Nu-sources Analysis Methods

Leo Seen

IceCube Summer School 2026

Introduction

- What is Nu-Sources?
  - Exactly what it sounds like: A subset of neutrino astrophysics related to neutrino sources
  - Outstanding questions
    - How are high-energy neutrinos produced?
    - What environments are necessary for particle acceleration?
    - What types of sources produce neutrinos?
    - Where are the neutrinos coming from?
    - How is neutrino emission tied to other messengers?
    - Whatever else YOU are curious about!
  - Ties to both theory and experiment
  - Broad field that has connections and implications for other fields in astronomy and particle astrophysics
  - Part of a growing field of Multi-Messenger Astrophysics

So you want to find a neutrino source

Step 1: Ideas

- What do you think produces high-energy neutrinos? What is the motivation? Theory driven? Experimental observation driven?
- Is it a single source? Multiple sources?
  - What is special about that single source? Are you looking for emission from a specific source class?
- Is (Are) the source (sources) extended?
  - Do we expect neutrino emission localized to a point in the sky or is the neutrino emission more spread out
- Is (Are) the source (sources) Galactic or extragalactic?
- Is (Are) the source (sources) time-dependent or steady-state?
  - Do we expect a burst of neutrinos at a specific time or do we expect constant emission for a long time

Step 2: Choose your dataset

- What type of data do you want to analyze? What best fits your physics goals?
- IceCube Datasets
  - Enhanced Starting Track Event Selection [ESTES]
  - Deep Neural Network Cascades [DNN Cascades]
  - Northern Tracks [NT]
  - GeV Reconstructed Events with Containment for Oscillation [GRECO]
  - Matter-Enhanced Oscillations with Steriles [MEOWS]
  - Gamma-ray Follow-Up [GFU]
  - Point Source Tracks [PS Tracks]
  - IceCat
- Many multi-dataset analyses nowadays

Step 3: Data analysis [Part I]

- Use statistical methods [See talk by Sam tomorrow]
  - Many methods, some model dependent, some model independent
- In IceCube, we utilize hypothesis testing
  - Must define the null hypothesis and alternative hypothesis
  - From statistics, we can use a maximum log-likelihood approach
  - L(Model Parameters|Data) vs. P(Data|Model Parameters)
- In Nu-Sources, we used an unbinned maximum log-likelihood approach
  - We write the likelihood in terms of probability density functions (PDFs)

$$L(n_s, \gamma) = \prod_i \left(\frac{n_s}{N} \underline{S(\alpha_i, \delta_i, E_i, \sigma_i, \alpha_{\text{src}}, \delta_{\text{src}}, \gamma)} + \left(1 - \frac{n_s}{N}\right) \underline{B(\delta_i, E_i)} \right)$$

Signal PDF   Background PDF

Step 3: Data analysis [Part II]

- Log-likelihoods have nice properties. Define the test-statistic, which tells you how far your observed data deviates from the null-hypothesis

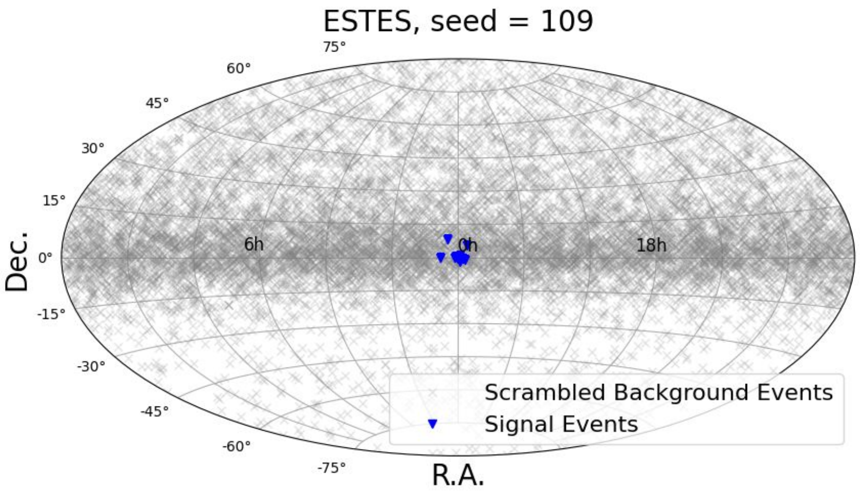
$$\mathrm{TS} = -2\ln\left(\frac{L(n_s = 0, \gamma)}{L(n_s, \gamma)}\right)$$

$$\mathrm{TS} = 2\sum_i \ln\left(\frac{n_s}{N}\left[\frac{S(\alpha_i, \delta_i, E_i, \sigma_i, \alpha_{\mathrm{src}}, \delta_{\mathrm{src}}, \gamma)}{B(\delta_i, E_i)} - 1\right] + 1\right)$$

- Example of TS that may be more familiar is z-score, z = (x-μ)/σ

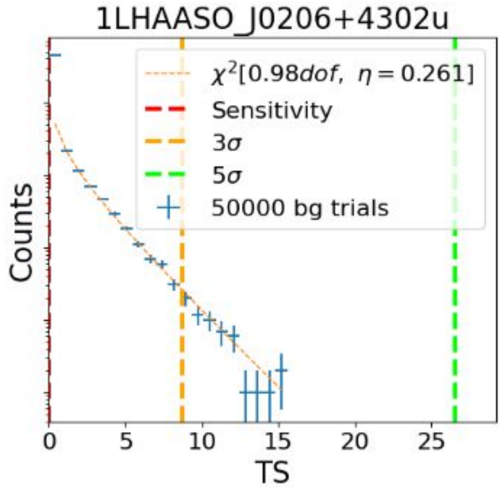
Step 3: Data analysis [Part III]

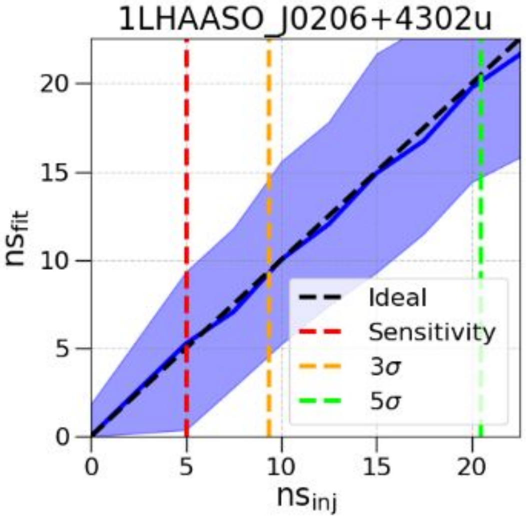
- Each dataset has two sets, Data and Simulation (Monte-Carlo [MC])
  - Run pseudo-experiments on scrambled data and MC
  - We use scrambled data to prevent biases, this is what we call “blind analyses”
  - In Nu-Sources, MC is what we use to simulate our signal
- How to create a pseudo-experiment?
  - Putting neutrinos in the sky [injection]
  - For background, we inject data events scrambled in RA
- For each pseudo-experiment, we fit a TS



Step 3: Data analysis [Part IV]

- Run many pseudo-experiments
  - Background trials - Inject no signal, only scrambled data
  - Signal trials - Inject some signal in addition to background according to your signal hypothesis
- Common plots to make
  - Background TS distribution
    - Used to quantify how likely it is to obtain a TS value by pure chance
  - Signal injection/recovery test
    - Make sure your fitter can recover the truth
  - Sensitivity/3σ/5σ
    - This will tell you, under your signal hypothesis, what set of model parameters we need to detect the signal at varying levels of significance





Step 3: Data analysis [Part V]

- Now this all sounds like a lot of work to build. Fortunately, previous generations have already built software packages that do this
  - csky, SkyLLH, Skylab, Flarestack, i3mla
- All you have to do, is figure out what is happening in the code
- Make sure to document all your work on a github and wikipedia page

Step 3a: New data analysis methods

- Machine learning/AI
  - Transformers, Boosted Decision Trees
  - Anomaly detection
- Joint-fits
  - Utilize data from multiple different detectors

Step 4: Review

- Two stage review process
  - Working group review - Someone in the Nu-Sources working group will look over your wikipedia page and give you comments
  - Collaboration review - Someone in the IceCube collaboration will look over your wikipedia page and give you comments
- Reproducibility review
  - Someone will run your code/scripts to make sure your results are reproducible
- Must outline unblinding plan before unblinding

Step 5: Unblind

- Once you get approval to unblind, you can now look at real data
- Perform analysis on real data and any post-unblinding checks you stated you would do

Step 6: Write paper

- You now have finished your analysis and have your result, it's time to write it up and publish a paper
- Every paper using IceCube data is required to go through the collaboration
- Once you have been in IceCube for a while, ask your advisor to be added to the IceCube author list

Welcome aboard!

- The race is on, good luck!
- Remember to always ask questions
- Exercise is in this github repository
    - cp -r /data/user/lseen/2026_IceCube_Summer_School_Nu_Sources_Tutorial .