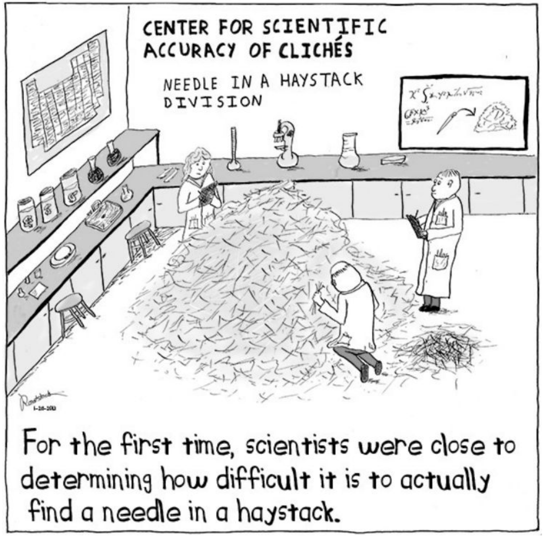
Point Source Likelihood Techniques

"Finding Needles in Haystacks"



Jesse Osborn, IceCube Summer School 2023

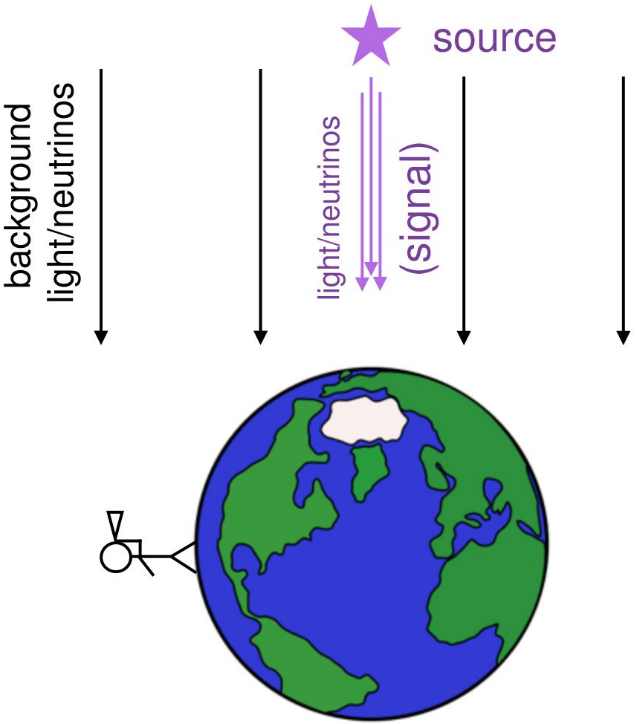
(Copied from IceCube Bootcamp 2022)

The Problem

- Astronomy would be easy if we didn't have background
- For some messengers (high energy photons, neutrinos) we can't turn backgrounds off, but we still want to find sources.

How do we find sources on top of background?

A few definitions

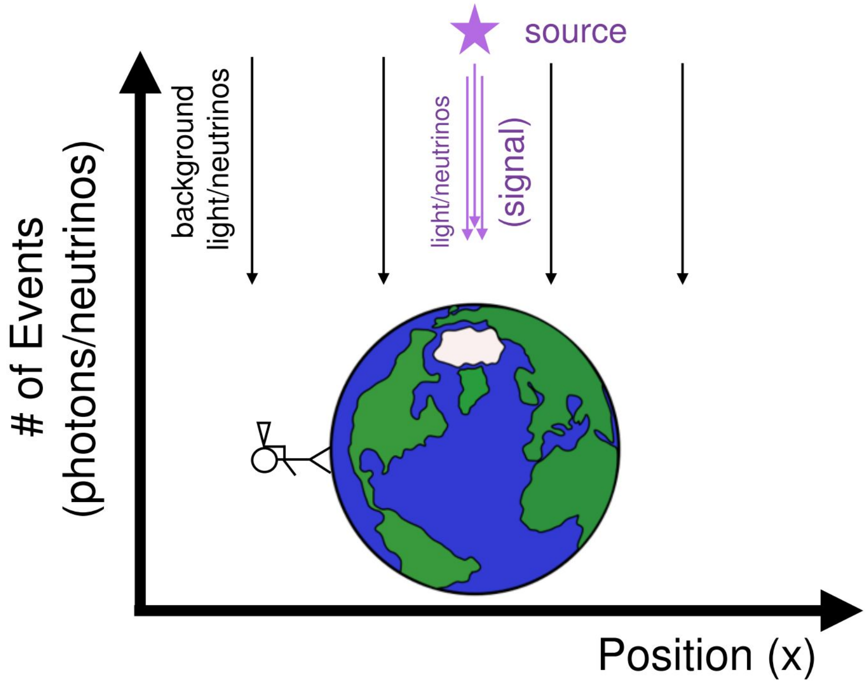


- **Signal** is a particle that came from the source you're looking for
- **Background** is a particle that did not come from the source (but looks identical to a particle emitted by the source)

    Example: a photon or neutrino with same energy as one from the source
- **Event** a detected particle. Can be a photon, neutrino, etc.

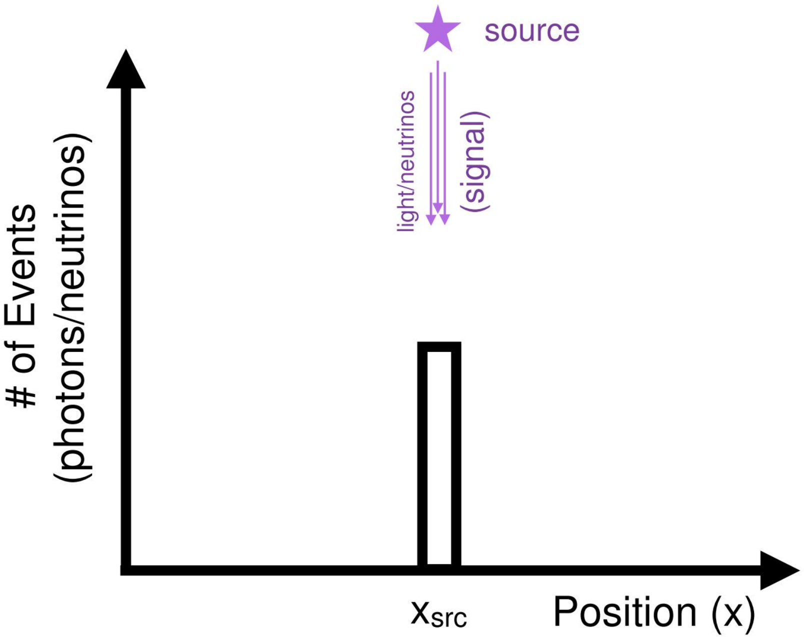
What can we do? Start with spatial distributions:

What would a signal look like in our detector? (Specifically, its spatial distribution)



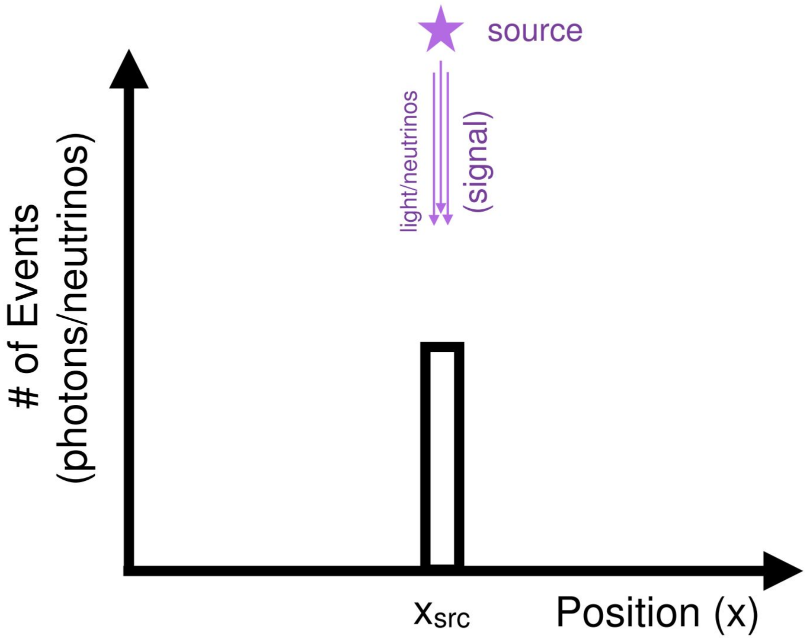
Spatial Distributions: Signal

What would a signal look like?

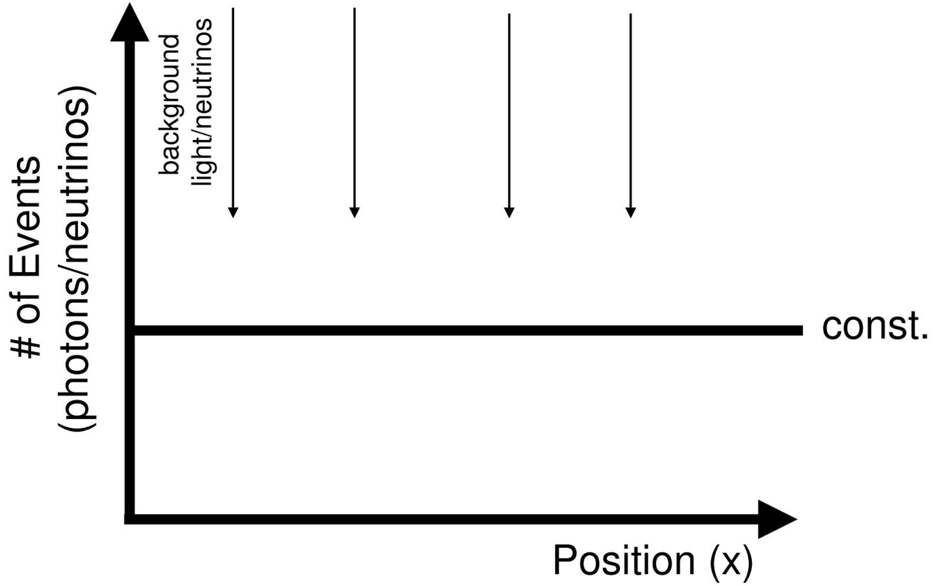


Spatial Distributions: Signal and Background

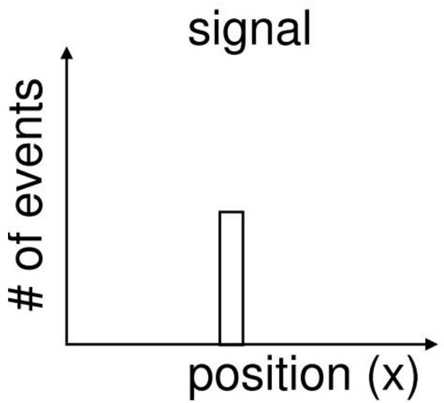
What would a signal look like?

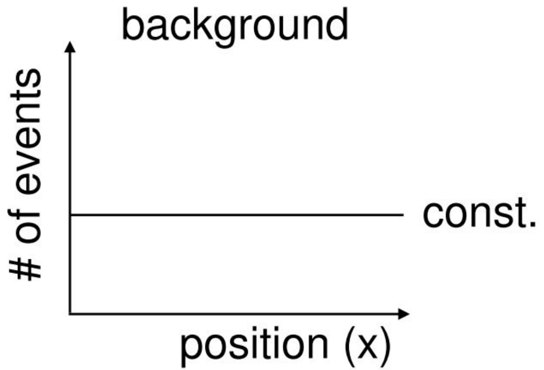


What would our background look like?

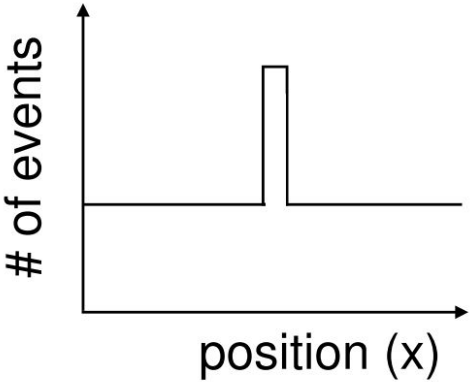


Total Spatial Distribution





Data = signal + background



Needle in a haystack!

Formalism: Probabilities

Now that we know what signal & background distributions look like, we can formulate them in terms of probabilities:

Probability is the chance of getting a given result out of the total number of outcomes.

- ➔ ranges 0 to 1 (never to always)
- ➔ sum of all outcomes must be 1

This way we can ask the question: what is the probability that our data are consistent with **background + signal** versus the case of **background only**?

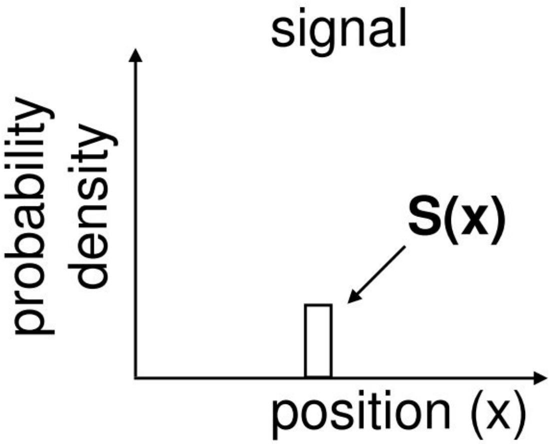
- ➔ Lets us quantify if there is a source in our data

Probability density: Signal

Ok, let's turn our distributions of events into **probability densities** (this means to scale our probabilities such that integral of the distribution is 1)

S(x) = probability density of finding signal at x

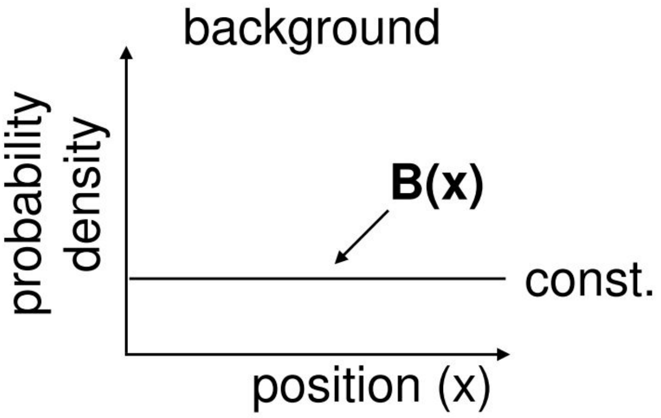
S(x) dx = probability of finding signal within dx of x



Probability density: Background

B(x) = probability density of finding background at x

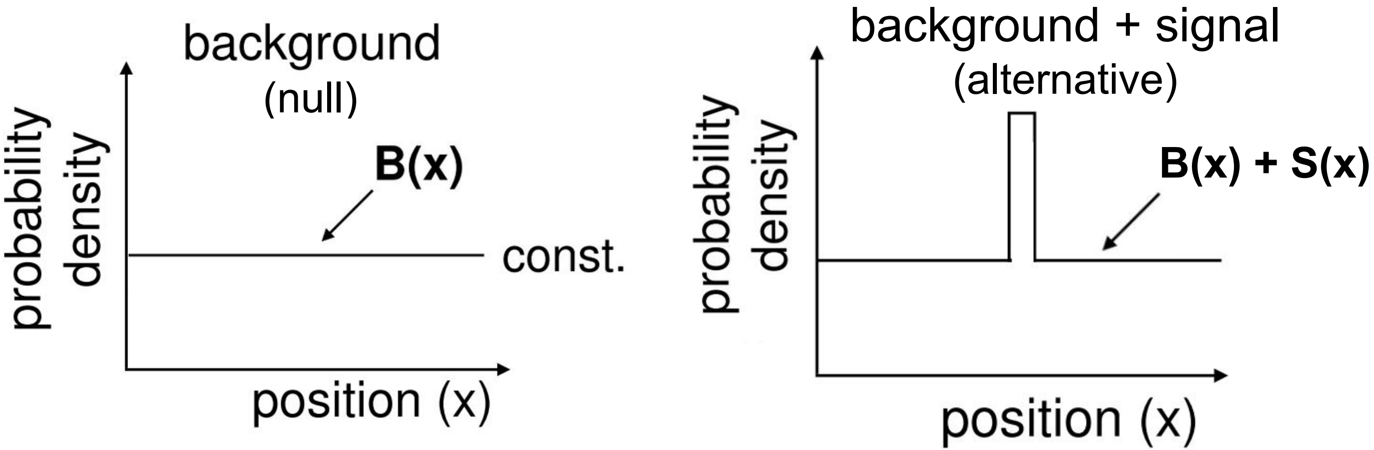
In astronomy, we typically work on the surface of a sphere, so our uniform **B(x)** = $1/4\pi$



Looking for a signal: Hypothesis testing

Null: All events are isotropically distributed and background-like (no signal events)

Alternative: Events are clustered
- Can be clustered in either time or space
- Can follow a particular expected source distribution (spatial prior/extended source - we'll talk about these in a few slides)



Bringing it all together: the likelihood function

For a dataset with:

- **N** total events (**known**)
- **n_s** signal events (**unknown**)
- **x_i** is the position where we detect the ith event ($i \in [1, N]$) (**known with error**)

total ith signal prob.

total ith background prob.

$$\mathcal{L}(n_s) = \prod_{i=1}^{N} \frac{n_s}{N} * S(x_i) + \left(1 - \frac{n_s}{N}\right) * B(x_i)$$

Probability that the ith event is signal

Probability density of signal at xi

Probability that the ith event is background

Probability density of background

The best estimate for the true value of n_s is the value which maximizes the likelihood.

How do we use the likelihood in an analysis?

Working with ratios of likelihoods has some nice statistical properties.

Using the **log of the likelihood** allows us to add instead of multiply, and the log likelihood is maximized at the same place as the likelihood function.

Then we can define our **Test Statistic (TS).** Finding the value of $\mathbf{n_s}$ which maximizes TS is equivalent to maximizing the likelihood. TS = 0 means consistent with background only, while high TS values (~25) can be proof of a source.

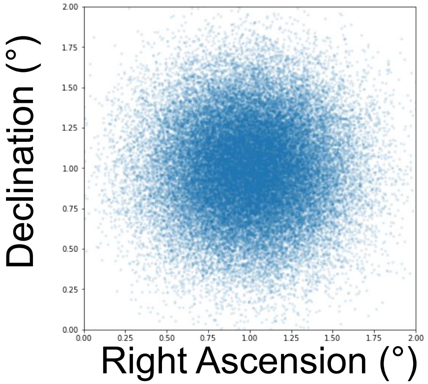
$$TS = 2\log\left[\frac{\mathcal{L}(n_s)}{\mathcal{L}(n_s = 0)}\right] = 2\sum_{i=1}^{N}\log\left[\frac{n_s}{N} * \frac{S(x_i)}{B(x_i)} + \left(1 - \frac{n_s}{N}\right)\right]$$

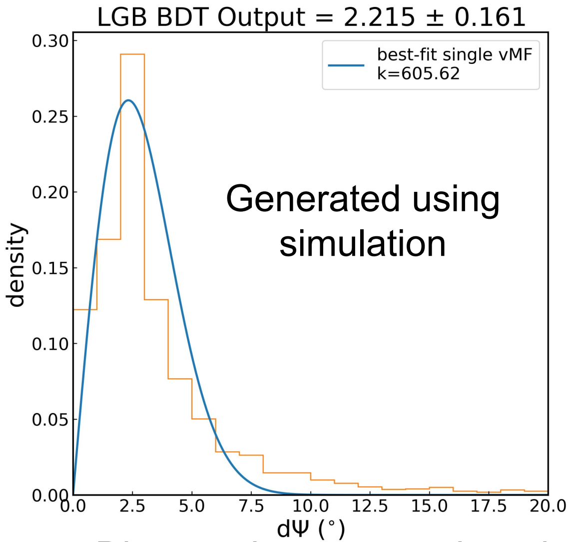
Brief overview of different source types we look at,
and the collaboration-specific code packages we commonly use

Point Source Searches in IceCube

Angular Error Reconstruction

- Do not have perfect reconstructions of our events
- Goal: estimate the true PSF
- Point Spread Function: function that describes the distribution of distances between the reconstructed and true direction
- Usually simplify and use a 2d gaussian (more formally: Kent/von Mises Fischer)
    - We know there are errors to this approximation
- Have now created more advanced methods to approximate PSF (see SkyLLH)





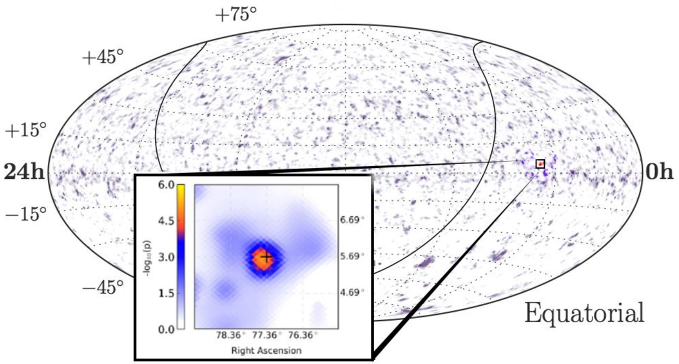
Distance between truth and reconstructed directions

Nu-sources code packages

- Skylab / csky / SkyLLH / FlareStack: likelihood frameworks for point source analyses
  - Examples for each of these codes in the `docs/` folders on GitHub
  - Many people start with csky
  - Csky has multiflare stacking, SkyLLH has KDE PSF modeling
- grbllh / psLab: legacy LLH code
- FIRESONG: simulation package for extragalactic neutrino sources
  - Examples in `notebooks/` folder on GitHub

Source types: Point Sources

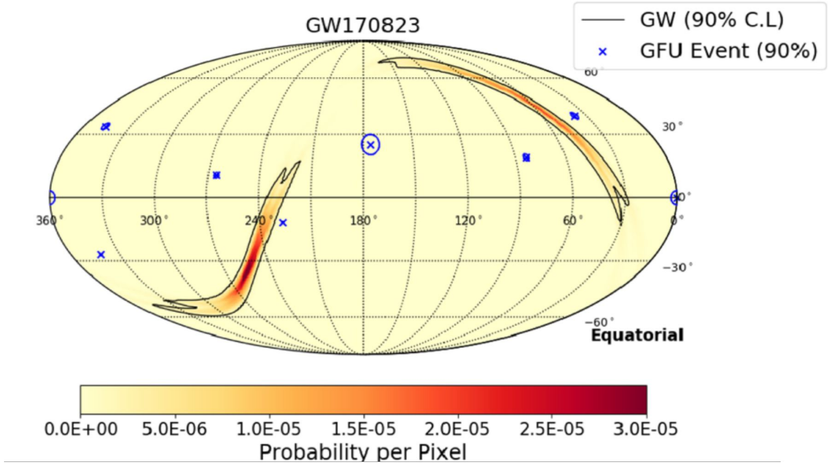
- Looking for a single, delta-function like source in our data
- Often do this by comparing IceCube data to a particular source (or catalog of sources), or taking a specific dataset and doing a **Hotspot scan** or **all-sky scan**
  - Hotspot scan (or all-sky scan): searching across the entire sky for the hottest spot in our data



Example: TXS 0506+056,
from the point source follow up wiki

Source types: Spatial Prior

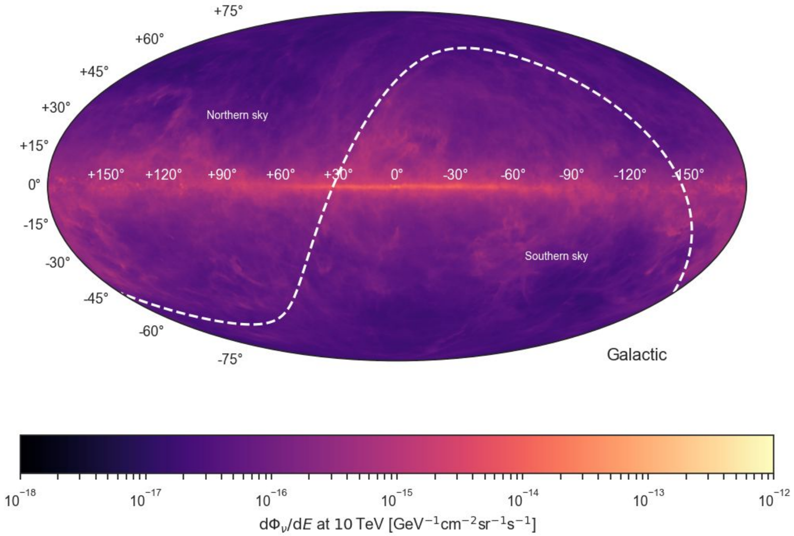
- Searching for a point source in a **localization region** (sometimes called a spatial prior)
- Still looking for a delta function-like point source, but we have a probability region to search for it
- Examples: gravitational wave follow ups, using sources with spatial uncertainty (e.g. CHIME FRBs)
- Do this by adding a spatial weight term to our likelihood (note that w_L is less than 1. so the log is negative): $\mathrm{TS}' = \mathrm{TS} + 2\ln(w_{\mathrm{L}})$



M. G. Aartsen *et al* 2020 *ApJL* **898** L10

Source types: Extended source

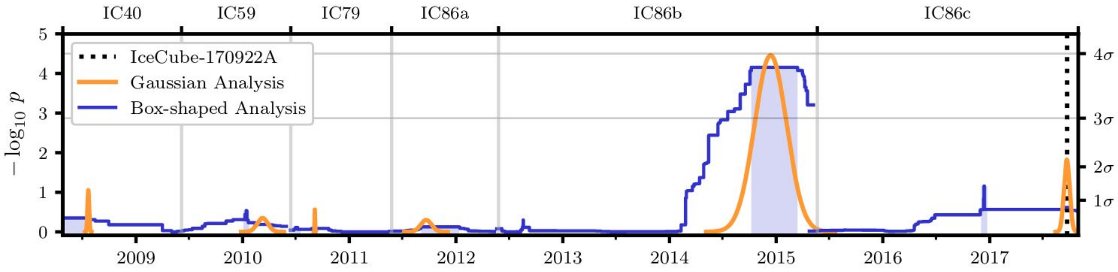
- Looking for a source that emits more widely (not a delta function)
- Example: Galactic Plane Diffuse Neutrino Emission, Extended TeV Gamma Ray Sources, Nearby Galaxies



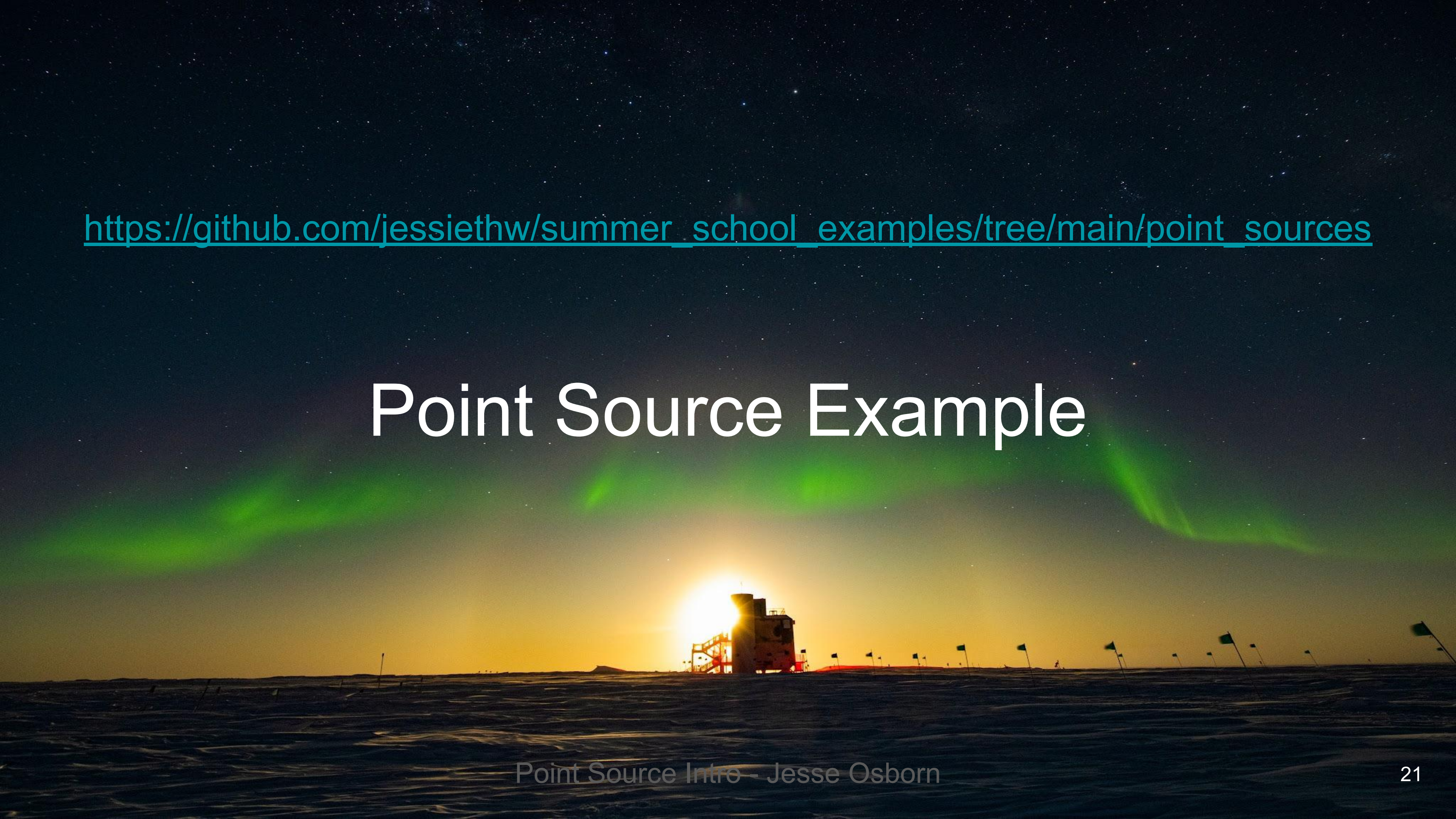
Fermi π0 model of neutrino emission in the galactic plane (based on Fermi-LAT diffuse gamma-ray π0 contribution)

Other types of searches

- Time-dependent search - looking for emission with a particular time window (usually the time window is either theory driven, or based on maximizing our sensitivity, or a combination of both)
- Time-integrated search - uses many years (or entire livetime) of data
- Flare search - looking for multiple events corresponding to a flare in our data
- Template search - looking for emission corresponding to a template



Plot stolen from Will Luszczak, [Multiflare stacking applied to TXS 0506+056 flares]



https://github.com/jessiethw/summer_school_examples/tree/main/point_sources

Point Source Example