# Statistical Challenges for GLAST/LAT Data

Jim Chiang
GSSC/CRESST/UMBC

## Properties of LAT Data

- **Energy coverage 30**–$3 \times 10^5$ **MeV**
- **Strongly energy-dependent PSF:**

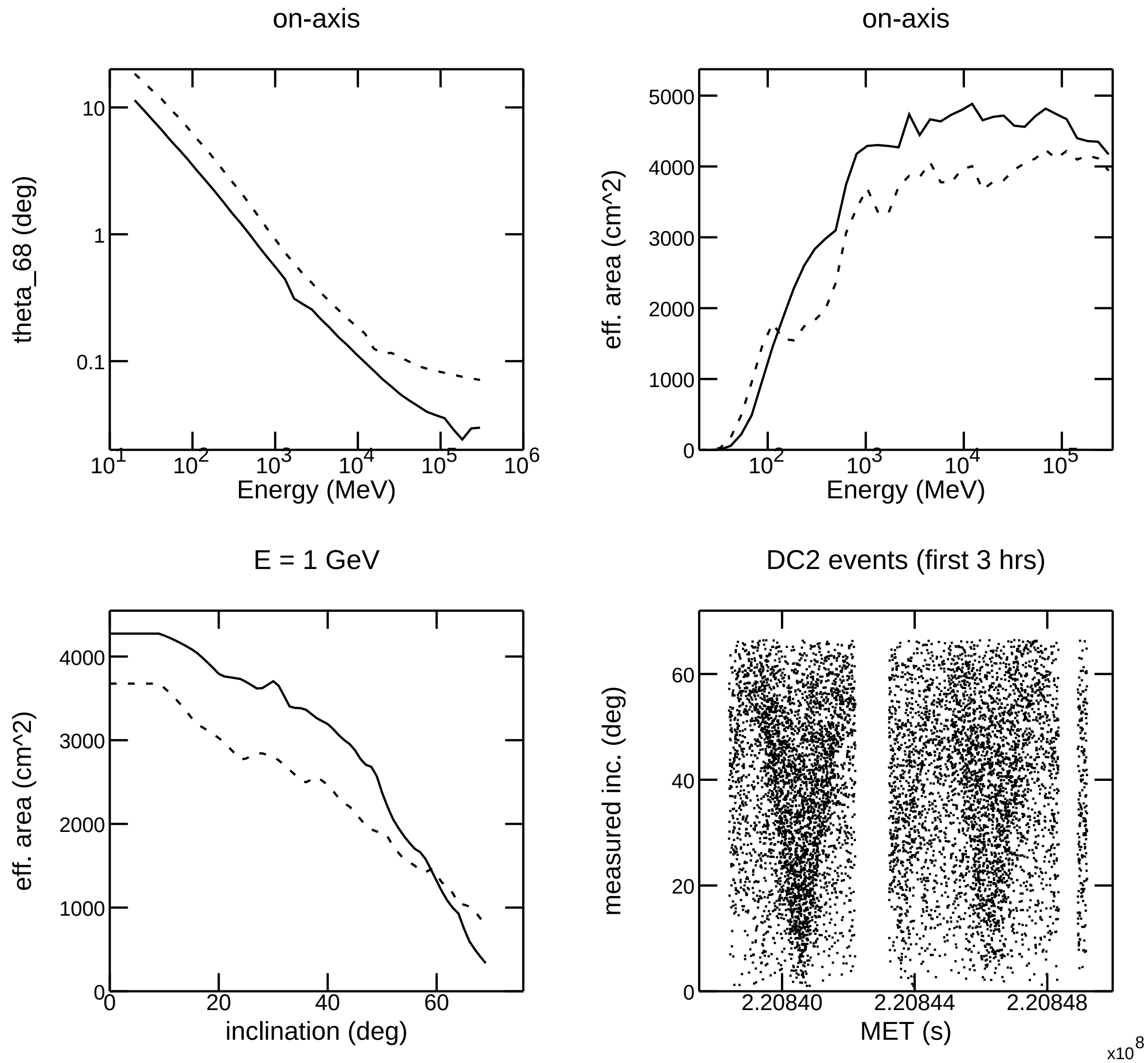
$$\begin{aligned} \theta_{68} &< 3.5^\circ \qquad \textbf{at 100 MeV} \\ &< 0.15^\circ \qquad \textbf{at 10 GeV} \end{aligned}$$

- **Effective area** $\sim 10^4$ **cm**$^2$ **at 10 GeV,** $\underset{\sim}{\propto} \cos\theta$**.**
- **Energy Resolution** $\lesssim 10\%$
- **Field-of-view** $\gtrsim 2\,$**sr**

$\Rightarrow$ **total count rate** $\sim 1$–**2 events** $\mathrm{s}^{-1}$

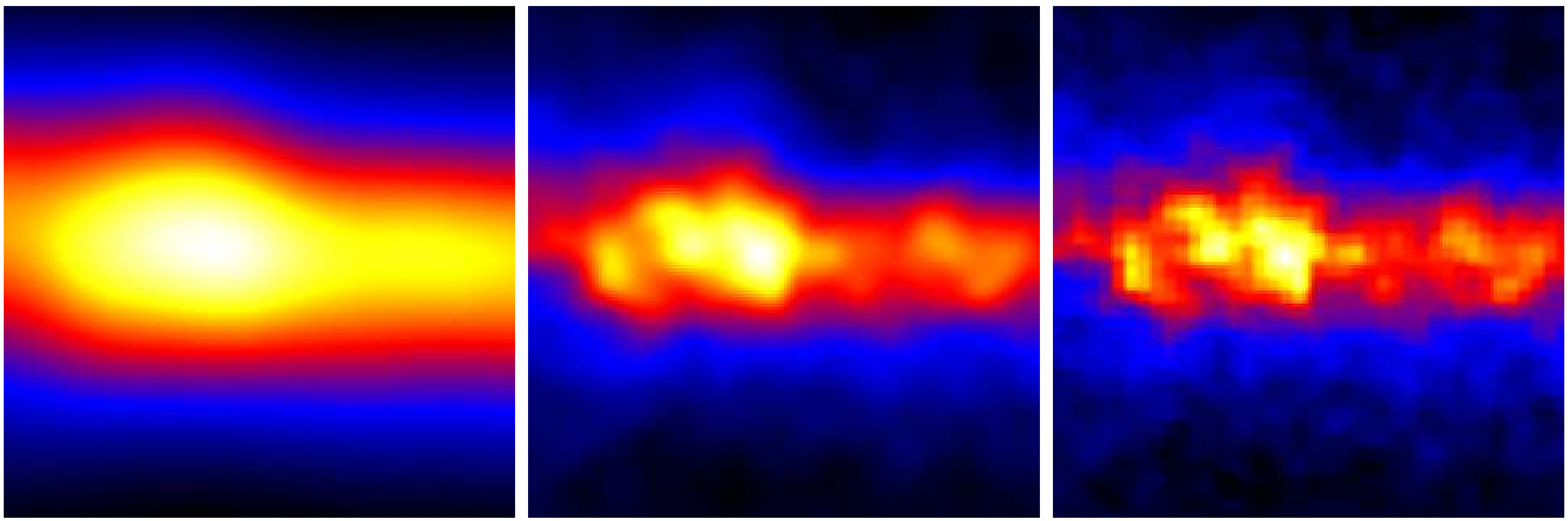
- **Default mode of operation: scanning with 95 min. orbit and** $\pm 35$ **rocking** $\Rightarrow$ **continuous aspect changes of** $\gtrsim 4^\circ$ **min**$^{-1}$**.**

# Instrument Response Functions



## Central Problem

**Discerning and measuring point sources and extended structures on angular scales smaller than that of the PSF in the Poisson limit.**



100 MeV 1000 MeV 10000 MeV

GALPROP model of Cygnus region convolved with the LAT PSF.

## Statistical Issues

- **Parameter estimation and confidence limits**

$\star$ **The LRT null distribution and the applicability of Wilks Theorem**

- **Source detection**
- **Source localization and identification**

$\star$ **Image reconstruction and the significance of extended features**

$\star$ **Detecting and characterizing variability**

- **Source-specific analyses:**
  - **GRB (blind search, timing)**
  - **Pulsar blind search**
  - **Dark Matter detection**
- **Track reconstruction**

## LRT and the Null Distribution

- **We wish to find the reference distribution under the null hypothesis. Assuming a model similar to**

$$p(x;\eta,\theta) = (1-\eta)f(x;\theta) + \eta\psi(x), \tag{1}$$

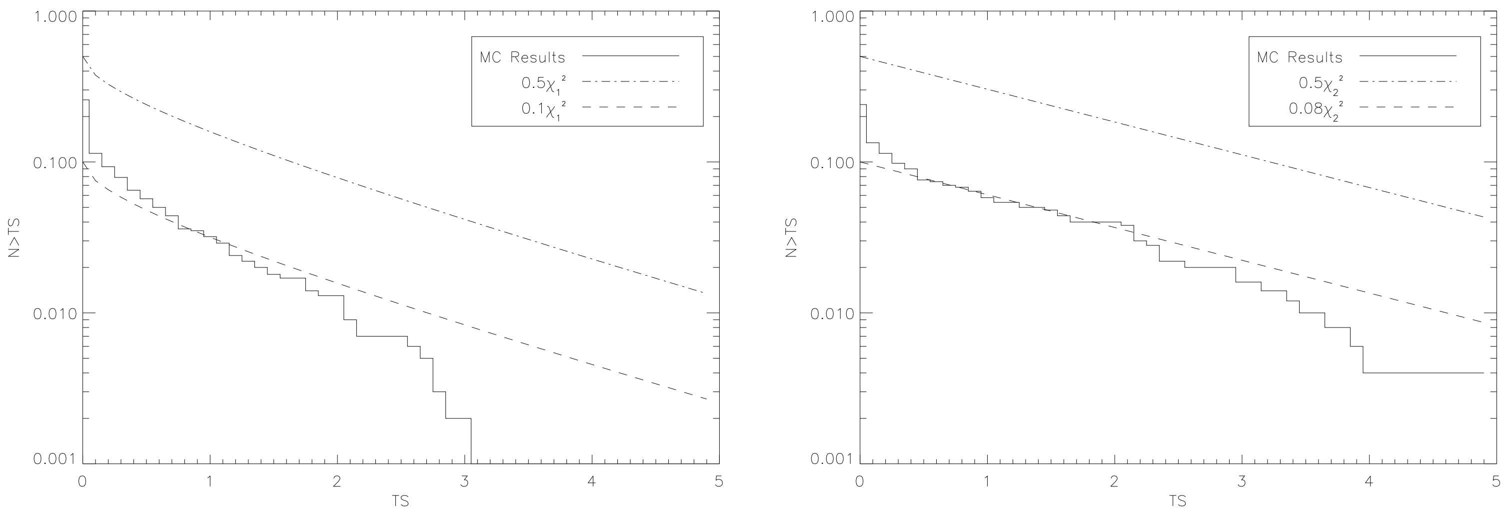
**Mattox et al. (1996; also Protassov et al. 2002) found that a mixture model of the form**

$$\chi_0^2/2 + \chi_1^2/2 \tag{2}$$

**is the desired distribution.**
**We consider a somewhat more complex model. Effectively, we have**

$$p(x;\eta,\theta,\lambda) = (1-\eta)f(x;\theta) + \eta\psi(x;\lambda) \tag{3}$$

**Here $\lambda$ represents the energy dependence ($\sim 1/E$) of the PSF. Our simulations yield distributions that differ significantly from the simple mixture model.**



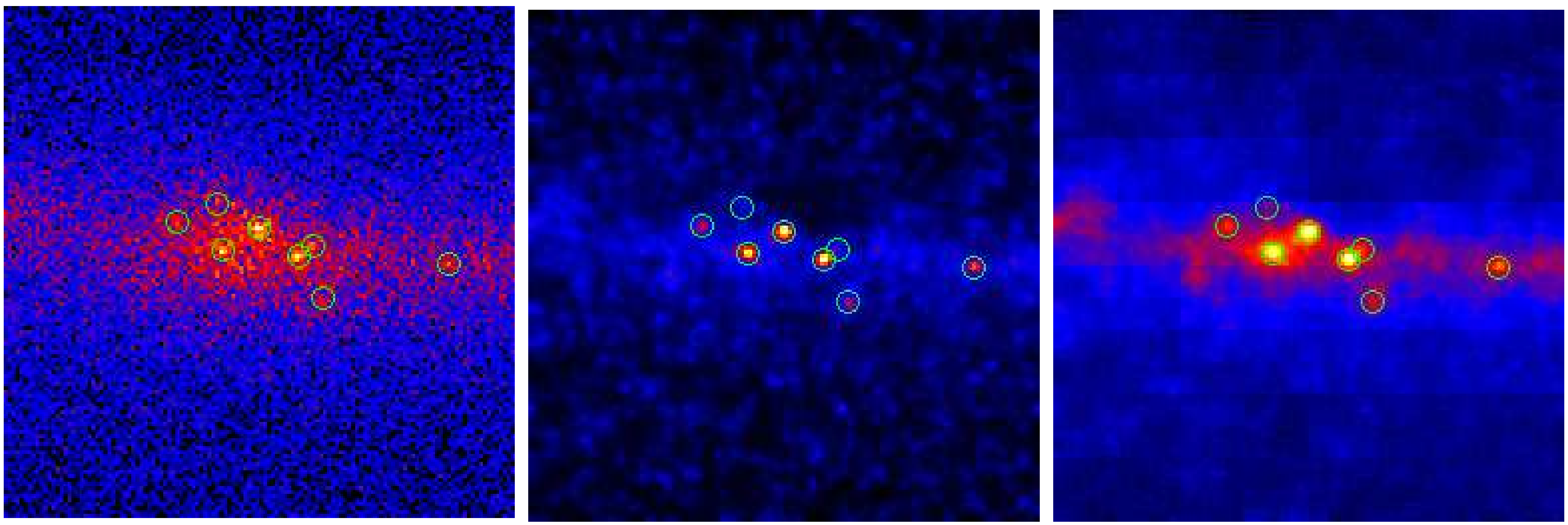

- **Possible solution: Pilla, Loader, & Taylor (2005, PRL, 95) propose a new test statistic based on a score process that is asymptotically equivalent to LRT. They can calculate the asymptotic reference distribution as a sum of weighted chi-square distributions if given the null $f(x;\theta)$ and signal $\psi(x;\lambda)$ densities.**

## Image Reconstruction

**Interstellar emission comprises the bulk of the celestial gamma-rays we expect to detect. Based on radio observations of the gas content along lines-of-sight through the Galaxy, the interstellar emission is expected to have significant features on all scales, including those of the PSF. However, we do not have a perfect model of the interstellar component.**

- **How do we distinguish point sources from diffuse sources?**
- **How can we best deconvolve the data from the instrument response?**
- **What is the significance of the underlying features that we find?**

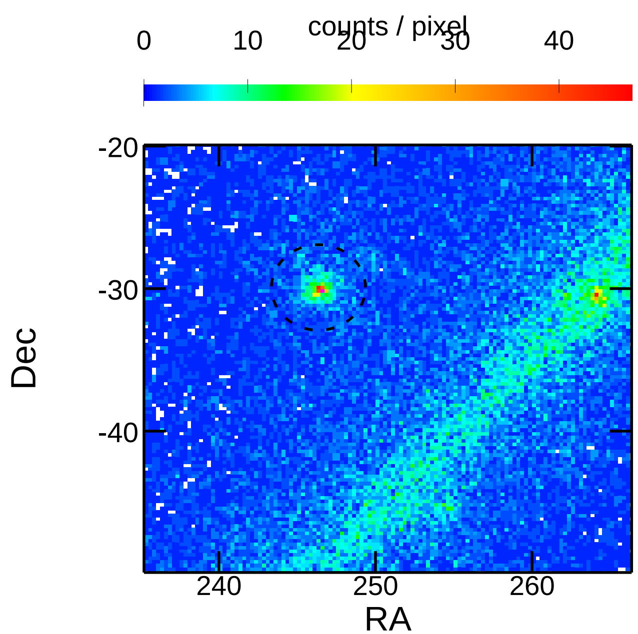
**Consider a LAT simulation of the Cygnus Region (with 3EG sources), integrated from 30 to 300 GeV:**

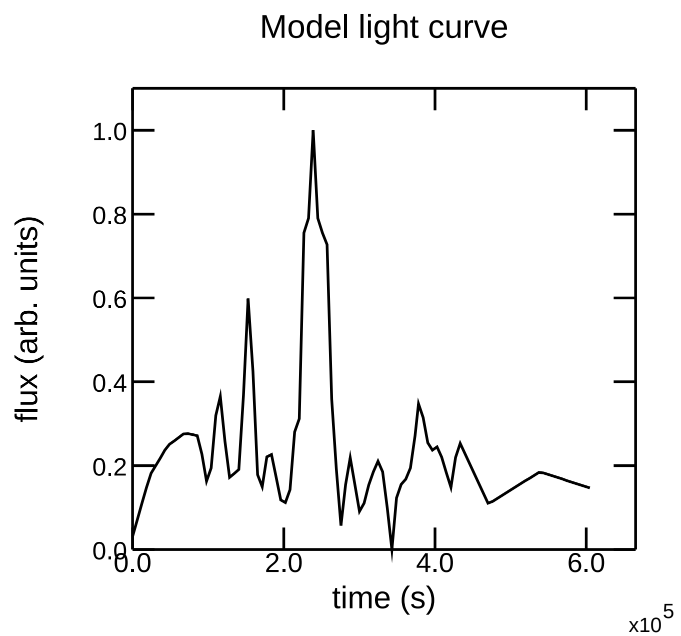


Simulated counts map RL reconstruction Multiscale reconstruction

# Transient Source Detection

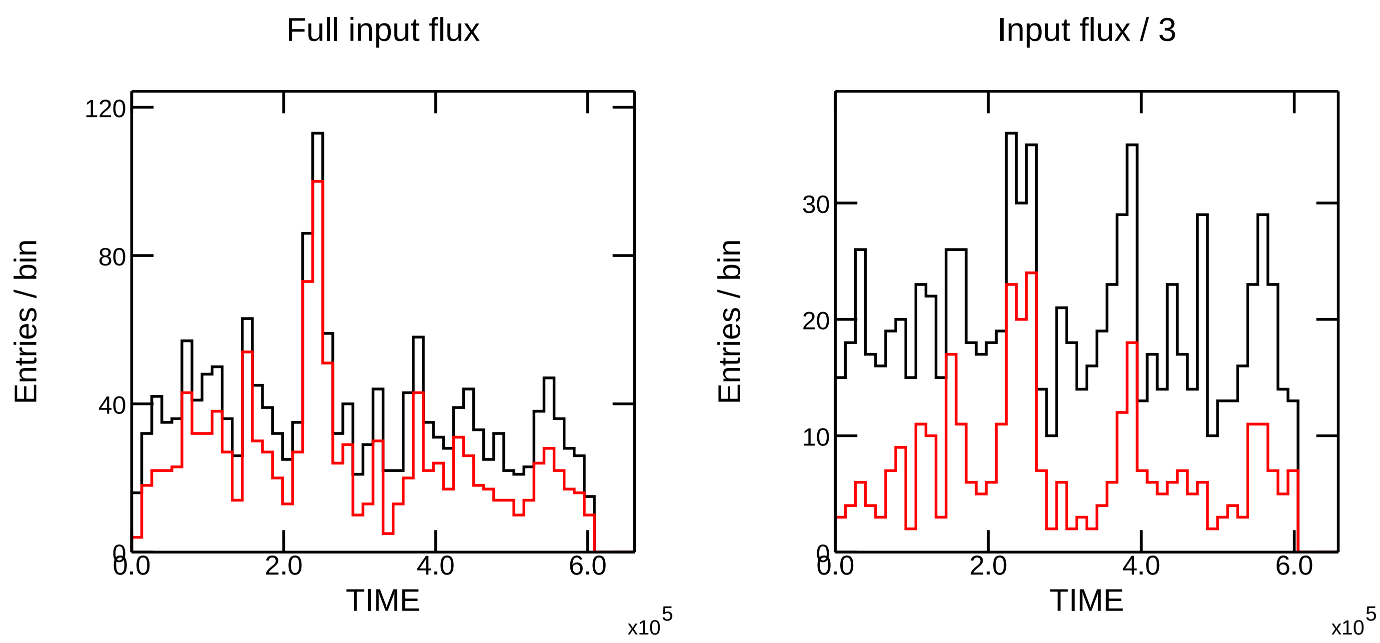
- GLAST is an all-sky monitor. We need an efficient algorithm for detecting variability from both known and previously unknown sources. The scanning and orbital motion and inclination dependence of the effective area will cause significant count rate variations even for a non-varying sky.
- If we monitor count rates for patches of the sky, how sensitively can we detect variability within a given patch? (In principle, we can compute the exposure variation as a function of time very precisely.)

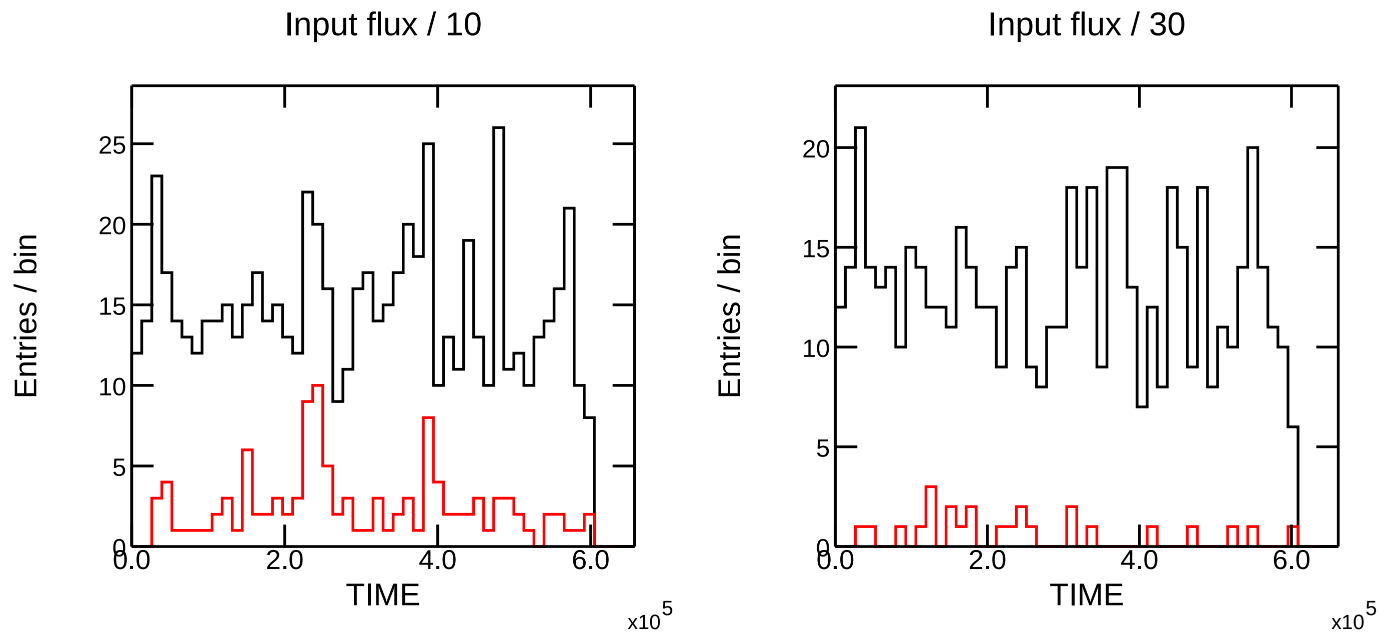



Monitor a 3 deg radius circle about this source, PKS 1622-297

- 1 week simulations, with interstellar, extragalactic diffuse and 3EG point sources as steady components + 1 flaring blazar.
- Note extremely low count rates: 81 counts/day for steady sources and 170 counts/day (mean) for flaring blazar at full strength.




Input flux / 10
Entries / bin
TIME
x10^5
0
5
10
15
20
25
0.0
2.0
4.0
6.0
Input flux / 30
Entries / bin
TIME
x10^5
0
5
10
15
20
0.0
2.0
4.0
6.0

# Statistical Issues Recap

- **Parameter estimation and confidence limits (P8.5, P19.23, P19.32, P19.40)**
- **The LRT null distribution and the applicability of Wilks Theorem**
- **Source detection (P13.7, P19.9, P19.14, P 19.22, P19.38)**
- **Source localization and identification (P5.4, P13.4, P19.6)**
- **Image reconstruction and the significance of extended features (P2.6, P3.1, P8.3, P8.4)**
- **Detecting and characterizing variability**
- **Source-specific analyses:**
  - **GRBs (blind search, timing) (P8.7, P16.1, P16.11, P18.5, P19.20)**
  - **Pulsar blind search (4.3, P3.2, P8.8, P19.15)**
  - **Dark Matter detection (P19.40, P19.32)**
- **Track reconstruction (P8.6, P19.39)**