **I**ce**C**ub**E** **M**ulti-flavor **A**strophysical **N**eutrino sample
- New method for characterizing in-ice particle shower point spread functions

ICEMAN Dataset

- ICEMAN - **I**ce**C**ub**E** **M**ulti-flavor **A**strophysical **N**eutrino sample
- Event morphologies
  - Through-going muon tracks from the northern sky (Northern Tracks)
    - For point source analysis, we will use Kernel Density Estimators (KDEs) to parameterize per-event point spread functions
  - **All-sky In-ice particle showers (DNN Cascades ICEMAN)**
    - Charged current ν_e, ν_τ and neutral current all flavor interactions
  - All-sky starting muon tracks (Enhanced Starting Track Event Selection [ESTES])
- Over 12 years of livetime
- More information on the ICEMAN dataset and Galactic Plane results was presented by Tianlu Yuan in the previous talk on <u>Thursday afternoon</u>

Cascade point spread function

- Cascades have good energy resolution (~3% at 10 TeV) but poor angular resolution (~10° at 10 TeV)
- In [IceCube Science, 2023](#), the point spread function (PSF) was obtained through a neural network
- Different types of data require different machine learning frameworks
  - Example: Training on analysis level vs. pulse level data
- Boosted decision trees have better interpretability compared to neural networks

Event 118545/63733662-0
Time 2011-08-09 12:23:18 UTC
Duration 19851.0 ns

PreferredFit:
Type/Status: unknown.OK
Zenith/Azimuth: (68.5, 327.9) deg
Vertex: (-201.1, 100.3, 120.8) m
Time: 9649.466 ns
Energy: 958 TeV
Speed: 0.299792 (1.000c)

Predicting per-event point spread functions

- Use LightGBM's gradient boosted decision tree (GBDT)
- Train with 50% of monte-carlo before final level cuts
- Use 14 reconstructed quantities as input features
  - Reconstructed x, y, z, Reconstructed energy, etc.
  - Reconstructed qualities: rlogl, square-residual, chi-squared
- Weigh each event to a $E^{-2.7}$ spectrum
- During the training process, a loss function is minimized
  - The minimum of the loss function is used to evaluate the model performance after each iteration

Loss function parameterization

- Align loss function with spatial component of the IceCube point source likelihood

$$f(\sigma, \psi) = \begin{cases} \frac{1}{2\pi\sigma^2} e^{-\frac{\psi^2}{2\sigma^2}} & \sigma < 7^o \\ \frac{1}{4\pi\sigma^2 \sinh\left(\frac{1}{\sigma^2}\right)} e^{\frac{\cos(\psi)}{\sigma^2}} & \sigma > 7^o \end{cases}$$

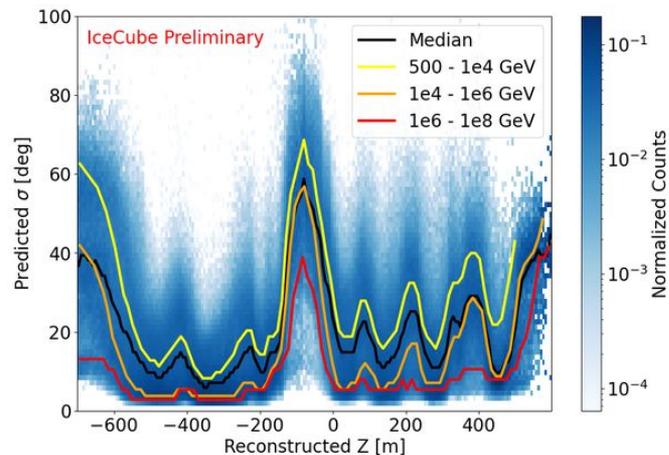- Since the von Mises-Fisher distribution converges to a 2D gaussian at low angular errors, we use the von Mises-Fisher distribution for entire angular error range
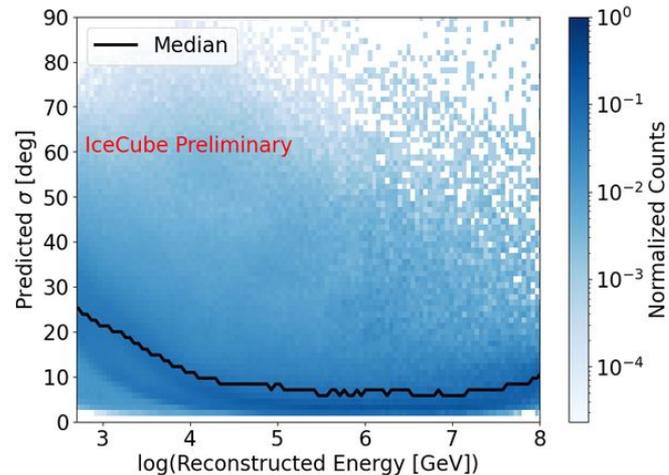
$$L(\sigma, \psi) = -\ln(f(\sigma, \psi)) = \ln(4\pi) + \ln(\sigma^2) + \ln\left[\sinh\left(\frac{1}{\sigma^2}\right)\right] - \frac{\cos(\psi)}{\sigma^2}$$

- LightGBM doesn't naturally allow for bounded parameters. We reparameterize our target variable in terms of $\eta = \ln(\kappa)$ where $\kappa = 1/\sigma^2$ resulting in a numerically stable loss function,

$$L(\eta, \psi) = e^\eta + \ln(1 - e^{-2e^\eta}) - \eta - e^\eta \cos(\psi) \qquad \nabla_\eta(L(\eta, \psi)) = e^\eta \coth(e^\eta) - 1 - e^\eta \cos(\psi)$$

GBDT performance on test set

- 2 of the top 3 most important training variables was energy and reconstructed depth
  - Reconstructed depth is in detector coordinates where z = 0 is near the center of the detector
- Was able to learn both energy and ice layer dependencies
  - Best median PSF is 5.97° between 290 TeV - 13.8 PeV
  - Significant worsening in the PSF due to the dust layer in the region -200m < z < 0m
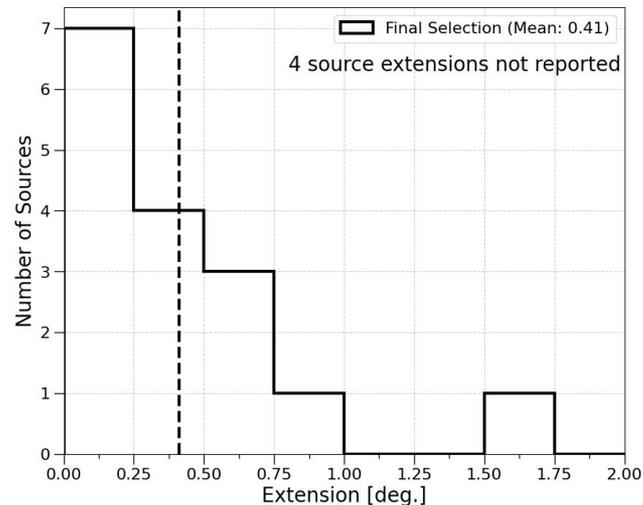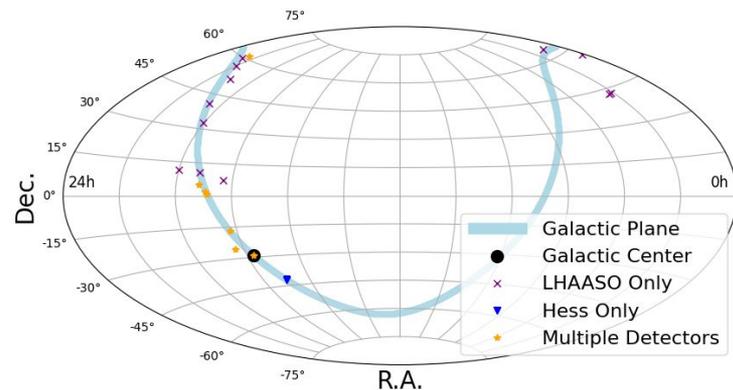
Point source, stacking, and source template analyses

- Individual point source search
  - Source list
  - Sensitivities
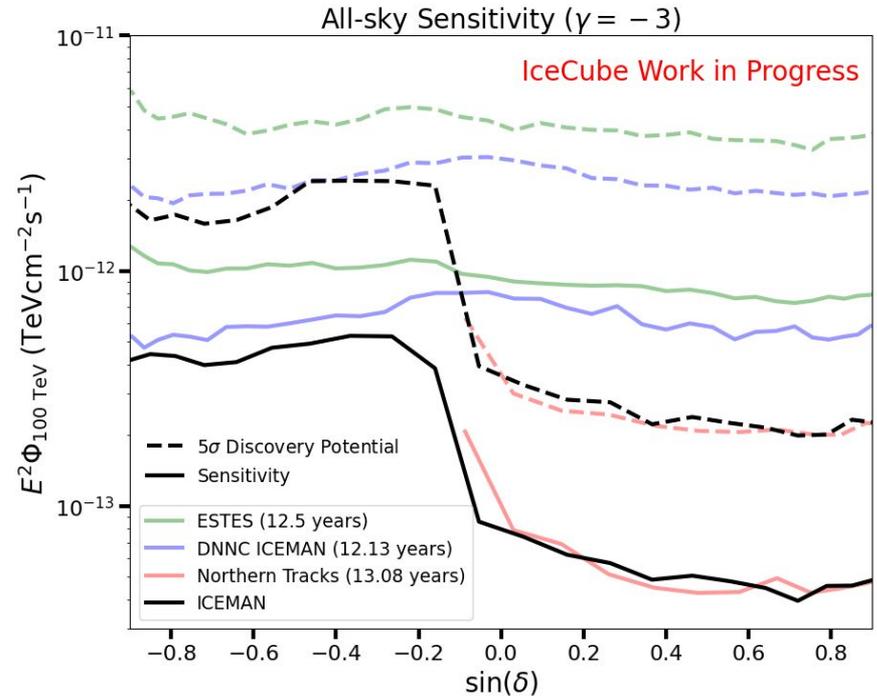- Cygnus Cocoon region
- Stacking catalog search

Source list

- Gathered a list of 49 gamma-ray sources with emission >100 TeV
  - 28 have been detected by LHAASO and HAWC
  - 19 have been detected by LHAASO only
  - 1 has been detected by HESS only, and 1 by HESS and HAWC
- Mostly point sources
- Conduct an individual source and stacking analysis
  - For individual sources, we select 20 sources that are most likely to be hadronic by removing sources near Pulsars/PWN and ones that are believed to be leptonic in literature
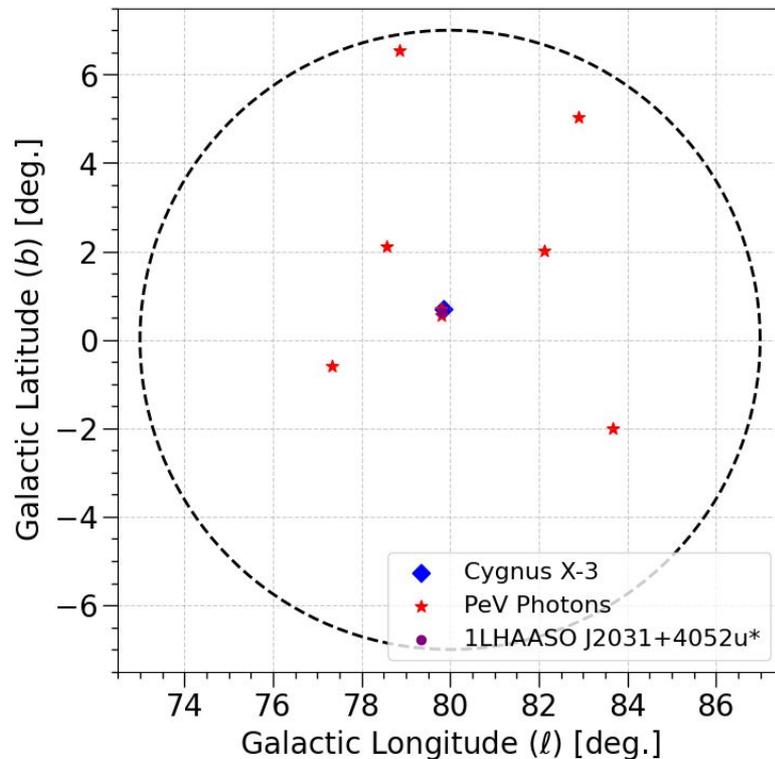
All-sky point source sensitivity

- Northern tracks contribute the
  majority of the sensitivity in the
  northern sky
- Gain in sensitivity from combining
  DNN Cascades ICEMAN and ESTES
  in the southern sky
- Assumes a single power law source
  spectrum model for the entire energy
  range
  - Difficult to compare with individual
    sources due to the sensitive energy ranges
    of gamma-ray detectors
  - Poor extrapolation outside of the sensitive
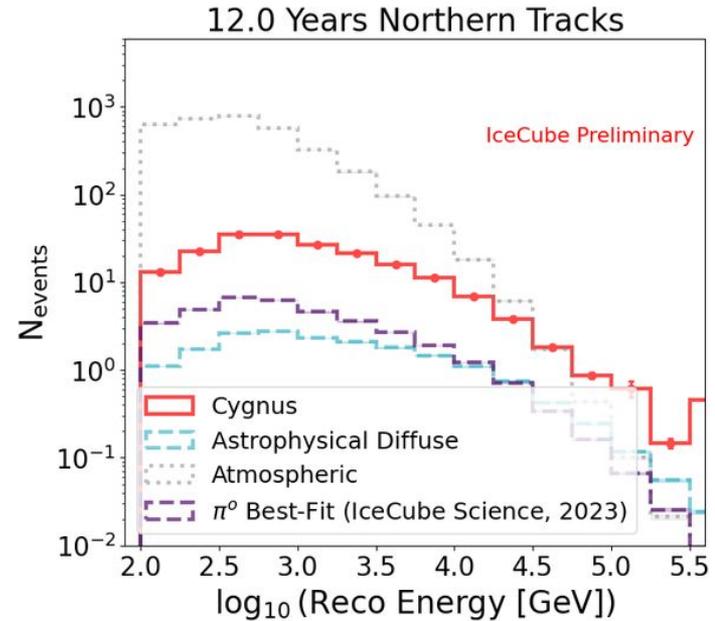   energy range

All-sky Sensitivity ($\gamma = -3$)

IceCube Work in Progress

$E^2 \Phi_{100\ \text{TeV}}$ (TeV cm^{-2} s^{-1})

- - - 5σ Discovery Potential
—— Sensitivity

—— ESTES (12.5 years)
—— DNNC ICEMAN (12.13 years)
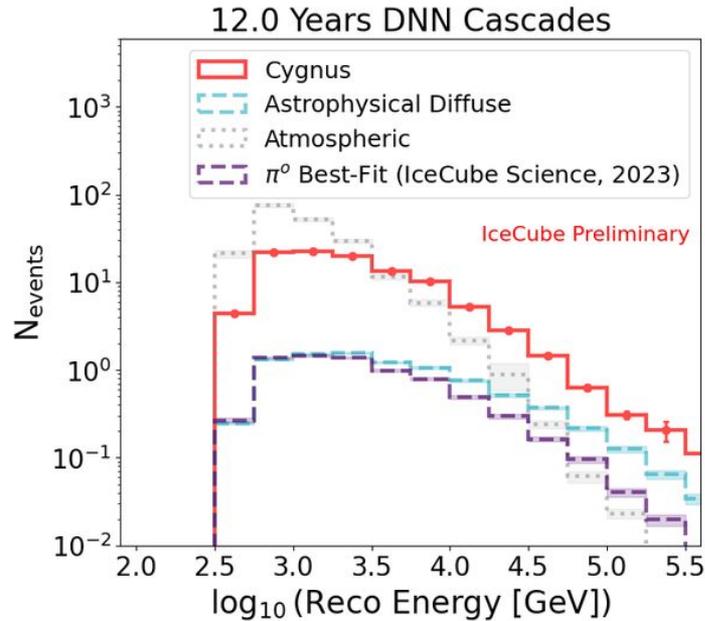—— Northern Tracks (13.08 years)
—— ICEMAN

sin(δ)

*Northern Tracks is sometimes below the combined
sensitivity/discovery potential due to statistics

14

Cygnus Cocoon

- Prominent candidate as a Galactic PeVatron
- LHAASO reported 8 PeV photons from the region in addition to 6° bubble
- HAWC reported a 2° emission region with energies between 1-100 TeV
- Extremely complex star formation region containing microquasars, supernova remnants, massive star clusters, molecular clouds, etc.
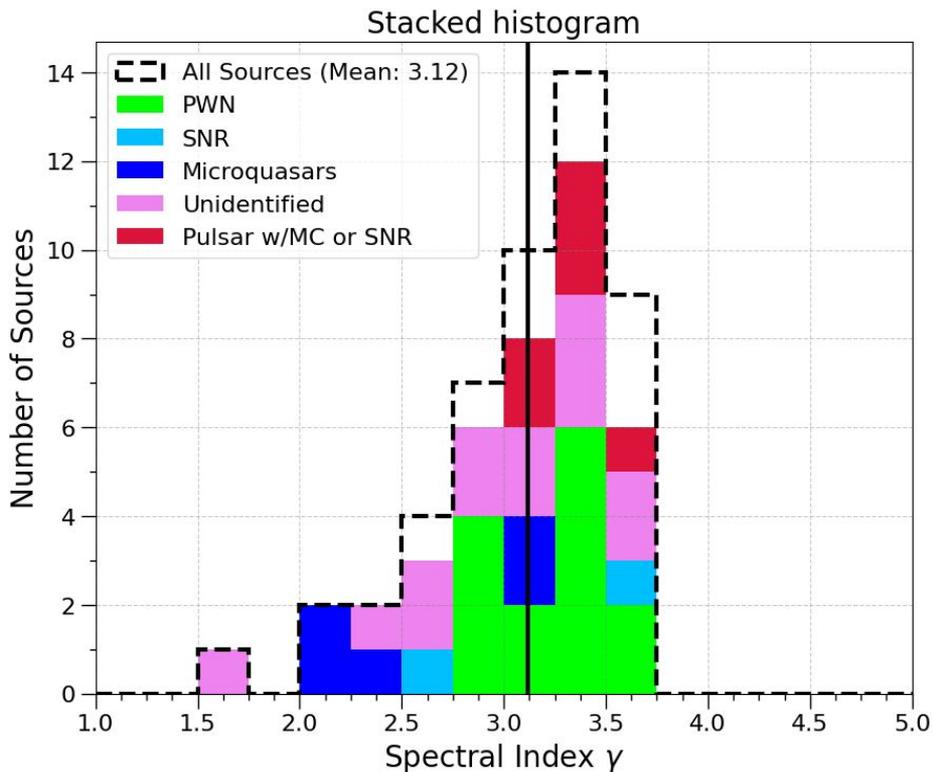
Expected events from the Cygnus Cocoon region

12.0 Years DNN Cascades

12.0 Years Northern Tracks

- Few signal events above background at tens of TeV energies in both DNN Cascades and Northern Tracks

16

Stacking

- Stack 5 different catalogs
- Learn about certain populations of sources
- If one individual source isn't bright enough, we may be able to detect collectively
- PWN and Pulsars have soft spectral shapes at high energies

Stacked histogram

Summary

- ICEMAN all-sky all-flavor dataset
  - New PSF characterization with custom loss function and GBDT
- Individual point source search
  - Selected 20 potentially hadronic sources located along the galactic plane
  - Prospects of neutrino emission originating from the Cygnus Cocoon region
- Stacking analysis
  - Search for emission originating from PWN, SNR, Microquasars, Unidentified sources, and Pulsars coincident with SNR or Molecular clouds
- Ongoing work
  - Road to results: Individual point source and stacking searches
  - Point source skymap with ICEMAN analysis, Cygnus cocoon extended source analysis
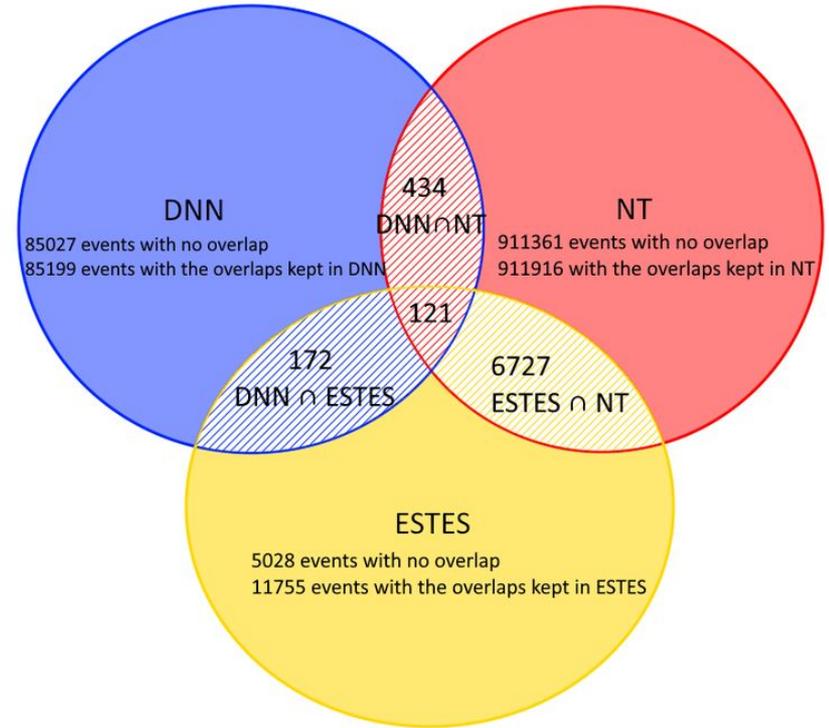
Backup

Motivation

- Galactic cosmic rays
- Galactic PeVatrons
- Breakthroughs in gamma-ray astronomy
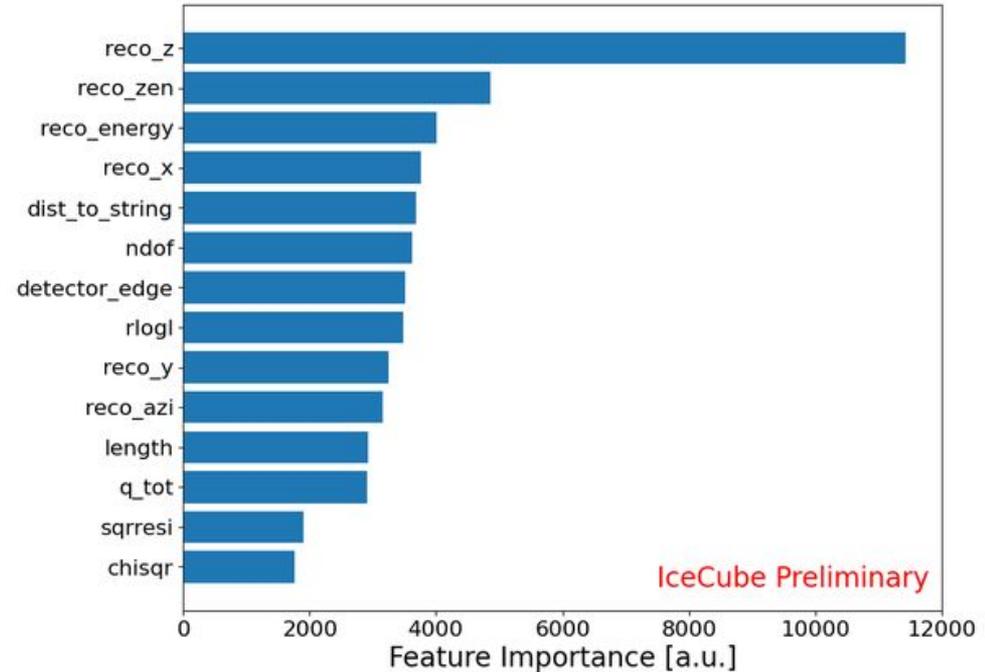- Breakthroughs in neutrino astronomy

ICEMAN dataset specs

- Individual dataset livetimes
    - Northern Tracks (13.08 years), DNN Cascades ICEMAN (12.13 years), ESTES (12.15 years)
- ESTES v DNN -> Remove From ESTES
- DNN v NT -> Remove From DNN
- NT v ESTES -> Remove From NT

DNN
85027 events with no overlap
85199 events with the overlaps kept in DNN

NT
911361 events with no overlap
911916 with the overlaps kept in NT

434
DNN∩NT

121

172
DNN ∩ ESTES

6727
ESTES ∩ NT

ESTES
5028 events with no overlap
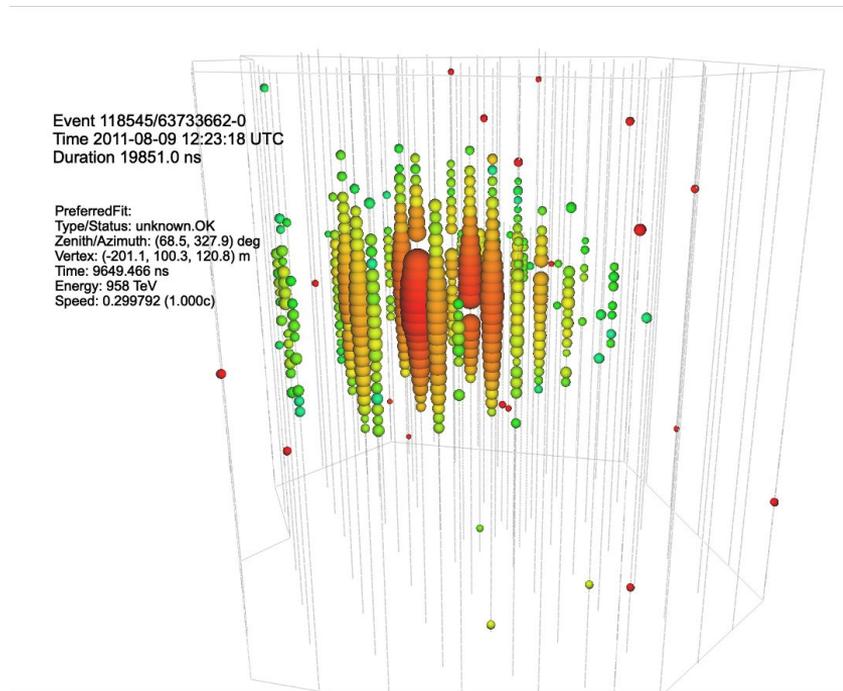11755 events with the overlaps kept in ESTES

GBDT feature importance
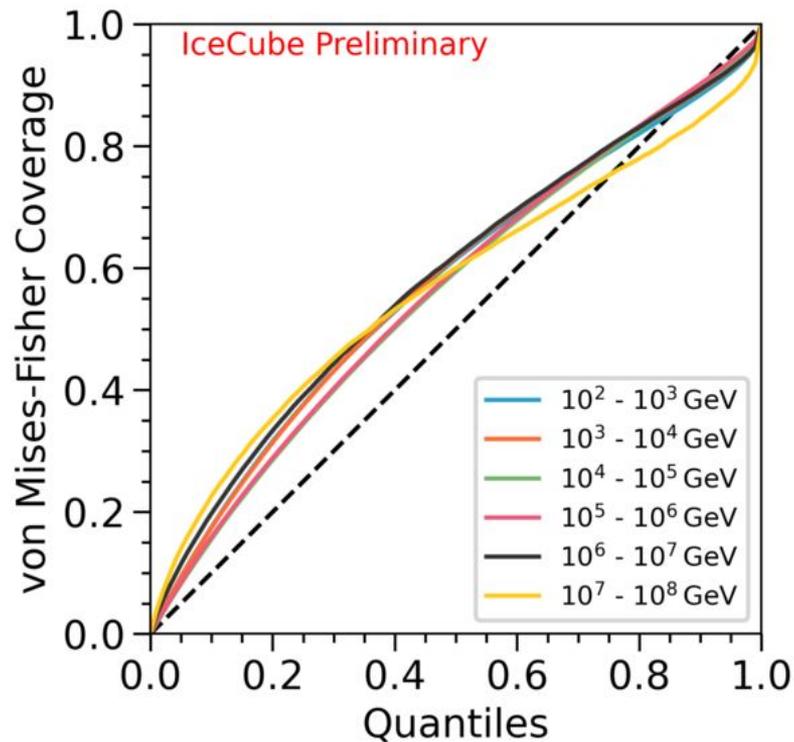
- Feature importance for parameters used during training

Cascade point spread function

- Cascades have good energy resolution (~3% at 10 TeV) but poor angular resolution (~10° at 10 TeV)
- In IceCube Science, 2023, the point spread function (PSF) was obtained through a neural network
- While still debated, studies have shown that tree-based models outperform neural networks when trained on tabular data
  - Tabular data usually has *N* samples with *d* features

Event 118545/63733662-0
Time 2011-08-09 12:23:18 UTC
Duration 19851.0 ns

PreferredFit:
Type/Status: unknown.OK
Zenith/Azimuth: (68.5, 327.9) deg
Vertex: (-201.1, 100.3, 120.8) m
Time: 9649.466 ns
Energy: 958 TeV
Speed: 0.299792 (1.000c)

PSF coverage

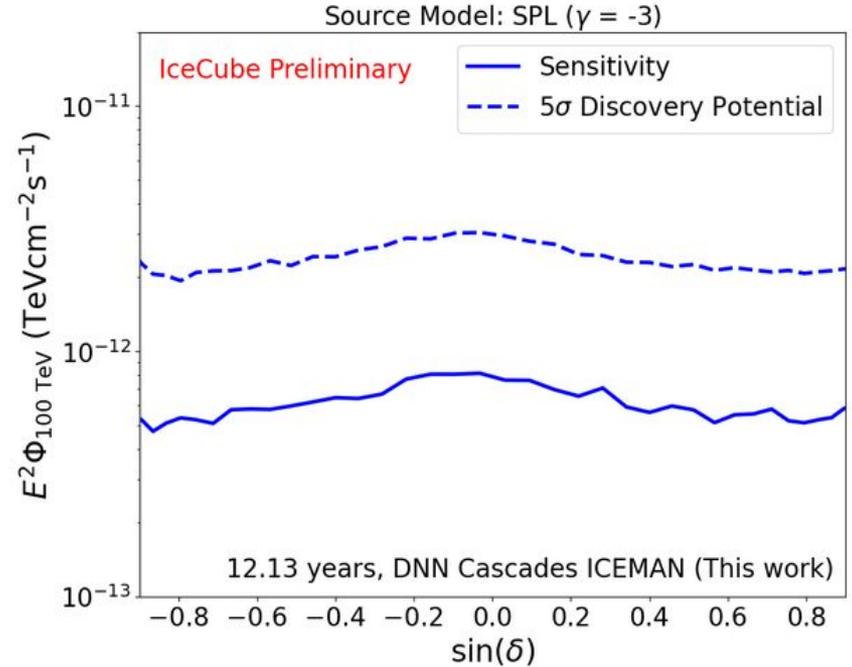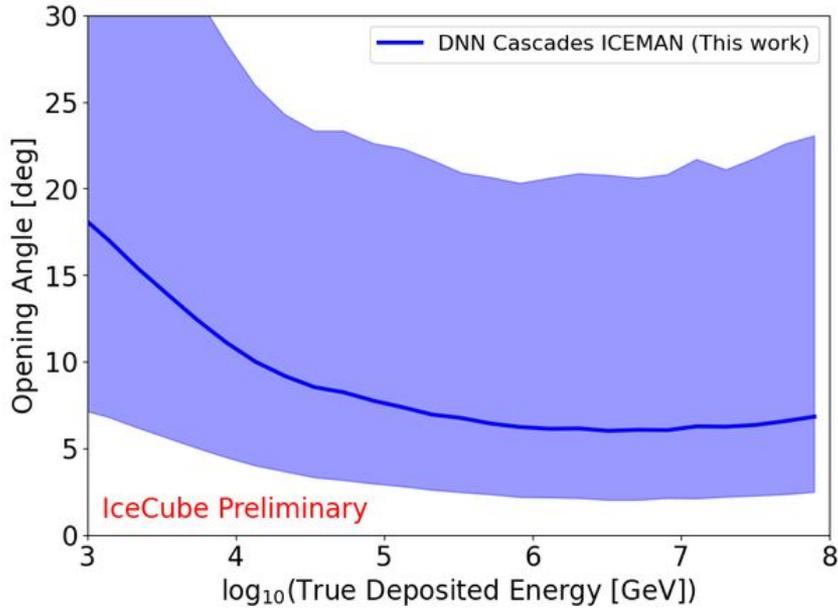- Overcoverage is seen for most of the distribution
- Slight undercoverage at tail of highest energies

GBDT Full list of input features

- Total deposited charge, Reconstructed energy, Reconstructed zenith and azimuth, ndof, Length between two cascades, Distance from reconstructed vertex to closest string, Shortest distance from reconstructed vertex to detector edge

DNN Cascades ICEMAN performance

Improvements in reconstruction

- Choose better reconstruction between Monopod and Taupede to create a PreferredFit reconstruction
  - Monopod is a maximum likelihood fit assuming a single cascade
  - Taupede uses results from monopod as seeds for its fit
- PreferredFit reconstruction is closer to the DirectFit reconstruction (Most accurate reconstruction)

Methods

- Point source likelihood
- Background modeling and PDF construction

IceCube point source likelihood

- Standard point source likelihood

$$L(n_s, \gamma) = \prod_j^M \prod_{i=1}^N \frac{n_s a^j}{N^j} S^j(\delta_i^j, \alpha_i^j, E_i^j, \sigma_i^j, \gamma) + \left(1 - \frac{n_s a^j}{N^j}\right) B_i(\sin(\delta_i^j), E_i^j)$$

- Signal subtracted likelihood

$$\mathcal{L}(n_s, \gamma) = \prod_j^M \prod_{i=1}^N \frac{n_s a^j}{N^j} S^j(\delta_i^j, \alpha_i^j, E_i^j, \sigma_i^j, \gamma) + \tilde{D}_i^j(\sin(\delta_i^j), E_i^j) - \frac{n_s a^j}{N^j} \tilde{S}_i^j(\sin(\delta_i^j), E_i^j, \sigma_i^j, \gamma)$$

- Both signal and background term contain spatial and energy components
  - Background spatial pdf: 1D histogram of data binned in sin(dec)
  - Signal spatial pdf : Gaussian at angular errors <7° and vMF at angular errors >7°, or Kernel-Density-Estimator (KDE)
  - Background energy pdf: 2D histogram of data binned in sin(dec) and $\log_{10}$(reconstructed energy)
  - Signal energy pdf: 2D histogram of MC binned in sin(dec) and $\log_{10}$(reconstructed energy) weighted to a specific spectrum

Background modeling and PDF construction

- Background in this analysis is the atmospheric and diffuse astrophysical neutrino fluxes
  - Due to IceCube's geometry, these only depend on declination
  - First method: simulate by randomizing data events in right ascension
  - Second method: Weight MC to atmospheric and diffuse astrophysical fluxes in pseudo-experiments
- Background spatial pdf can then be constructed in two ways
  - First method: Binning data events in sin(dec)
    - Add signal subtraction term to account for the contamination from signal into the background PDF
      - Marginalize the source pdf over right ascension, currently only possible for analytical functions
  - Second method: Directly constructed by expectations in MC

Selected methods

- For DNN Cascades ICEMAN and ESTES, we use the signal subtracted likelihood and data scrambling as background characterization
  - Both use the gaussian + vMF description of the PSF which allows for an analytical closed form of the signal subtraction term
  - Large background from atmospheric muons in the southern sky
- For the Northern Tracks dataset, we use the non-signal subtracted likelihood and MC background weighting as background characterization
  - A KDE is used for the spatial signal pdf. Therefore, there is no closed analytical form for the signal subtraction term
  - Minimal atmospheric muon background from the northern sky due to earth absorption

SED characterization

- **Why is it hard to compare gamma-ray expected flux to sensitivity?**
  - Single power law from LHAASO KM2A over predicts at lower energies, bad extrapolation
- **Ways to tackle this spectrum mismatch**
  - Float spectral index in the fit
    - Fit an average gamma, loss sensitivity due to mismatch in spectral shape
  - Cut signal pdf to certain energy ranges where gamma-ray measurements are valid and fix spectral index
    - Lose events outside energy range, potentially gain sensitivity due to correctly modeling spectral shape

1LHAASO J2108+5153u SPL+E_{Cut}

Legend:
- Energy>100 GeV: $N_o = 5.74e-10$, $\gamma = 1.95$, $E_c = 18.28$ TeV
- Energy>1 TeV: $N_o = 8.65e-10$, $\gamma = 2.01$, $E_c = 46.47$ TeV
- Energy>10 TeV: $N_o = 1.97e-08$, $\gamma = 2.36$, $E_c = 220.89$ TeV
- Energy>100 TeV: $N_o = 3.97e-07$, $\gamma = 2.63$, $E_c = 243.13$ TeV
- LHAASO Best Fit: $N_o = 1.38e-19$, $\gamma = 2.97$
- Hadronic Model
- LHAASO Pivot Energy (50 TeV)

x-axis: γ-ray Energy [GeV]
y-axis: dN/dE [GeV^{-1}cm^{-2}s^{-1}]

Cygnus technical specs

- Atmospheric diffuse model is using SIBYLL 2.3c hadronic interaction model and Gaisser H3a cosmic ray model
- Source model is derived from LHAASO's 6° log parabola spectrum from Cygnus
- Astrophysical diffuse assumes a single power law of the form $dN/dE = N_o(E/E_0)^{-\gamma}$ with $N_o = 1.44 \times 10^{-8}$ $\text{GeV}^{-1}\text{cm}^{-2}\text{s}^{-1}\text{sr}^{-1}$, $E_0 = 10^5$ GeV, $\gamma = 2.37$
- Even with older versions of the individual datasets, scaled to 12 years of livetime, we still see an excess in the signal over background in DNN Cascades and also Northern Tracks at around 10 and 30 TeV respectively

12.0 Years DNN Cascades

Cygnus
Astrophysical Diffuse
Atmospheric
π^o Best-Fit (IceCube Science, 2023)

IceCube Preliminary

N_{events}

$\log_{10}$ (Reco Energy [GeV])

12.0 Years Northern Tracks

IceCube Preliminary

Cygnus
Astrophysical Diffuse
Atmospheric
π^o Best-Fit (IceCube Science, 2023)

N_{events}

$\log_{10}$ (Reco Energy [GeV])