Time-Dependent Maximum Likelihood Analysis

Jim Braun

Spring 2009 IceCube Collaboration Meeting

Outline

- Review of UW time-dependent likelihood searches
  - **Triggered** searches with an assumed time dependence
  - **Untriggered** searches for an unknown time dependence

- Problem encountered with the untriggered search, along with solution (~30% gain in discovery potential)

- 5σ discovery potential comparison of triggered and untriggered searches

Triggered Search

- Believe the signal has some assumed time dependence (e.g. TeV γ flares, GRB, etc.)

$$\mathcal{S}_i(\vec{x}_s, \vec{x}_i, \gamma) = \underbrace{\frac{1}{2\pi{\sigma_i}^2} e^{\frac{|\vec{x}_i - \vec{x}_s|^2}{2{\sigma_i}^2}}}_{\text{Space Angle}} \underbrace{P(E_i|\gamma)}_{\text{Energy}} \underbrace{\frac{1}{\sqrt{2\pi}\sigma_t} e^{\frac{|t_i - t_\circ|^2}{2{\sigma_t}^2}}}_{\text{Time}}$$

- Dependence could be Gaussian, rectangular, etc.

$$\mathcal{L}_i(\vec{x}_s, \vec{x}_i, \gamma) = \prod_{i=1}^{N} \left(\frac{n_s}{N} \mathcal{S}_i + (1 - \frac{n_s}{N}) \mathcal{B}_i \right) \qquad \mathcal{B}_i = \frac{P(E_i|\Phi_{atm})}{\Omega T_L}$$

- Maximize likelihood with respect to signal strength and spectral index

$$\lambda = -2 \cdot sign(\hat{n}_s) \cdot log \left[\frac{\mathcal{L}(\vec{x}_s, 0)}{\mathcal{L}(\vec{x}_s, \hat{n}_s, \hat{\gamma})} \right]$$

Untriggered Search

- Assume the signal has some unknown time dependence
  - Assume clusters of signal-like events are flares
  - Reduce background and thus the number of events needed to reach 5σ
  - Tightly clustered signals can be discovered which would be missed in the time integrated search

- Assume time dependence is a single flare with a Gaussian distribution in time

- Maximize likelihood with respect to time and width along with signal strength and spectral index

$$\lambda = -2 \cdot log \left[\frac{\mathcal{L}(\vec{x}_s, n_s = 0)}{\mathcal{L}(\vec{x}_s, \hat{n}_s, \hat{\gamma}, \hat{\sigma}_t, \hat{t}_o)} \right]$$

Problem with Untriggered Search

- Two symptoms suggest the method can be improved
  - The background-only distribution of the test statistic is not χ^2 distributed
  - High significance background fluctuations are overwhelmingly pairs rather than triplets, etc.

- We know there is a trial factor we do not account for
  - We select one time out of the data livetime as our flare time, but any other time is equally likely

Problem with Untriggered Search

- Our signal hypothesis is subtly incorrect:
  - Current: "What is the likelihood of observing a flare at the best fit time"

$$\lambda = -2 \cdot log\left[\frac{\mathcal{L}(\vec{x}_s, n_s = 0)}{\mathcal{L}(\vec{x}_s, \hat{n}_s, \hat{\gamma}, \hat{\sigma}_t, \hat{t}_o)}\right]$$

  - Correct: "What is the likelihood of observing a flare at ANY time"
  - Suggests we should marginalize likelihood with respect to time:

$$\mathcal{L}(\vec{x}_s, \hat{n}_s, \hat{\gamma}, \hat{\sigma}_t) = \int_t P(Data|\hat{n}_s, \hat{\gamma}, \hat{\sigma}_t, t)P(t)dt$$

$$\lambda = -2 \cdot log\left[\frac{\mathcal{L}(\vec{x}_s, n_s = 0)}{\mathcal{L}(\vec{x}_s, \hat{n}_s, \hat{\gamma}, \hat{\sigma}_t)}\right]$$

  - P(t) time prior is uniform: $P(t) = 1/T_L$

Problem with Untriggered Search

- The integral over time is numerically inconvenient
- Can approximate the integral by assuming all contributions are from the vicinity of the likelihood maximum, which itself is approximately Gaussian:

$$\int_t P(Data|\hat{n}_s, \hat{\gamma}, \hat{\sigma}_t, t) P(t) dt \simeq \frac{\sqrt{2\pi}\sigma_{t_o}}{T_L} P(Data|\hat{n}_s, \hat{\gamma}, \hat{\sigma}_t, \hat{t}_o)$$

- σ_{t_o} is the uncertainty on the best fit mean time
- New test statistic improves discovery potential ~30% and is χ^2 distributed

$$\lambda = -2 \cdot log \left[\frac{T_L}{\sqrt{2\pi}\sigma_{t_o}} \frac{\mathcal{L}(\vec{x}_s, n_s = 0)}{\mathcal{L}(\vec{x}_s, \hat{n}_s, \hat{\gamma}, \hat{\sigma}_t, \hat{t}_o)} \right]$$

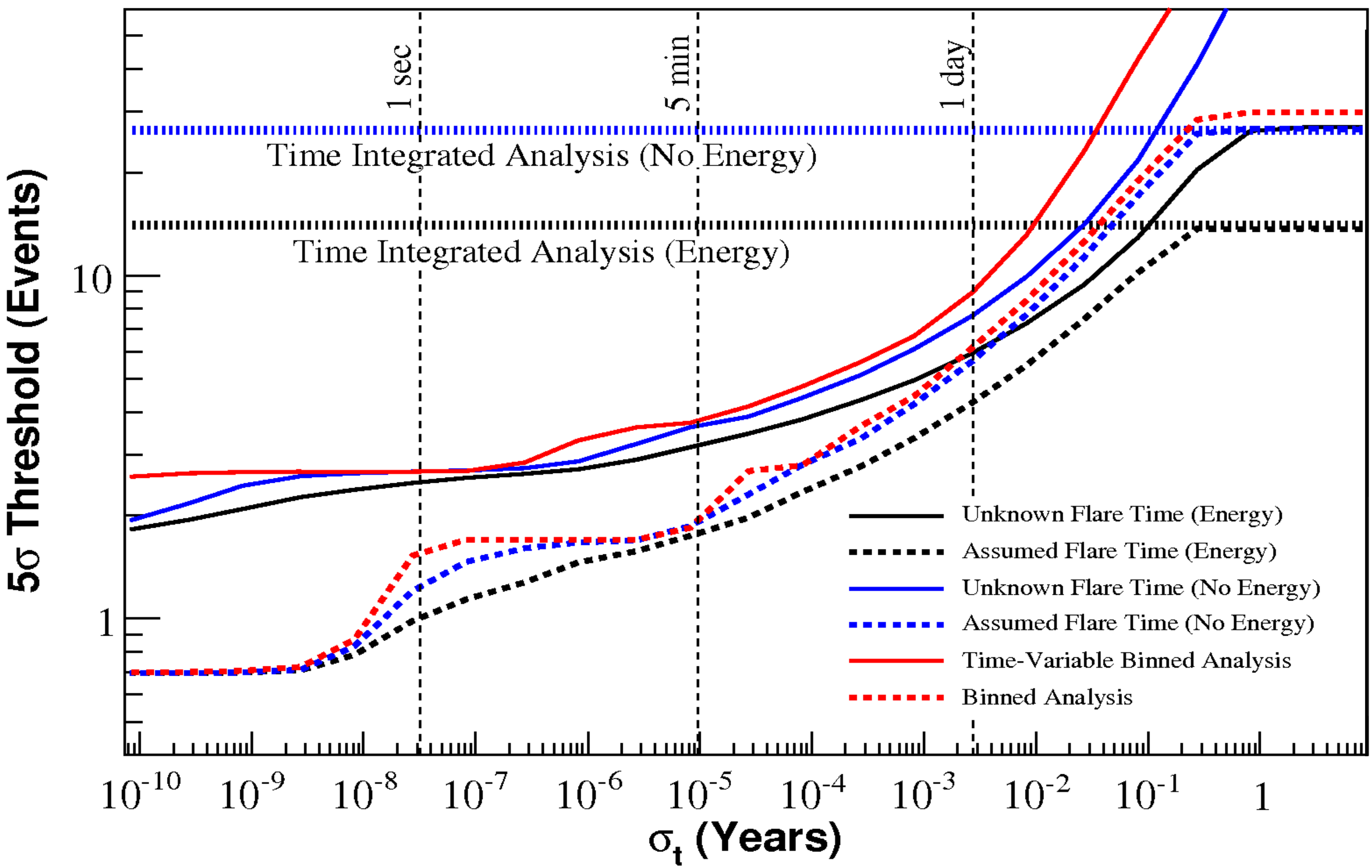
Time-Dependent Analysis for km^3

- Simulate a ~km^3-scale detector skymap with 67,000 events (~1 year, 0.7^{o} angular resolution)

- Compare the mean Poisson number of events needed for 50% chance of 5σ discovery
  - Both triggered and untriggered methods
  - With and without energy information
  - Compare to binned analyses as a reference

- Compare over a large range of flare widths

Time-Dependent Analysis for km³



The All-Sky Trials Factor

- All sky search:
  - Find maximum test statistic value over the sky

$$\lambda = -2 \cdot log\left[\frac{\mathcal{L}(\vec{x}_s, n_s = 0)}{\mathcal{L}(\vec{x}_s, \hat{n}_s, \hat{\gamma}, \hat{\delta}, \hat{\alpha})}\right]$$

  - Measure post-trials significance by comparing to many skymaps randomized in RA
  -
- Apply the untriggered search solution to all-sky analysis
  - Marginalize with respect to RA and declination

$$\lambda = -2 \cdot log\left[\frac{\Omega}{2\pi\sigma_r^2}\frac{\mathcal{L}(\vec{x}_s, n_s = 0)}{\mathcal{L}(\vec{x}_s, \hat{n}_s, \hat{\gamma}, \hat{\delta}, \hat{\alpha})}\right]$$

  - Would give χ^2 distribution for post-trials significance
  - No need for thousands of randomized trials
-
- Currently working to verify this