Likelihood Methods for the Detection
and Characterization of Gamma-ray Pulsars
with the *Fermi* Large Area Telescope

Matthew Kerr

A dissertation submitted in partial fulfillment
of the requirements for the degree of

Doctor of Philosophy

University of Washington

2010

Program Authorized to Offer Degree: Department of Physics

University of Washington

Graduate School

This is to certify that I have examined this copy of a doctoral dissertation by

Matthew Kerr

and have found that it is complete and satisfactory in all respects,
and that any and all revisions required by the final
examining committee have been made.

Chair of the Supervisory Committee:

_____

Thompson Burnett

Reading Committee:

_____

Thompson Burnett

_____

Scott Anderson

_____

Miguel Morales

Date: _____

University of Washington

Abstract

Likelihood Methods for the Detection
and Characterization of Gamma-ray Pulsars
with the *Fermi* Large Area Telescope

Matthew Kerr

Chair of the Supervisory Committee:
Professor Thompson Burnett
Physics

The sensitivity of the *Large Area Telescope* (LAT) aboard the *Fermi Gamma-ray Space Telescope* allows detection of thousands of new γ-ray sources and detailed characterization of the spectra and variability of bright sources. Unsurprisingly, this increased capability leads to increased complexity in data analysis. Likelihood methods are ideal for connecting models with data, but the computational cost of folding the model input through the multi-scale instrument response function is appreciable. Both interactive analysis and large projects—such as analysis of the full gamma-ray sky—can be prohibitive or impossible, reducing the scope of the science possible with the LAT.

To improve on this situation, we have developed *pointlike*, a software package for fast maximum likelihood analysis of *Fermi*-LAT data. It is interactive by design and its rapid evaluation of the likelihood facilitates exploratory and large-scale, all-sky analysis. We detail its implementation and validate its performance on simulated data. We demonstrate its capability for interactive analysis and present several all-sky analyses. These include a search for new γ-ray sources and the selection of LAT sources with pulsar-like characteristics for targeted radio pulsation searches. We conclude by developing sensitive periodicity tests incorporating spectral information obtained from *pointlike*.

TABLE OF CONTENTS

LIST OF FIGURES

iv

LIST OF TABLES

GLOSSARY

LAT: Large Area Telescope, a pair-conversion telescope

HE: high energy; a class of γ rays with energies between about 100 MeV and 100 GeV; the passband of the *Fermi*-LAT

GEV: 10^9 electron Volts; in the text "GeV" is also used synonmously with "HE" for describing the *Fermi*-LAT passband, e.g., *Fermi*-LAT is a GeV telescope

IRF: instrument response function; a statistical description of the measurement outcomes for an incident photon, including (a) the probability of an interaction and successful event reconstruction and (b) the distribution of measured variables about their true values

PSF: point spread function; a probability density function describing the random spread in reconstructed position of photons incident from a point source

TKR: the tracker, a section of *Fermi*-LAT comprising interleaved tungsten foils and silicon strip detectors

CAL: the calorimeter, a section of *Fermi*-LAT comprising a hodoscopic array of CsI crystals

ACD: anticoincidence detector; a component of *Fermi*-LAT comprising segments of scintillator designed to signal the passage of a charged particle into the detector

FITS: Flexible Image Transport System, a standard for tabular data and images

CALDB: a scheme for storing calibration information for instruments and telescopes

FT1: a high-level *Fermi*-LAT data product in FITS format; columns give reconstructed quantities (e.g., energy, time) for events

FT2: a high-level *Fermi*-LAT data product in FITS format; contains information about the spacecraft's orientation and position as a function of time

L: Galactic longitude, a measure of azimuth with origin at Earth referenced to the Galactic center

B: Galactic latitude, a measure of polar angle with origin at Earth referenced to Galactic center

S/C: spacecraft, in particular the vehicle of the *Fermi*-LAT

GTI: Good Time Interval, a tuple of a start and end time; events with reconstructed times lying within the interval pass some set of criteria to be "good" events

ML: maximum likelihood; a statistical technique for estimating parameters of a distribution

CT: conversion type, a classification for a reconstructed *Fermi*-LAT event indicating whether the event converted in a thin tungsten foil in the "front" of the TKR ($CT = 0$) or in a thick tungsten foil in the "back" of the TKR ($CT = 1$)

SNR: signal-to-noise ratio; N.B. not supernova remnant!

ACKNOWLEDGMENTS

It's said it takes a village to raise a child. I don't know how many villages it takes to raise a grad student—in many ways still a child—but I've been lucky enough to have many.

My advisor, Toby Burnett, has made my course as a grad student remarkably straight and smooth. He allowed me the freedom to choose and pursue my own projects while guiding their direction and providing insight on the research at hand. He has gone above and beyond in supporting me, including allowing for travel to connect with LAT Collaboration members and other scientists in the field, an important part of joining the research community.

The science and performance of the Large Area Telescope is driven by the LAT Collaboration, and to them I owe much. The members are happy to share advice, expertise, and time, and I learned much about the process of science by having the opportunity to regularly present results to a group of interested experts. Within the collaboration, I owe thanks to a few individuals in particular:

- to David Smith, for having what turned out to be a rather important breakfast with me at Denny's in Cocoa Beach, Florida, where we were awaiting the launch of *Fermi*. He convinced me over greasy food that doing Galactic science with *Fermi* was far too promising and exciting to miss out on, and this set the course of my graduate career. Since then he has offered advice, friendship, and hospitality, and I am grateful for these and more.

- to Eric Grove and Paul Ray for their enthusiasm about and support of the development of *pointlike*.

DEDICATION

For my parents

Chapter 1

OVERVIEW

In Chapter 2, we present an introduction to γ-ray astronomy, including the physics of γ-ray interaction in matter and the resulting constraints on γ-ray telescopes. We review the major components of the *Fermi* Large Area Telescope (LAT) and briefly describe the implementation of similar components in past experiments. We conclude with an overview of some of the major sources of γ rays.

In the next chapter, we detail the challenges of analysis of $Fermi - LAT$ data. The primary consideration is the complex instrument response function (IRF), e.g. the power law dependence of angular resolution on energy. We suggest the use of likelihood techniques which incorporate the full IRF *ab initio*. We present the likelihood appropriate for photon-counting instruments and, by considering the IRF of the LAT in detail, develop a version suited to the LAT. We note that the expressions require multi-dimensional integrals and are generally computationally expensive.

In Chapter 4, we discuss the implementation of *pointlike*, a package for maximum likelihood spectral analysis of *Fermi*-LAT data. By adopting a binning scheme that scales with the detector resolution, we achieve considerable compression of the data at low energy. Moreover, we make controlled approximations allowing for accurate but rapid evaluation of the likelihood. We detail these approximations formally and show they adhere to a goal of 1% accuracy. By implementing a fast likelihood package, we open up new science through both interactive and large-scale analysis.

Chapter 5 gives an overview of the validation of *pointlike* at the top level. We verify that the code accurately reproduces Monte Carlo data from the model, i.e. that the approximations developed in Chapter 4 are sufficiently accurate. We then validate the most important functionality of *pointlike*, viz. estimating spectral parameters and parameter uncertainties,

by performing fits of ensembles of simulated sources.

To demonstrate the potential of *pointlike*, we present in Chapter 6 some actual *pointlike* analyses. We work through an "interactive" analysis of sources in the Cygnus region, conveying the importance of an exploratory approach in identifying new and necessary components of the source model. We develop the machinery for an all-sky analysis in which the spectra for all sources in the sky are determined consistently. The product of this analysis—an excellent model of the GeV sky—enables other large-scale analyses, and we detail three: the construction of "test statistic" maps for the detection of new sources, the generation of useful visual representations of the data via kernel density estimation, and the selection of unidentified LAT with pulsar-like properties. These pulsar-like sources have been targeted in radio pulsation searches, and we briefly present the results of two surveys.

We switch gears somewhat in Chapter 7. We review statistics used for pulsation searches (Z_m^2 and H_m) and argue that their use of only time-domain information only is inadequate. We modify the statistics to include a weight for each photon, and we use *pointlike* to calculate a weight giving the probability a photon originates from the source being tested. By leveraging the additional information in the photon energy and position via the spectral analysis, the statistic becomes appreciably more resilient to Type I and Type II error. We demonstrate the capabilities with an ensemble of Monte Carlo pulsars and show the weighted statistics improve the sensitivity by 50–100%.

Finally, in the appendices, we present a derivation of the asymptotic distribution of the H test[42]. This new result obviates the need for Monte Carlo calibration. We also provide an extension of the methods presented in Chapter 7 to sources with periods long compared to the time scale on which the LAT orientation changes, allowing sensitive searches for orbitally-modulated emission from binary systems.

Chapter 2

GAMMA-RAY ASTRONOMY AND THE *FERMI* LARGE AREA TELESCOPE

Gamma-ray astronomy is the study of light in the limit of very few, very energetic photons. The gamma-ray band extends from soft γ rays with energies of 100 keV—see e.g. the instruments aboard the INTEGRAL[85] observatory—up to 100s of TeV, the domain of ground-based imaging air Čerenkov telescopes, e.g., VERITAS[47], HESS[58], and MAGIC[46]. In between these two extremes, from 100 MeV to 100 GeV, lies the high-energy (HE) γ-ray band, which we shall also occasionally refer to as "GeV γ rays". To study this light is the purpose of the *Fermi* Large Area Telescope and this work.

To order of magnitude, a blackbody spectrum peaking at 1 MeV indicates a source temperature of 10^{10}K. Such temperatures occur in only the most extreme and ephemeral processes—e.g., core collapse supernovae[86]. In general, persistent HE γ rays are emitted through *non-thermal* processes—inverse Compton scattering, Bremsstrahlung[1], and π^0 decay[65]. In §2.3, we consider some of the sources hosting such processes and shining in the GeV.

Before considering emission, however, we begin with the most important facet of γ-ray (or any!) astronomy: reception. Earth's atmosphere is opaque to γ rays, so GeV telescopes are necessarily balloon-borne or space-based. HE source fluxes are generally low, and long integration times and reduced atmospheric background favor space telescopes (currently, AGILE[80] and the *Fermi* Large Area Telescope). In the following section, we give an overview of the relevant detector physics and resulting principles of operation of GeV telescopes.

---

[1]We include radiation induced by magnetic fields—synchrotron and curvature—in this category.

2.1 Gamma-ray Telescope Principles

Just as the phenomenology of low-frequency radiation—diffraction, refraction, e.g.—determines the design of radio and optical telescopes, so the physics of high-energy particle interactions shapes the design of γ-ray telescopes. Accordingly, we begin with an overview of the interactions of photons, particularly at high energy, in matter.

2.1.1 γ rays in Matter: Pair Production and Bremsstrahlung

At long wavelengths, light has a well-defined phase and interacts with matter as prescribed by classical electrodynamics[59]. At optical wavelengths, the energy associated with photons becomes comparable to atomic binding energies and the photoelectric effect is the dominant process. For heavier elements, the binding energy of core electrons reaches X-ray energies, allowing for continuing resonances in the photoelectric cross section, while for lighter elements Compton scattering beings to dominate the cross section at energies of a few keV. At 1 MeV (precisely, twice the electron mass), it becomes possible for a photon to produce an electron and a positron in the Coulomb field of a nucleus, the field absorbing the necessary four-momentum to ensure four-momentum conservation. Stronger Coulomb fields provide for a larger cross section with a more rapid onset at threshold. At sufficiently high energies (about 100 MeV for heavy elements), the pair-production cross section (a) dominates the total cross section (see Figure 2.1.1) and (b) is essentially flat (see Figure 2.1.1).

The flat cross section implies that the interaction rate is scale free, and there is effectively a mean free path for high energy photons[2]. This quantity depends only on the material, and is commonly called the *radiation length*[3], denoted X_0. In this same high-energy régime, the dominant interaction for electrons is Bremsstrahlung via interaction with nuclear Coulomb fields. (At the low-energy limit of this range, energy loss due to ionization of the material becomes important; this *critical energy*[87] is material-dependent but is roughly 10 MeV for heavy elements.) At first order, the energy loss rate for Bremsstrahlung is $dE/dx = -E/X_0$, i.e. electrons emit on average all but $1/e$ of their initial energy after traversing a distance

---

[2]Equivalently, the energy loss rate is linear in the incident energy.

[3]A radiation length is 7/9 of the photon mean free path[87].
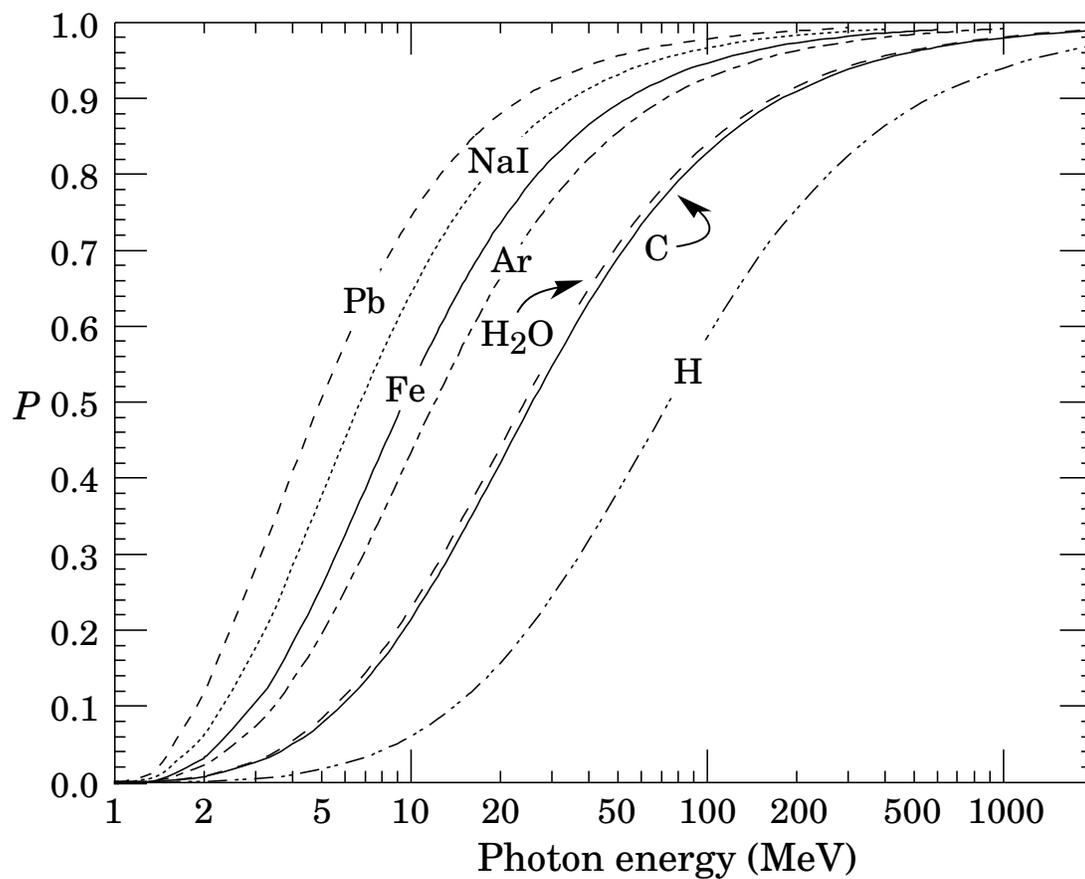
Figure 2.1: The probability that the first interaction of an incident photon in the given material is pair production. By 100 MeV, the cross section for photon interaction in heavy elements is dominated by pair production. Reproduced from [87].
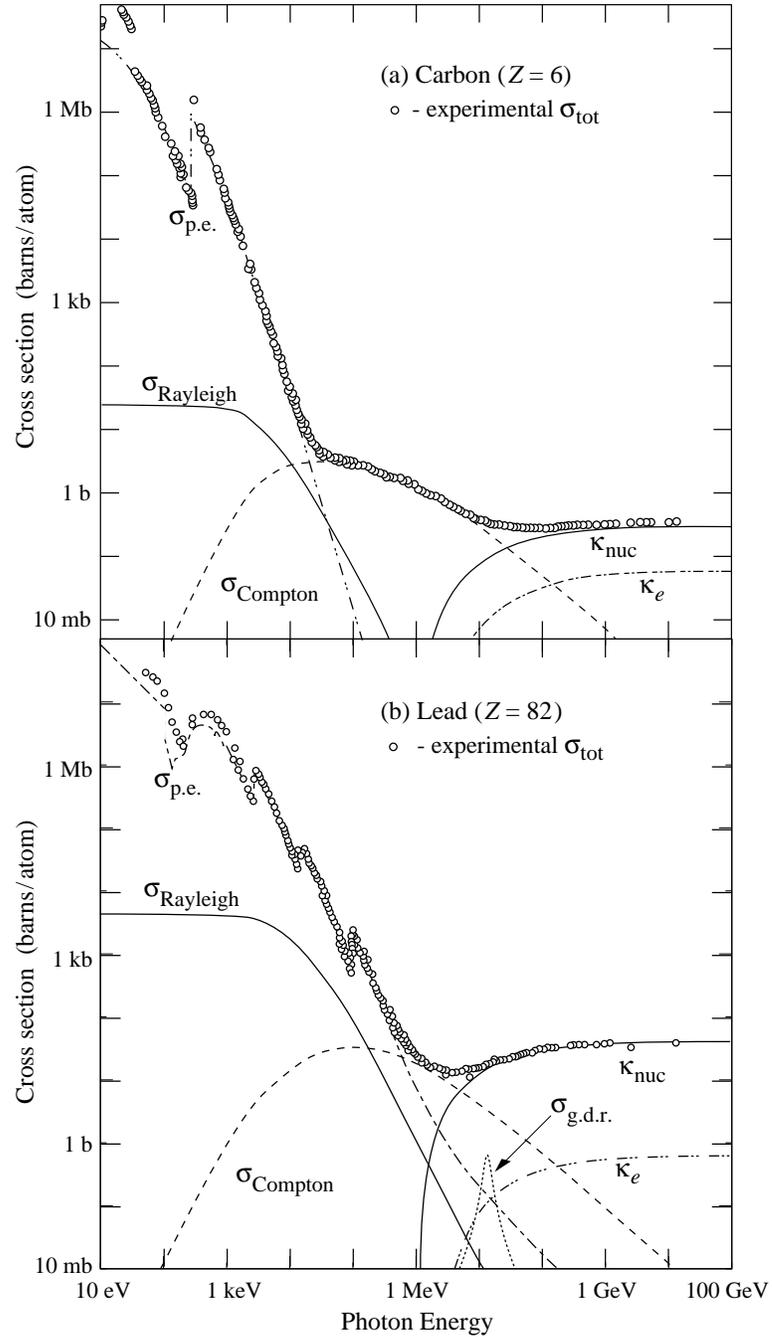
Figure 2.2: The photon cross section in carbon (Z=6) and lead (Z=82). The greater binding energy of atomic electrons in lead allows photoelectric resonances to X-ray energies, while Compton scattering is the dominant process in X-rays for carbon. In both materials, pair production on nuclei becomes dominant at high energies, where the cross section is essentially flat. The absolute value of the cross section is significantly higher for lead, as $\sigma \propto Z^2$. Reproduced from [87].

X_0.

These conditions lead naturally to an electromagnetic shower in which an incident photon produces an e^-/e^+ pair, each member of which in turn radiates high energy photons, leading to an exponentially-growing cascade of electrons and photons. As long as the next-generation particles remain well above the pair-production threshold (photons) or critical energy (electrons), the shower continues to develop exponentially. Since the produced e^-/e^+ are highly relativistic, the pair-production cross section in the lab frame is sharply peaked along the initial photon trajectory. The Bremsstrahlung is in turn beamed in a narrow cone about the electron trajectory. The resulting well-contained showers allow inference about the properties of the incident photon.

An electron traveling in matter is subject to Coulomb scattering from surrounding atoms. Unlike the significant momentum change from Bremsstrahlung spurred by close approach to a nuclear field, the interactions with distant charges are mere "love taps" perturbing the path of the electron. For a sufficiently thick piece of material but small overall perturbations, this multiple Coulomb scattering is like a random walk, and the particle exits the material with an approximately Gaussian distribution about its initial direction. The standard deviation of this distribution is given in the relativistic limit by[87]

$$\theta_0 \approx 8° \sqrt{X} \, \frac{100 \; MeV}{E}, \tag{2.1}$$

with X the length of material traversed measured in radiation lengths.

These basic physics considerations mean that a GeV telescope must be a full-fledged particle detector capable of tracking charged particles and measuring the energy of showers. In the next section, we consider the components a typical GeV telescope needs to accomplish these tasks.

2.1.2 Components of a GeV Telescope

The nature of interaction of GeV photons with matter sets strong constraints on telescope design. The *only* available detection process is conversion of the incident photon into a pair, i.e., there is no focusing of γ rays. We have the following results:

- converting the bulk of incident photons requires sufficient radiation lengths of material;

- maximizing the pair-production cross section (relative to, e.g., Compton scattering) and the electron critical energy requires a high-Z converter;

- obtaining a good estimate of the photon trajectory requires measurement of the position of the first generation pair before it has suffered significant multiple scattering;

- obtaining a precise measurement of the energy requires converting, containing, and measuring the bulk of the electromagnetic shower, i.e., many radiation lengths for conversion and additional material to measuring ionizing charged particles and radiation.

These constraints essentially design γ-ray telescopes for us, and nearly all successful GeV telescopes have used some form of the following components:

- A combination tracker/converter module. To minimize multiple scattering and premature *Bremsstrahlung*, thin, high-Z metal converter foils are interwoven with active material capable of tracking charged particles. To increase the fraction of converted photons, multiple conversion foils and tracking layers may be stacked.

- A calorimeter. After having obtained position information, it is desirable to contain and measure as much of the shower as possible. The calorimeter should bring high energy particles out of the radiation range (i.e., be many radiation lengths deep) and then measure the energy of the ionizing daughter particles.

Additional components require consideration of the telescope environment. The charged particle number flux—from primary cosmic rays and secondary cosmic rays trapped in the earth's magnetosphere—is about from 10^3 to 10^5 that of the HE γ-ray flux, and this considerable background must be rejected. Second, the earth is a bright source of γ rays (see below and Chatper 4), and it is desirable to reject γ rays entering the instrument from below. Thus, we require two more components:

- An anticoincidence detector that signals the passage of a charged particle.

- A mechanism to reject upward-going (zenith bound) γ rays.

Having now identified the guiding physical principles and the major components of GeV telescopes, we consider the implementation of the most sensitive GeV telescope to date, the *Fermi* Large Area Telescope.

2.2   The Fermi *Large Area Telescope*

As we observed in the last section, a GeV telescope is naturally modular, and accordingly we organize this section around the components of the *Fermi* Large Area Telescope (LAT). To provide context and to acknowledge the significant scientific achievement of previous GeV experiments, we intersperse discussion of the LAT components with remarks on the implementation adopted in foregoing telescopes and the discoveries made possible by the technical advances.

2.2.1   The Tracker

The tracker/converter (TKR[21]) is the heart of the LAT. Enabled by advances in solid state particle tracking technology, it provides the sharpest view to date of the GeV sky. To balance the tension between efficiency (governed by the total radiation lengths of converter material) and angular resolution (governed by multiple scattering in the converter material), the TKR is composed of 16 alternating layers of tungsten (Z=74) and tracking components[4]. The first 12 foils—referred to as the "front" of the TKR—are $0.028X_0$, while the final 4 foils—the "back"—are $0.18X_0$. Eq. 2.1 thus indicates that photons converting in the back layers suffer approximately twice the angular deviation of those converting in the front. The total number of converted photons are roughly equally apportioned between the front and the back by the ratio of X_0 in the two TKR sections.

Following each tungsten layer is a silicon strip detector (SSD) with a pitch of 230 microns. The SSD comprises two planes with orthogonal strip orientations, allowing for a

---

[4]An additional two layers of tracking planes without conversion foils lie at the bottom of the TKR, for a total of 18 tracking layers.

measurement of the transverse (to the detector plane) position of a charged particle. The fine pitch and numerous layers require nearly one million readout channels[22].

The TKR is not monolithic. It is segmented into a 4x4 array of 16 identical modules, each in turn made up of the 16 (18) conversion (tracking) layers described above.

Although using modern technology, the LAT TKR is also evolutionary. The first γ-ray satellite telescope, OSO-3, had no tracking capabilities and instead relied on a rough collimation based on the anticoincidence veto, giving it a $\approx 20°$ angular resolution. Even this rough discrimination was sufficient to detect diffuse emission from the Milky Way[63] (see below), but it was not until spark chambers were incorporated in SAS-2[49] and COS-B[27] that sufficient resolution and sensitivity to detect point sources was acquired. The spark chamber reached its zenith in EGRET[62, 83], where the 28 layers of tantalum-interleaved spark chambers (with 0.8mm wire spacing) achieved a high-energy resolution of 0.4°.

2.2.2 The Calorimeter

To measure the energy from an incident photon, the LAT is equipped with a hodoscopic calorimeter (CAL[60]). As with the tracker, it is segmented into 16 autonomous modules. (Together, the tracking and calorimetry elements and their electronics form a "tower".) Each of these modules is formed from 96 $2.0 \times 2.7 \times 33.6$ cm cesium iodide (CsI) cyrstal scintillator bars arranged in an 8 (depth, along instrument axis) by 12 (breadth) array with the long dimension in the same plane as the TKR layers. The orientation of the long dimension shifts by 90° in alternating layers.

The crystals serve both to continue to seed pair production and Brehmsstrahlung and, once the electromagnetic cascade has reached its maximum (particles are at the critical energy/ pair production threshold), the crystals scintillate as the main energy loss mechanism becomes ionization the CsI atoms. Two photodiodes attached to each end of the crystal's long dimension allow for readout and a measurement of the transverse position of energy deposition by light asymmetry.

The total CAL is $8.4X_0$ on-axis[5], which depth becomes insufficient to fully contain show-

---

[5]The CAL contains about 1800kg of CsI, constrained by the payload limit of the *Fermi* launch vehicle.

ers initiated at high energies. The segmentation allows characterization of the longitudinal profile of the shower and hence an estimate for the fraction of energy that escapes the CAL. The segmentation also allows an accurate measurement of the centroid of energy deposition which serves as an anchor point in track reconstruction algorithm[22].

While the depth of the CAL is similar to the calorimeter flown on EGRET, the greater transverse area of the CAL enables the wide field-of-view of the LAT. Beyond facilitating track reconstruction and energy measurement, the imaging capabilities of the CAL also obviate the need for a dedicated time-of-flight coincidence system for rejecting upward-going γ rays; EGRET, SAS-II, COS-B, and OSO-III all made use of such a system.

2.2.3   The Anticoincidence Detector

The charged particle background flux can exceed that of the desired γ-ray signal by up to 10^5[68]. While the TKR and CAL can reject hadronic showers with relatively high efficiency, cosmic ray electrons generate cascades very similar to those initiated by photons. These too, of course, can be vetoed at the cost of efficiency by requiring "empty" TKR planes above the conversion point of the photon. However, in the face of complex analysis and extremely high backgrounds, it makes much more sense to attempt to detect charged particles independently. This function is provided by the Anticoincidence Detector (ACD), a plastic scintillator-based detector covering the top and sides of the LAT.

The ACD comprises 89 scintillator tiles, each with independent and redundant readout. Spaces between tiles are filled with ribbon scintillators. By segmenting the ACD, self-veto from "backsplash"[6] is significantly decreased[68]. The segmented ACD of the LAT is a significant improvement on those of previous experiments, which were largely monolithic and thus suffered from appreciable self-veto. The sensitivity of EGRET, e.g., was essentially zero above 50 GeV[68].

---

[6] "Backsplash" denotes the phenomenon of soft X-rays produced in the electromagnetic cascade of a high-energy (above $\approx$ 10GeV) photon traveling back through the TKR and into the ACD. Some of these X-rays will Compton scatter, causing an ACD trigger. A segmented calorimeter allows for a fine-grained veto, as veto can be restricted to only events with a track extending to a triggered ACD tile.

2.2.4 Orbital Environment and Operating Modes

The LAT was launched on June 11, 2008 from Cape Canaveral into an orbit with an altitude of ≈ 565 km inclined $\approx 26°$ from the equator. The orbit precesses with a period of about 53 days. The S/C typically operates in a "sky survey" mode in which with each orbit it rocks away from the zenith by $\pm D°$, alternately viewing the northern and southern orbital hemispheres. The initial rock angle was $D = 35°$ but has since increased to $D = 50°$ to reduce the operating temperature of the S/C battery.

If the onboard software of the LAT or the GBM (Gamma-ray Burst Monitor) detects a signal consistent with a gamma-ray burst, the S/C can execute an Autonomous Repoint Request in which the sky survey mode is abandoned and the S/C maintains an attitude to keep the putative burst site near the center of the field-of-view[7]. Since such pointing strategies allow the S/C axis to approach the limb of the earth, these periods are typically excised in standard analysis.

As noted in the previous section, the charged particle background for the LAT is considerable. The typical trigger rate is ≈ 2.2kHz, but can increase by about a factor of 2 during some orbits as the earth's magnetosphere rotates in the plane of the S/C orbit[8]. Reading out the LAT takes about 26 μs, leading to deadtime of order 5%. On the majority of orbits, the S/C encounters the "South Atlantic Anomaly", or SAA. Within this region, the particle background is significantly higher than typical for the LAT's orbit, making nominal observations impossible. When the S/C enters the SAA, the LAT ceases to trigger, and the time spent traversing the region is lost to normal science operations. This episodic deadtime decreases the LAT duty cycle by about 10% and leads to a modest north-south asymmetry in the exposure.

---

[7]As with many telescopes, viewing a source directly on axis is a poor stategy since the "cracks" in the detector are then aligned with incident photon trajectories.

[8]Background leakage is then also time-dependent on < 1d timescales; care is required when searching for periodic signal from short-period binaries!

2.3   Sources of Gamma Rays

To complete the overview of γ-ray astronomy and *Fermi*, we take a brief look at the classes of γ-ray sources filling the high-energy heavens.

2.3.1   Diffuse Emission from Cosmic Rays

Cosmic rays produce GeV radiation via three primary mechanisms[18]. At energies below and to the edge of the LAT passband, the emission is dominated by electronic processes, particularly Bremsstrahlung (induced by gas) and inverse Compton (IC) scattering of ambient light, e.g. the cosmic microwave background. Coincident with the low end of the LAT passband, the contribution of hadronic processes turns on rapidly. The inelastic scattering of high energy protons in gas nuclei produces π^+, π^0, and π^- particles in roughly equal numbers, and the dominant decay process for π^0 (98.8% branching ratio[87]) is to two photons. The γ spectrum depends on the π^0 spectrum, which in turn depends on the cosmic ray proton spectrum and the cross section for inelastic nucleon scattering. This cross section turns on rapidly at the threshold for π^0 production[51, 61], leading to the characteristic rapid rise of the observed diffuse γ spectrum in the LAT band[6].

At the high end of the LAT passband (> 100 GeV), electronic processes can again become important via IC scattering of the interstellar radiation field.

In all of these processes, the γ-ray production depends on both the cosmic ray energy density and the "target" density—gas for Bremsstrahlung and radiation for IC scattering, leading to significant variations of the appearance of the HE sky as a function of energy, a topic we take up in a brief discussion of emission in our own Milky Way.

The Milky Way

As can be appreciated from the representation of the γ-ray sky in Figure 6.3, the dominant feature is diffuse emission associated with the structure of the Milky Way. The intensity of the low latitude emission is a projection effect as the emissivity scales as the product of the cosmic ray density and the target density. This sharp peak in intensity[14] leads to difficulty in the analysis of point sources in the plane[12]. Emission from higher latitude

represents cosmic ray interactions in our local environment[6].

Starburst Galaxies

If supernovae shocks provide the accelerators for cosmic rays, starburst galaxies[64], with their intense patches of high star-formation rates, should be modestly bright γ-ray sources, as the high supernova rate and enhanced dust concentration (target material) lead to significantly elevated γ-ray production relative to, e.g., the Milky Way. The detection by *Fermi*-LAT of two starburst Galaxies[64] confirms this idea, providing both a new opportunity for the study of cosmic ray production and insight into the observed extragalactic γ-ray background.

The Earth

As with much other emission from the Solar System (e.g., the moon and the quiescent sun[30], emission of γ rays from Earth has its origin in cosmic rays[4]. Due to its proximity, the limb of the earth is in fact the brightest source of γ rays seen by Fermi, requiring care in separating its signal from that of celestial sources, see e.g. Figure 4.1.2 and discussion in §4.1.2.

2.3.2 Active Galactic Nuclei

The center of most galaxies are known to harbor massive $\geq 10^6 M_{sun}$ black holes. While most are quiescent (e.g. our own Sgr A*), accretion at $\approx 10\%$ of the Eddington rate onto these black holes can fuel powerful relativistic, collimated outflows known as jets. These jets are often observed to have apparent superluminal velocities (e.g. [29]) implying the jet material has Lorentz factors of order 10 and the motion is largely along the line of sight. Such motion leads to a Doppler boost in both luminosity and apparent energy, making emission visible from cosmological distances. GeV emission is produced through IC scattering of seed photons, which may be produced by synchtron emission by the electrons in the magnetic field of the jet or from sources external to the jet, e.g., the bright accretion disk[77]. AGN bright in the GeV are known as blazars and represent the class of AGN

with strong jets aligned very closely with the line of sight. The class breaks down into flat-spectrum radio quasars and BL Lac objects, and Fermi sees roughly equal numbers of each[16]. An appreciable component of the observed isotropic diffuse emission[15] is likely contributed by unresolved blazars.

2.3.3 Pulsars

Pulsars are rapidly-rotating, highly-magnetized (10^8 to 10^{14} gauss) neutron stars. The intense currents induced lead to radiation emission over a broad energy range. Although pulsars have a storied history in radio[57], *Fermi*-LAT has essentially opened up the study of pulsars in GeV, e.g., with the detection of millisecond pulsars[2], the detection of tens of potentially radio-quiet pulsars[3, 75], and thus we concentrate on their high-energy emission.

In canonical emission models, the acceleration of charged particles to very high energies and subsequent production of γ rays through synchrotron/curvature radiation or inverse Compton scattering relies on the "gap" paradigm. A relative vacuum—a gap—develops in a magnetosphere otherwise filled with charged particles configured such that $\mathbf{E} \cdot \mathbf{B} \approx 0$[52]. The unshielded potential in the gap is suffcient to accelerate particles to ultrarelativistic energies, powering pair cascades which produce the observed GeV emission and allow current to flow.

Emission models are primarily distinguished by the site of the gap (Figure 2.3.3). Polar cap models[38, 39] place the gap immediately above the polar cap, the small patch on the neutron star surface defined by the footprint of the open magnetic field lines. (A field line is open if it crosses the light cylinder (LC), the surface at which a corotating particle would attain light speed.) Outer gap models[34, 36, 74, 73] invoke a gap at higher altitude, along the last closed field lines. The gap extends from the null charge surface—where the corotating charge density $\propto \mathbf{\Omega} \cdot \mathbf{B} = 0$ changes sign[52]—to the LC. Slot gap models[56] postulate narrow gaps encircling the PC and rising along the last closed field lines to the outer magnetosphere. The two-pole caustic model[45] is essentially a slot gap model with emission from a vanishingly thin surface along the last closed field lines. The annular gap model of [23] illuminates field lines along a thin annulus interior to the last closed lines.
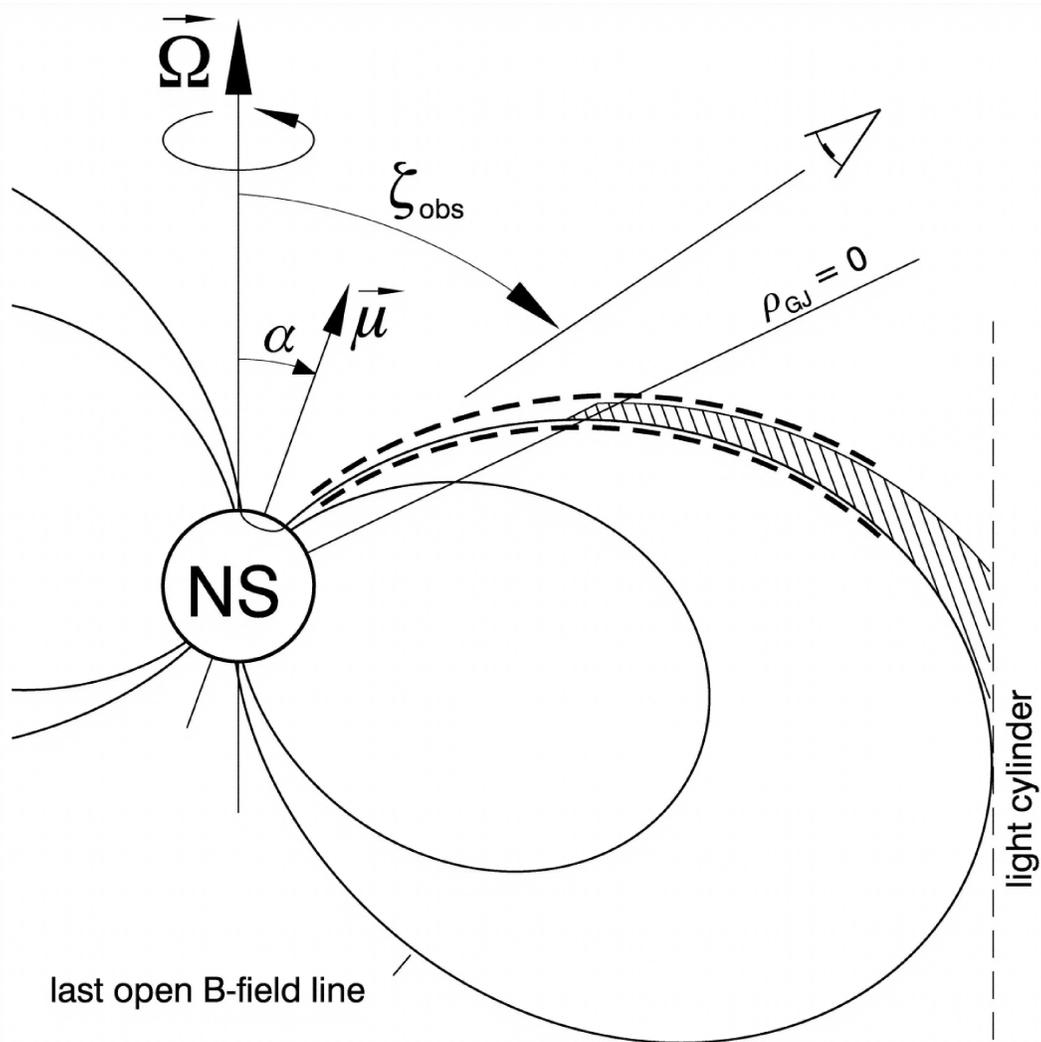
Figure 2.3: The canonical representation of a pulsar magnetosphere, reproduced from [45]. The magnetic moment of the neutron star (NS), μ, is inclined to the NS spin axis, $\mathbf{\Omega}$, by an angle α. The field is assumed to be dominantly dipolar, though higher multipoles may be important[88]. The Goldreich-Julian equilibrium charge density[52] ρ_{GJ} vanishes along the null charge surface. The small patch on the NS surface at the foot of the open $\mathbf{B}$ field lines is the polar cap. The slot gap (dashed lines) and outer gap (hatching) lie along the last closed (open) field line. HE emission is beamed along $\mathbf{B}$ field lines, and the observer's inclination ζ determines which field lines are visible, strongly affecting the resulting light curve.

Relativistic effects, such as aberration, become important in outer-magnetospheric emission and naturally furnish the characteristic cusped light curves of many HE pulsars[34, 5].

A second important consideration is the assumed structure of the magnetic field. Most calculations employ the solution of Deutsch[43], which posits an infinitely-conductive star and a vacuum exterior. At the other extreme are force-free magnetospheres in which the star exterior is modeled as a highly-conductive fluid and numerical solutions are derived with magnetohydrodynamic codes[79]. The true solution must combine elements of both. The two approaches differ most in the outer magnetosphere, where the co-rotation velocity becomes relativistic and the modification of field lines by plasma becomes important. Since LAT results point to a primary emission site in the outer magnetosphere[5, 2], this distinction becomes even more important. In both models, the field geometry depends strongly on α, the inclination of the dipole axis to the spin axis (Figure 2.3.3).

2.4   Summary

The myriad sources of γ rays—which we have only touched on here—offer opportunities to study some of the most extreme processes in the universe. However, the physics of γ-ray interactions require non-traditional telescopes, viz. particle detectors. By taking advantage of advances in solid state technology, the *Fermi*-LAT offers unprecedented angular resolution and sensitivity, particularly at energies > 10 GeV via suppression of self-veto. However, as we shall see in the following sections, the relevant physics still set fundamental limits on the capabilities of the detector. In particular, multiple Coulomb scattering leads to an angular resolution that varies by more than two orders of magnitude over the LAT passband. Dealing with this multi-scale response requires care, and in the next chapter we introduce the principle of maximum likelihood, leading to techniques optimally suited for analysis of *Fermi*-LAT data.

Chapter 3

LIKELIHOOD ANALYSIS FOR HIGH-ENERGY GAMMA-RAY TELESCOPES

Analysis of data from *Fermi*-LAT and other HE gamma-ray telescopes (e.g., EGRET[66]) requires techniques not normally found in analyses of data from other wavebands. In optical astronomy, for instance, observations are typically made of a narrow range of wavelengths, and the angular resolution of the instrument is sufficient to resolve the astrophysical background. To determine the flux from a star, e.g., one identifies the pixels in the CCD illuminated by the image of the star (aperture photometry), measures the collected charge, and divides by the telescope aperture, time of observation, and efficiency of the filter and CCD. In practice, of course, these steps may be quite involved, but the point remains that *physical* quantities, e.g., source intensities, can generally be extracted directly from the data. In addition, time-dependent instrumental effects are typically small, as sources can be tracked with great precision, resulting in a stable instrument response over the integration.

Such methods fail for *Fermi*-LAT data for two reasons. First, very few *Fermi*-LAT sources appear isolated. Looking at Figure 3, we note the absolute scale of the PSF at energies of order 0.1 GeV is about 5°, and an immediate consequence is source confusion. Suppose we want to extract the flux between 0.1 and 0.3 GeV for a particular source. Except for the brightest blazars at high Galactic latitudes, there is generally no circle we can draw that is dominated by emission from the source of interest. The situation is dire in the Galactic plane, as the projected density of point sources increases dramatically along with the diffuse background from cosmic rays. At higher energies, the PSF scale drops to $< 1°$ and the situation improves. By carefully calculating the integral of the PSF over an extraction radius, one could reliably extract fluxes for sources at high Galactic latitude by photon counting. Sources in the Galactic plane, however, still suffer strong contamination. In general, then, any technique that relies on simply apportioning photons to a source will

deliver poor results.

Second, detecting and characterizing GeV sources may require integrations of a year or more. The orientation of the instrument with respect to the source is constantly changing as the S/C orbits the earth, so there is no single "number" characterizing the instrument characteristic during the observation. In essence, we are using a different telescope every few minutes. Even if we could estimate accurately the number of photons we have observed from a particular source, we cannot straightforwardly convert it to a flux.

These considerations and others discussed below motivate likelihood techniques for the analysis of *Fermi*-LAT data. Such techniques work in the "forward-folded" space, i.e., not with physical quantities but with instrumental quantities, and they naturally incorporate the complexity of the energy-dependent PSF and time-dependent observation. Below, we introduce likelihood techniques in general and then specialize to GeV instruments and the LAT.

3.1 Characterizing and Modeling Sources

We characterize a source by its photon flux density. (In the following, we neglect polarization, since it is very difficult for a pair-conversion telescope to measure the polarization of an incident beam.) The particle flux density—the rate of photons incident per unit energy/area/time from a solid angle $d\Omega$ about the position $\vec{\Omega}$—completely encapsulates the properties of a source and is denoted $\mathcal{F}(E, t, \vec{\Omega})$. Its arguments are the observables associated with a photon, namely energy (E), time of arrival (t), and direction of origin ($\vec{\Omega}$), i.e., $-c\vec{p}/E$, with $\vec{p}$ the photon momentum vector. Although the distinction is seldom of interest, t is understood to be the retarded time.

Unlike infrared, optical, or even X-ray sources, the mechanisms that power HE gamma-ray sources almost universally produce broadband spectra, that is, spectra spanning multiple decades of energy and possessing no features and modest curvature. Such spectra can be modeled with only a few parameters, and often a simple power law suffices:

$$\mathcal{F}(E, t, \vec{\Omega}; N, \Gamma, E_o) = N(t, \vec{\Omega}) \left(\frac{E}{E_0} \right)^{-\Gamma(t, \vec{\Omega})} f(\vec{\Omega}). \qquad (3.1)$$
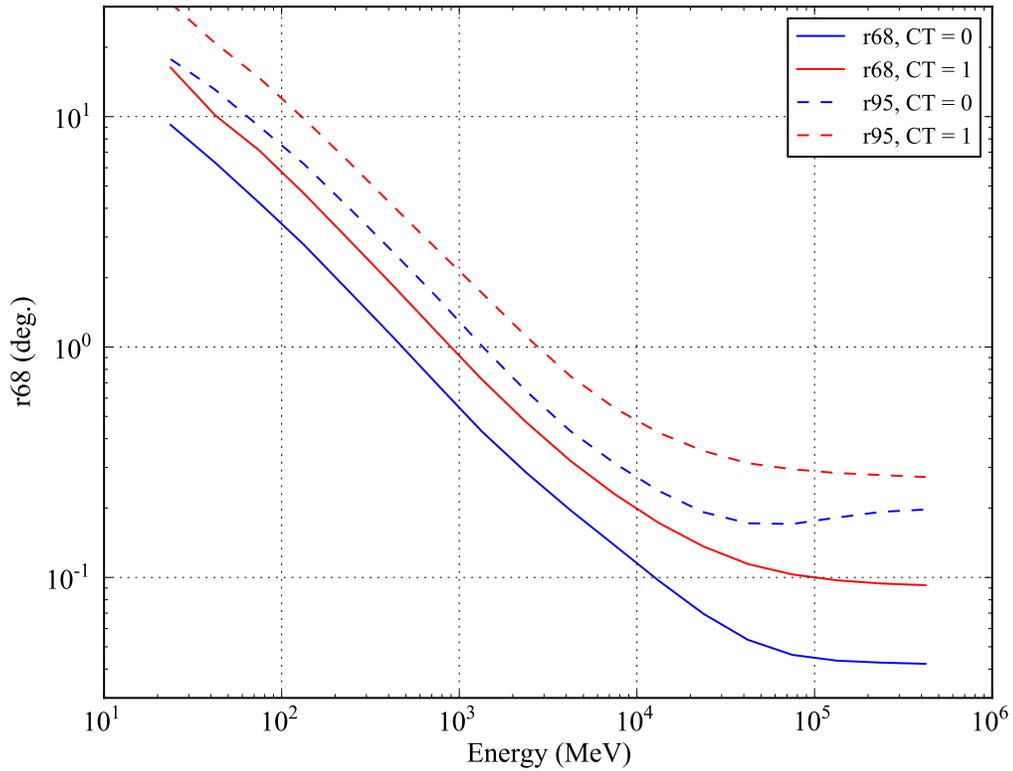
Figure 3.1: The *Fermi*-LAT point-spread function as characterized by "r68" and "r95", the angular separation from a point source within which 68% and 95% of photons from the source lie. The description here is for photons arriving close to the instrument's boresight; the angular resolution drops by up to 50% at the edge of the field of view (see Figure 4.1.2). As discussed in Chapter 2, the energy-dependence originates from multiple scattering of pair-produced electrons and positrons in the tungsten foils, and the asymptote is set by the pitch of the silicon strips in the tracking layers.

A simple extension, the exponential suppression of the flux above a cutoff energy,

$$\mathcal{F}(E, t, \vec{\Omega}; N, \Gamma, E_c, E_o) = N(t, \vec{\Omega}) \left(\frac{E}{E_0} \right)^{-\Gamma(t, \vec{\Omega})} \exp \left(-\frac{E}{E_c(t, \vec{\Omega})} \right) f(\vec{\Omega}), \qquad (3.2)$$

is adequate to characterize the emission from most γ-ray pulsars[5]. Here, $f(\vec{\Omega})$ is a normalized function ($\int d\Omega \, f(\vec{\Omega}) = 1$) describing the spatial morphology of the source. We are mostly interested in point sources, i.e., those that cannot be spatially resolved and for which the spatial dependence is given by the Dirac delta function, $f(\vec{\Omega}) = \delta(\vec{\Omega} - \vec{\Omega}_0)$, where $\vec{\Omega}_0$ is the true position of the point source. In general, we will not specify a particular form of spectral model and instead label it with a set of parameters, $\vec{\lambda}$, and express the modeled flux density for a source as $\mathcal{F}(E, t, \vec{\Omega}; \vec{\lambda})$.

3.2 Principles of Maximum Likelihood

With a model for the physical properties of sources in hand, the next step is to connect them to data and estimate the model parameters. Below we outline the method of maximum likelihood (ML) parameter estimation which shall be our primary tool for this task.

Suppose an experiment is designed to measure a random vector $\vec{X}$, that the random variables admit a probability density function (pdf), and that the pdf is parameterized by some vector $\vec{\lambda}$. We express the pdf as $f_{\vec{X}}(\vec{x}; \vec{\lambda})^1$, meaning that the probability to measure a value near $\vec{x}$ is just $f_{\vec{X}}(\vec{x}; \vec{\lambda}) \, d\vec{x}$. Viewed this way—being given $\vec{\lambda}$—the probability is a function of $\vec{x}$, the possible outcomes of our experiment. On the other hand, having performed the experiment and measured $\vec{x}$, we can view the same expression, $f_{\vec{X}}(\vec{x}; \vec{\lambda})$, as function of $\vec{\lambda}$, i.e., it characterizes the parameter values *most likely* to have yielded the measured values. When viewed as a function of the parameters, this quantity is called the *likelihood* and we denote it $\mathcal{L}(\vec{\lambda}; \vec{x})$, although we often suppress the explicit dependence on $\vec{x}$.

The likelihood is useful for the common statistical/analysis task of determining estimators for the parameters of a distribution from data: natural estimators are those which maximize the likelihood, known as the *maximum likelihood estimators*, or MLE. In many cases, the MLE are *consistent*, have an asymptotically normal distribution, and are efficient.

---

[1]The notation for pdfs is discussed more fully in Chapter 7.

That is, with enough data, the estimators converge to their true values, the parameter uncertainty can be estimated from the shape of the likelihood, and the uncertainties are the smallest possible.

For large data sets, the likelihood is likely to be a very small number. Since the logarithm of monotonic functions is again monotonic, we can instead maximize the "log likelihood", which is also often easier to compute.

We note in passing that, via Bayes' Theorem, the likelihood can be converted to the posterior probability by multiplication with the prior probability. That is, since $f(\vec{x}; \vec{\lambda})f(\vec{\lambda}) = f(\vec{\lambda}; \vec{x})f(\vec{x})$, the posterior probability density, $f(\vec{\lambda}; \vec{x})$, is proportional to $\mathcal{L}(\vec{\lambda}; \vec{x})f(\vec{\lambda})$. Here, $f(\vec{\lambda})$ is the *prior probability*, or the distribution for the parameters we would expect in the absence or any data, or in light of previous experiments. This interpretation allows one to interpret the statistical errors on the estimators as uncertainties on the true values, e.g., to construct credible intervals which have a 90% chance of containing the true value of the parameter.

3.3 The Observed Signal

As outlined above, likelihood provides a connection between model parameters (as close to physical measurements as we can hope to get given the experimental constraints) and data. Implicit (and perhaps hidden) in the simple notation we employed is the mapping of the source models onto the actual data. That is, if we want to calculate the probability to observe some set of data (in the case of GeV instruments, some collection of photons) given some model, we must first figure out what that model looks like to the detector. The first step is understanding how a signal from a source is processed by the instrument.

A GeV telescope is essentially a particle detector, and the fundamental data are events. The detector, of course, is imperfect: it (a) fails to generate observables for all incident photons, and (b) introduces error into the observables for successfully detected photons. To illustrate, suppose some number of photons from a steady, monoenergetic point source impinge upon the detector. For the LAT, the tungsten foils in the tracker provide about

$X_0 = 1.1$ radiation lengths[22], and $\approx 40\%$ of photons at normal incidence[2] will pass through the TKR without interacting[21]. These photons cannot be successfully reconstructed[3]. Of the photons that interact in the TKR, some will fail to produce an identifiable signal[4]. And of course, for those that interact *and* are successfully recognized as interacting photons, we will see a spread in the reconstructed energies and position.

These effects, both the *efficiency* for successful reconstruction of an incident photon and the *dispersion* of its true observables are characterized by the *instrument response function*, or IRF. In keeping with its two rôles, we explicitly factor the IRF into two pieces. The instantaneous sensitivity, with units of area and denoted $\epsilon(E, t, \vec{\Omega})$, is the quantiy by which the flux density $\mathcal{F}$ should be multiplied to determine the instantaneous count rate in bin of width $dE \times dt \times d\Omega$. The dispersion matrix, $P(E', t', \vec{\Omega}'|E, t, \vec{\Omega})$, gives the probability that a photon with true energy E, time of arrival t, and position $\vec{\Omega}$ will have a reconstructed energy E', time of arrival t', and position $\vec{\Omega}'$. The dispersion matrix is a true probability density function and is normalized such that $\iiint dE\, d\Omega\, dt\, P(E', t', \vec{\Omega}'|E, t, \vec{\Omega}) = 1$. The expected event rate in an infinitesimal bin centered on E', t', and $\vec{\Omega}'$ is then

$$r(E', \vec{\Omega}', t'; \vec{\lambda}) = \iiint dE\, d\Omega\, dt\, \mathcal{F}(E, \vec{\Omega}, t; \vec{\lambda})\, \epsilon(E, \vec{\Omega}, t)\, P(E', \vec{\Omega}', t'; E, \vec{\Omega}, t). \qquad (3.3)$$

The integral is over the support of $\mathcal{F}(E, \vec{\Omega}, t; \vec{\lambda})$.

This quantity—the source flux as processed by the instrument—is what we need to calculate the likelihood. Before we proceed, we first describe in some detail the *Fermi*-LAT IRF and make some simplifications to Eq. 3.3.

3.4   *The Instrument Response Function of the Fermi-LAT*

The calculation of the terms $\epsilon(E, \vec{\Omega}, t)$ and $P(E', \vec{\Omega}', t'; E, \vec{\Omega}, t)$ above requires some care for space-based gamma-ray telescopes such as the Fermi-LAT. First, we note that the response

---

[2]The apparent depth of a tungsten foil is inversely proportional to the cosine of the incidence angle.

[3]In principle, events can be reconstructed using the hodoscopic calorimeter only, but this technique is not yet part of standard *Fermi*-LAT analysis.

[4]The cross section for pair production in tungsten is essentially flat in energy from a few hundred MeV to 100s of GeV[87], so the increase in effective area (Figure 3.4) between 100 MeV and 1 GeV is primarily due to increasing performance of the event reconstruction algorithms.

of the LAT to an incident photon depends primarily on the geometry, viz. the angle between the photon's momentum vector and the instrument's boresight, which by convention defines the z-axis in the S/C frame, $\hat{z}$. While this dependence is both on the polar angle, $\vec{\Omega} \cdot \hat{z}(t) \equiv \cos\theta(t)$ and the azimuth, for events not too far from the boresight, the azimuthal dependence is subsidiary to the polar dependence. Furthermore, over long integrations[5], the azimuthal dependence tends to average out (see, e.g., Figure 3.4). Thus, we express the geometry solely in terms of the polar angle, which depends on the position and orientation of the S/C as it orbits.

We consider first the term $\epsilon(E, \vec{\Omega}, t)$, which is essentially the effective cross section of the detector at time t. (As noted above, not all events can be reconstructed, so the effective area is always less than the geometric cross section.) The term factors into the effective area, $A[E, \cos\theta(t)]$, with units of area, and a dimensionless efficiency, which we denote $e(E, t)$, a number between 0 and 1. The effective area accounts for the ability of the instrument and software to successfully reconstruct a photon with energy E incident at an angle θ. It gives *to first order* the true collecting area of the detector. We show the effective area for a particular version of the reconstruction algorithms, *P6_v3_diff*, in Figure 3.4. The efficiency comprises two terms. First, an energy-independent term depends on the trigger rate, as the instrument has a period of deadtime following each trigger while the electronics are read out. Second, there is an energy-dependent effect resulting from "pileup". For about 10% of triggers, an additional charged background particle is interacting in the detector, degrading the reconstruction of the triggering particle and decreasing the efficiency. The effect decreases with incident photon energy. Since the trigger rate is dominated by charged particles, both of these effects depend on geomagnetic latitude and hence on time.

Next, we consider the term $P(E', \vec{\Omega}', t'; E, \vec{\Omega}, t)$. We make the simplifying approximation that the three observables are statistically independent and write $P(E', \vec{\Omega}', t'; E, \vec{\Omega}, t) = f_{\text{psf}}[\vec{\Omega}'; \vec{\Omega}, \cos\theta(t), E] \times f_{\text{edisp}}[E'; \cos\theta(t), E] \times f_{\text{tdisp}}(t'; E, \vec{\Omega}, t)$, with f_{psf} the point-spread function, f_{edisp} the energy dispersion function, and f_{tdisp} the time dispersion. This assumption is certainly justified for t'. The position (energy) reconstruction is dominated by the

---

[5]For transients such as gamma-ray bursts, only a short selection of data is used, and this approximation clearly fails.
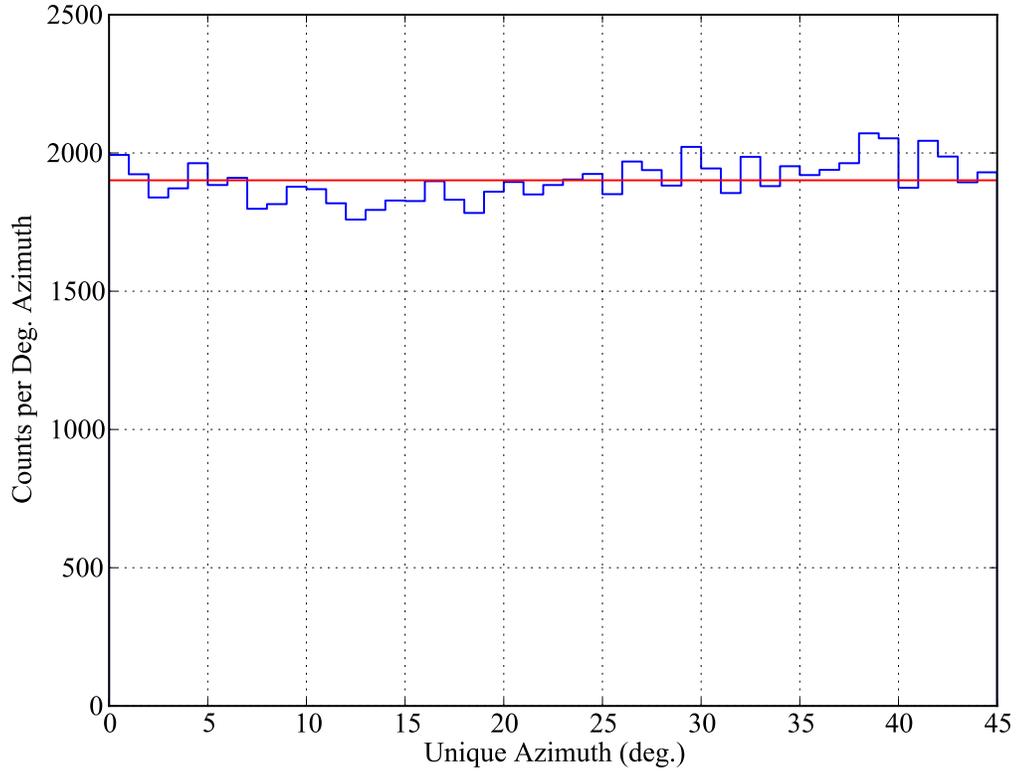
Figure 3.2: The distribution in azimuth for a set of reconstructed events (P6 "diffuse" class) with 100 MeV < E < 100 GeV in a 10° aperture centered on (l,b) = (0,-45), integrated from Aug. 4 2008 to Jul. 18 2010. Events with a reconstructed zenith angle > 105° have been discarded. (See Figure 4.1.2 which charts the zenith angle for this same data.) The *Fermi*-LAT has twofold bilateral symmetry, so we can obtain the "unique" azimuth angles, i.e., those that truly probe the dependence on the shape of the detector, by $\phi_{unique} \equiv \phi \bmod 90°$ and, further, reflecting events with $\phi_{unique} > 45°$ to $90° - 45°$. The red horizontal line indicates the mean, while the blue gives the reconstructed unique azimuth in 1° bins. The deviation from the mean is quite modest.
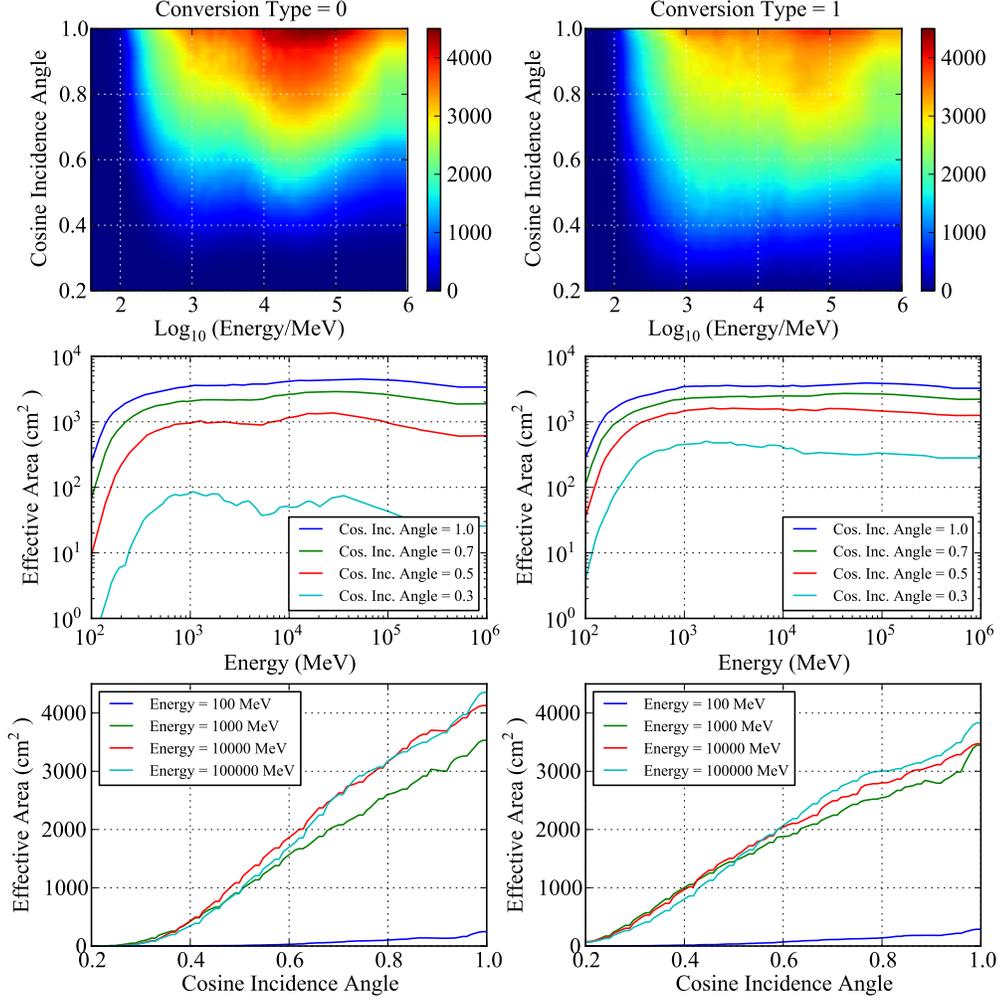
Figure 3.3: The effective area for *Fermi*-LAT as described by the *P6_v3_diff* IRF. The lefthand column shows the effective area of the front section of the detector (see §2.2.1) and the righthand column the back. The top row shows the joint dependence of the effective area on energy and incidence angle. The second row shows the dependence on energy for a set of incidence angles, while the third row shows the dependence on incidence angle for a set of energies. We note that effective area is essentially zero below 100 MeV, turns on rapidly, and becomes essentially flat at 1 GeV. This behavior is governed by the difficulty in reconstructing "soft" photons.

signal in the TKR (CAL), so to first order E' and $\vec{\Omega}'$ are also independent. We discuss each of these terms below.

3.4.1 The Point-spread Function

The point-spread function (PSF) lies at the heart of point source analysis. The spatial distribution of photons on the sky is the single most important handle for disentangling confused sources. Accordingly, we take some care in characterizing and (implementing) the PSF.

As with the effective area and energy dispersion, the PSF is determined by end-to-end Monte Carlo simulations of the detector. The Monte Carlo events are binned in (true) energy and (true) incidence angle[6]. The resulting distribution of reconstructed incidence angle is fit with a parameterized analytic function. These parameters describe the PSF over the instrument's phase space.

The PSF by definition is the probability density to observe an event in a small piece of solid angle $d\Omega$ centered on $\vec{\Omega}$ given that event originated from $\vec{\Omega}_0$. That is, for some energy and incidence angle, suppressed here, the PSF is

$$\frac{dN(\vec{\Omega})}{d\Omega} = \frac{dN(\vec{\Omega})}{d\phi\,d\cos\theta} \approx \frac{1}{2\pi}\frac{dN(\cos\theta)}{d\cos\theta} \approx \frac{1}{\pi}\frac{dN(\theta^2)}{d\theta^2},\tag{3.4}$$

where ϕ is the azimuthal angle and θ the polar angle measured with respect to the origin $\vec{\Omega}_0$, and where we have made the approximations of azimuthal symmetry and small angular deviations. We see that the natural variable is θ^2, the squared angular deviation, as isotropic sources ($dN/d\Omega$ constant) are flat in this quantity. This motivates the definition of $u \equiv \theta^2/2\sigma^2$, in terms of which

$$\frac{dN(\vec{\Omega})}{d\Omega} \approx \frac{1}{2\pi\,\sigma^2}\frac{dN(u)}{du}.\tag{3.5}$$

While we could in principle use the Monte Carlo histograms directly as a representation of the PSF, a fitted analytic function is much more convenient; we thus need a function $dN(u)/du$ that adequately characterizes the instrument performance. The functional form adopted for the PSF is the *King function* which arises in characterization of many-body

---

[6]The spectrum of incident events is uniform in logarithmic energy.

gravitational dynamics[28]. Mathematically, it is a precursor to the Gaussian distribution. That is, if the King function is defined as

$$f_k(u, \gamma) \equiv \left(1 + \frac{u}{\gamma}\right)^{-\gamma}, \tag{3.6}$$

then since

$$\lim_{\gamma \to \infty} \left(1 + \frac{u}{\gamma}\right)^{-\gamma} = \lim_{\gamma \to \infty} f_k(u, \gamma) = \exp(-u), \tag{3.7}$$

the King function can be seen as a "finite γ" version of the exponential. Since $u \equiv \theta^2/2\sigma^2$, we see that the limiting form is a Gaussian in the angular deviation. Thus, the smallness of γ characterizes how far we are from the "ideal" Gaussian PSF, and σ sets the overall angular scale. (Note that $\gamma > 1$ for the function to be well-behaved.) We see that, unlike the Gaussian, the King function has power law tails with slope $-\gamma$ for $u >> \gamma$, and is thus more suitable for characterizing a PSF with broad tails. A King function fit to Monte Carlo data is shown in the top panel of Figure 3.4.1.

Thus, for a given energy and incidence angle, the PSF is described by two parameters, σ and γ. In practice, these parameters are evaluted for bins in energy (4 bins per decade) and incidence angle (bins of width 0.1 in $\cos(\theta)$) and stored in a FITS table as a part of CALDB, the repository for instrument calibration information. In order to retain the smooth energy dependence of the underlying behavior and to reduce the steepness of the parameter dependence on energy, the σ parameter has a built-in energy dependence. That is, $\sigma \equiv \sigma_{\text{CALDB}} \times f_\sigma(E)$, with

$$f_\sigma(E)^2 \equiv \left(p_1(E/100\,MeV)^{-p_2}\right)^2 + p_3^2, \tag{3.8}$$

with p_1 and p_3 dependent on the conversion type of the event and $p_2 = 0.8$ in current versions of the IRF. p_3, the high-energy asymptote, is governed by the pitch of the silicon strips in the tracker. σ_{CALDB} is the parameter value stored in CALDB and is typically of order unity; p_1, p_2, and p_3 also appear in the CALDB.

The PSF must be normalized. By integrating $f_k(u)$ from 0 to ∞, we see $f_k(u)' \equiv (1 - 1/\gamma)f_k(u)$ is normalized to 1. Thus, the final (King function) form of the PSF becomes

$$\frac{dN}{d\Omega} \approx \frac{1}{2\pi\,\sigma^2} f_k'[u(\theta)] = \frac{1}{2\pi\,\sigma^2}\left(1 - \frac{1}{\gamma}\right)\left(1 + \frac{\theta^2}{2\gamma\sigma^2}\right)^{-\gamma} \tag{3.9}$$
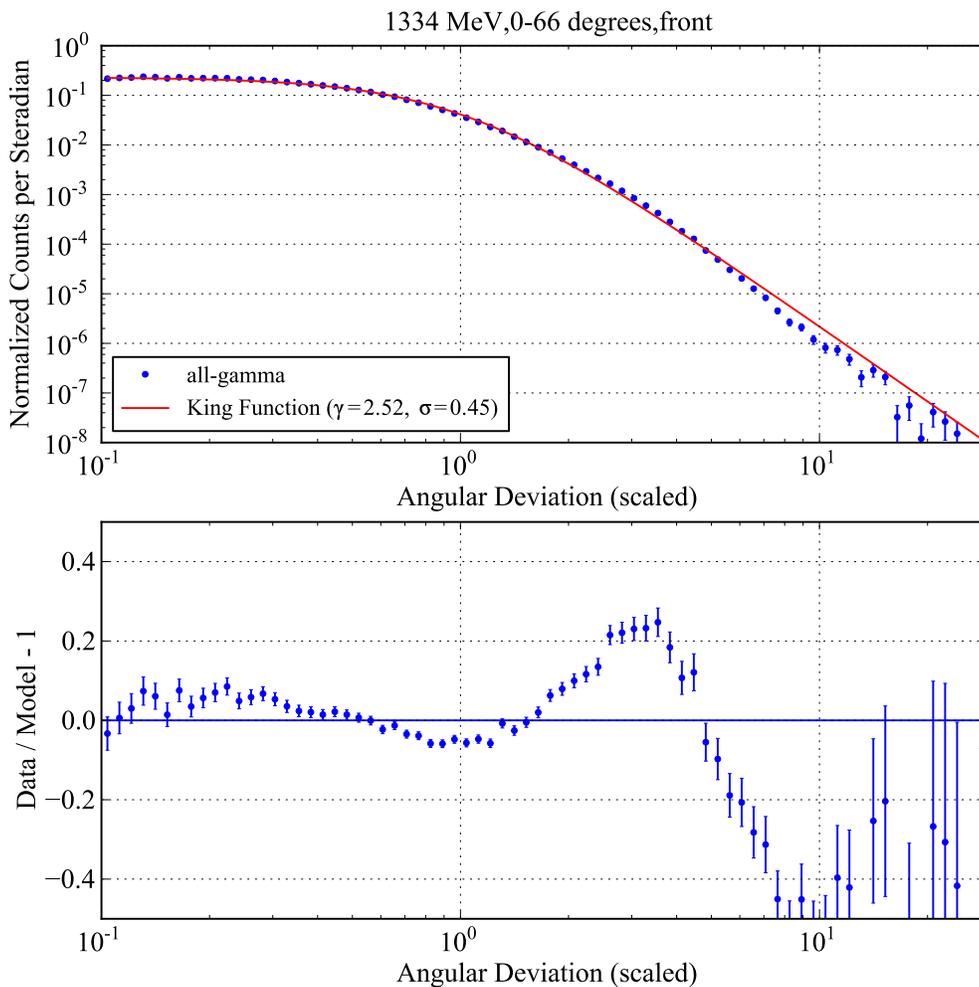
Figure 3.4: Monte Carlo data from a detector simulation for a bin centered on 1334 MeV for front-converting events. The events have been summed over incidence angle up to 66.4°. The behavior of the density function over many orders of magnitude is well-described by a single King function, i.e., that data show a sharp core and power law tails. However, there is a slight break in the slope of the power law tail, and accordingly the King function cannot provide a perfect fit, leaving residuals > 10%. In later versions of the IRF, this issue is ameliorated by the use of two King functions.

In addition to determining the asymptotic behavior at large u, γ also determines the transition point at $u = \gamma$. This dual rôle for γ sets a stringent relation between the tail steepness and the width of the core which is not always satisfied by the data (see Figure 3.4.1). E.g., improvements of the reconstruction algorithm delivered a PSF with an enhanced core and a "broken" power law tail. A single King function is unable to model this more complex shape. A new analytic function, the weighted sum of two King functions, on the other hand, adequately models the PSF. Thus, each energy/cos(θ) bin actually has 5 independent parameters, σ_c, σ_t, γ_c, γ_t, and w, i.e., the parameters for a core and a tail King function and the w, the relative weighting of the two.

3.4.2   Energy Dispersion

As with the reconstructed position, the reconstructed energy is subject to uncertainty. At the highest energies, above about 100 GeV, an appreciable fraction of the energy in the electromagnetic shower escapes the calorimeter, making an energy estimate difficult. At lower incident energies, a significant fraction of the photon energy may be deposited in the TKR. In between these two extremes, there is uncertainty due to shower particles missing the active elements of the calorimeter and the natural scatter induced when estimating the deposited energy from the crystal light yield.

The magnitude of these effects are shown in Figure 3.4.2, where it can be seen that the half-width at half-maximum (HWHM) at most energies is less than 10% but with appreciable tails and rather wider HWHMs at low energy, about $15 - 20\%$. This level of uncertainty is smaller than—but not *much* smaller than—the width of the energy bins introduced in Chapter 4, which have a full width of about 30% at 8 bins per decade.

In the following sections, we ignore the effects of energy dispersion by assuming it is negligible. While we see here that this is not the case, we make this simplifying assumption because it drastically reduces the computational burden of spectral analysis. (Including the effects of energy dispersion essentially requires an extra quadrature for all computed quantities.) We expect this approximation may not be too bad since, unlike the PSF, the energy dispersion does not itself depend strongly on energy. If our effective area were

independent of energy, then for power law sources, the only effect of energy dispersion would be a slight softening of the measured slope and a slight increase in the measured flux. Since the effective area rapidly rises for the first decade of the LAT passband, the energy dispersion does introduce some curvature in the measured spectra and causes *hardening* of measured spectra and *decrease* of measured flux, as we shall see in §5.1.5. However, the resulting bias on the measured parameters is found to be $< 10\%$.

3.4.3 Time Dispersion

The LAT uses GPS for a time reference and measures the absolute time of photon arrival with an accuracy better than 1μs [78, 22]. The error is negligible for nearly all applications, so we approximate $f_{\text{tdisp}} = \delta(t' - t)$.

3.4.4 Approximate Observed Event Rate

Given these approximations, the modeled observed event rate for a source becomes

$$r(E', \vec{\Omega}', t'; \vec{\lambda}) = \int d\Omega \, \mathcal{F}(E', t', \vec{\Omega}; \vec{\lambda}) \, A[E', \cos\theta(t')] \, e(E', t') \, f_{\text{psf}}[\vec{\Omega}'; \vec{\Omega}, \cos\theta(t'), E']. \quad (3.10)$$

For a point source, we obtain

$$r(E', \vec{\Omega}', t'; \vec{\lambda}, \vec{\Omega}_0) = \mathcal{F}(E', t'; \vec{\lambda}) A[E', \cos\theta(t)] \, e(E', t') \, f_{\text{psf}}[\vec{\Omega}'; \vec{\Omega}_0, \cos\theta(t), E']. \quad (3.11)$$

The simplification delivered for point sources by the accurate timing of the LAT and the neglect of energy dispersion is that we may express the observed rate without *any* integration. Diffuse sources, on the other hand, require a two-dimensional convolution to calculate the observed rate.

3.5 Poisson Likelihood for the Fermi-LAT

We now outline the formulation of the likelihood for the *Fermi*-LAT . As with all particle detectors, *Fermi*-LAT is a counting experiment. Events are binned by the observed quantities, particularly energy and position. They may also be binned in time, or accumulated into a single bin spanning the length of the observation. Thus, in the notation of §3.2, $\vec{X}$ is the set of counts observed in each bin, which we now relabel $\vec{N}$.
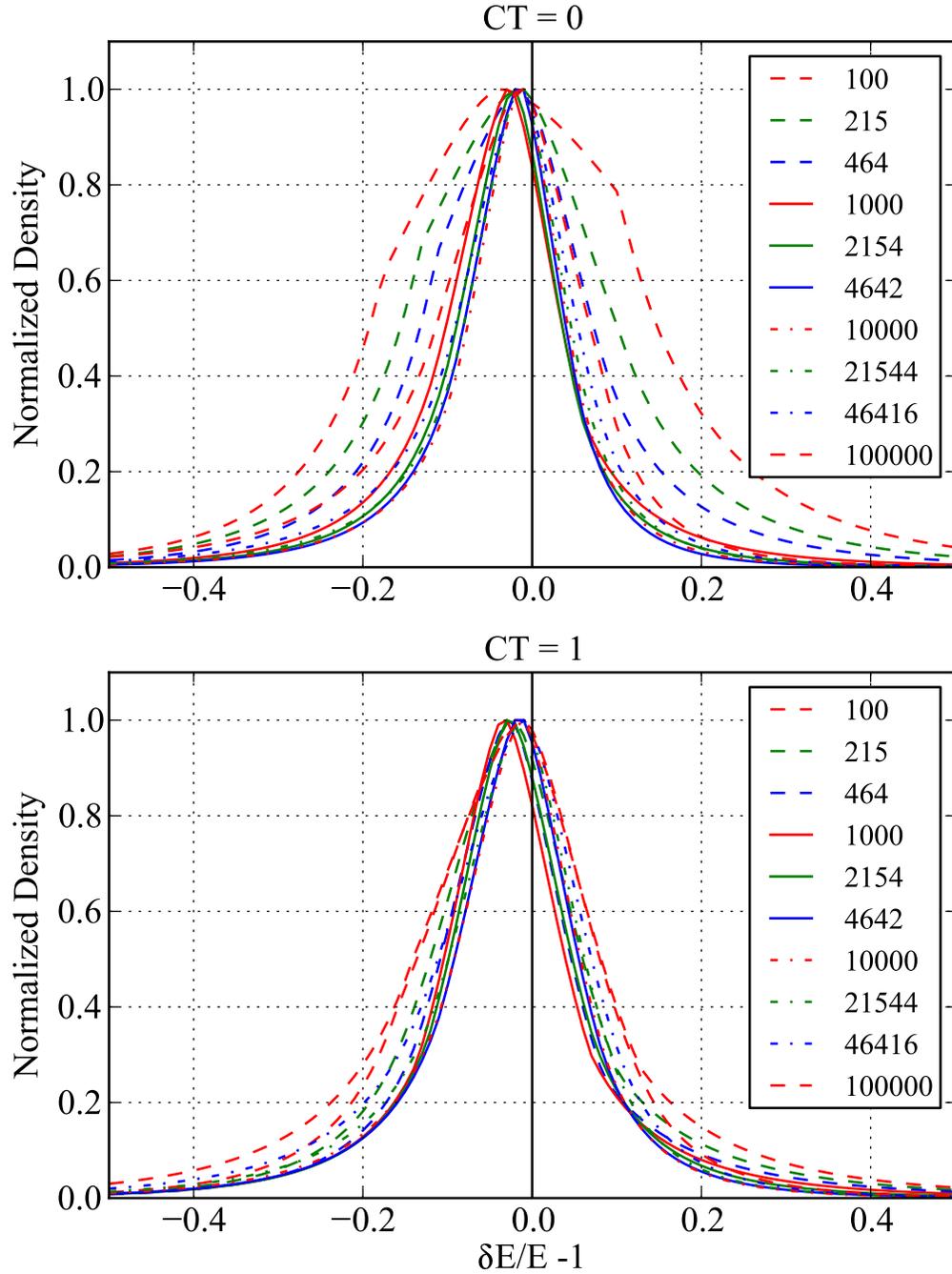
Figure 3.5: The energy dispersion (observed/simulated - 1) for the *P6_v3_diff* IRF for events incident on-axis. The top (bottom) panel shows events converting in the front (back) of the TKR. At low (order 100 MeV) energies, the resolution is significantly better for back-converting events since the shower commences closer to the calorimeter, i.e. less energy is deposited in or escapes from the TKR.

Each element of $\vec{N}$ is distributed according to a Poisson distribution with unknown mean, r_i. We model this mean by integrating either Eq. 3.10 (diffuse sources) or Eq. 3.11 (point sources) over the bin. Since the components of $\vec{N}$, N_i, are mutually statistically independent, the probabily mass function for the data is the product of Poisson distributions with rates r_i. We recall that for a Poisson distribution with mean r, the probability to observe N counts is

$$p(N; r) = \frac{r^N}{N!} \exp(-r).$$ (3.12)

As noted in the introduction to this chapter, with the broad PSF of the LAT and the strong diffuse background present in the GeV sky, the rate for a bin of phase space involves contributions from multiple sources. Thus, using the probability mass function in Eq. 3.12, we write the (logarithm of the) binned likelihood for all selected data by summing over all N_{bins} bins and all N_s sources:

$$\log \mathcal{L}(\vec{\lambda}; \vec{N}) = \sum_{i=1}^{N_{bins}} \left[-\iiint_{\text{bin}_i} \sum_{j=1}^{N_s} r_j(E', \vec{\Omega}', t'; \vec{\lambda}_j) + N_i \log \sum_{j=1}^{N_s} \iiint_{\text{bin}} r_j(E', \vec{\Omega}', t'; \vec{\lambda}_j) \right]$$ (3.13)

$$\equiv \sum_{i=1}^{N_{bins}} \left[-\sum_{j=1}^{N_s} R_{ij} + N_i \log \sum_{j=1}^{N_s} R_{ij} \right]$$ (3.14)

Here, N_i are the observed counts in the ith bin, and the triple integrals are over the energy, position, and time range specified for each bin. In the second line, we have defined R_{ij}, the expected number of counts in the ith bin from the jth source. We have discarded the $N!$ term from Eq. 3.12 since it is independent of the model parameters. This is the form of the likelihood we shall find useful in the following chapters.

Typically, the bins are continguous, and we define the *region of interest* (ROI) to be the portion of the total data (the volume of phase space) we have selected. The likelihood then becomes

$$\log \mathcal{L}(\vec{\lambda}; \vec{N}) = -\iiint_{\text{ROI}} \sum_{j=1}^{N_s} r_j(E', \vec{\Omega}', t'; \vec{\lambda}_j) + \sum_{i}^{N_{bins}} N_i \log \sum_{j=1}^{N_s} R_{ij}.$$ (3.15)

This formulation facilitates the introduction of *unbinned* Poisson likelihood, in which the bin widths are taken to be infinitesimal, such that only 0 or 1 counts can possibly be observed. This formulation sacrifices no information to binning but can become prohibitive for large

data sets. The unbinned likelihood, with $\vec{N}$ now containing *every* count, is given by,

$$\log \mathcal{L}(\vec{\lambda}, \vec{N}) = - \iiint_{\text{ROI}} \sum_{j=1}^{N_s} r_j(E', \vec{\Omega}', t'; \vec{\lambda}_j) + \sum_{i=1}^{N_{\text{events}}} \log \sum_{j=1}^{N_s} r_j(E'_i, \vec{\Omega}'_i, t'_i; \vec{\lambda}_j), \qquad (3.16)$$

where E'_i denotes the reconstructed energy of the ith event, etc.

3.6 Summary

We have motivated the use of likelihood principles in GeV astronomy by demonstrating that—at least for current telescopes—it is impossible to make reliable measurements of astrophysical quantities directly from the data. Instead, we noted that GeV sources can be easily modeled with only a few parameters and suggested the method of maximum likelihood to estimate these parameters. To calculate the likelihood, we fold the parameterized models through the instrument response function, thus naturally accounting for the complex IRF, source confusion, and the constantly-changing orientation of the instrument. We outlined the features of the IRF and provided a convenient analytic approximation for the PSF in the form of the King function, and we argued that we could ignore the effects of energy dispersion at the cost of some bias to our measured parameters. Finally, we provided an expression for the log likelihood appropriate for *Fermi*-LAT (Eq. 3.13) and for the expressions for the observed rates from diffuse sources (Eqs. 3.10 and 3.11) whose evaluation shall be our endeavor in the next chapter.

Chapter 4

THE *POINTLIKE* PACKAGE: DESIGN AND IMPLEMENTATION

As outlined in Chapter 3, likelihood analysis is the optimal method for characterizing γ-ray sources. (We discuss in Chapter 6 likelihood techniques for source *detection* as well.) Unfortunately, likelihood analysis is complex. One must manage source models, descriptions of the instrument, and data. Forward folding the models into "data space" can be computationally intensive. Analysis for a single source (which, due to the PSF-broadening, entails analysis of many sources!) can be time consuming. All-sky analysis requires large computer clusters to handle the computational burden. However, by binning the data with scalable pixels and sacrificing some accuracy for time-saving approximations, this situation can be ameliorated. In this section, we describe the *pointlike* software package, a collection of routines for performing maximum likelihood (ML) analysis of *Fermi*-LAT data, which aims for the canonical goal of "making easy things easy and hard thing possible".

At the highest level, *pointlike* is a collection of Python modules designed to be used both interactively (for analyzing a single source) and in batch mode (for analyzing many sources). *pointlike* also relies internally on a large C++ codebase via SWIG[1], a tool providing wrapper code to access compiled libraries from interpreted languages. The interface and organization are more properly left to documentation of the code, and in this chaper, we discuss the "nuts and bolts" of the implementation.

4.1 Fermi-*LAT Data*

For all the complexity of the instrument, the high-level *Fermi*-LAT data is simple, a list of meausured quantities from reconstructed events. In the following section, we provide details about this list, our method for binning the events, and some necessary cuts that must be made to reject background.

---

[1]http://www.swig.org

4.1.1 High-level Data Products

Information about each event is telemetered from the S/C to a batch farm at SLAC for event reconstruction, in which the low-level hardware readouts are converted into estimates of physical quantities[22]: the energy, position, and time of arrival of the putative γ ray. There are two high-level data products. The *FT1* file is an event list in FITS format with columns containing reconstructed quantities (energy, R.A., Decl., time, etc.) for each event. The *FT2* FITS file tabulates the position and orientation of the S/C at 30-second intervals, information needed to construct the time-dependent IRF.

4.1.2 Binning the Data: HEALPix

Binning *Fermi*-LAT data is not a trivial task. Due to the energy-dependence of multiple scattering in the tracker, the LAT's angular resolution varies by nearly two orders of magnitude over its energy range (Figure 3). If one uses only a single size for position bins—e.g., the EGRET choice of $0.5° \times 0.5°$[66]—then one faces a tradeoff: lose information at high energies by using bins much larger than the ultimate instrumental resolution, or increase the computational burden by using bins much too small for the poor angular resolution at low energies.

We address this problem with HEALPix[53], a scheme for tessellating the sky with equal-area pixels[2]. The base tessellation comprises 12 pixels, and finer binnings are achieved by subdividing the base pixels; if there are N_{side} subdivisions on a side of the base pixels, there are $12 \times N_{\mathrm{side}}^2$ pixels. N_{side} thus controls the granularity of the pixelization. (A useful relation is that a pixel is approximately $60°/N_{\mathrm{side}}$ on a side.) The finest binning possible is limited by computer architecture, as each pixel must be uniquely mappable to an integer. We enforce $N_{\mathrm{side}} \leq 8192$[3], yielding a maximum resolution of about $30''$, more than adequate for the finest resolution PSF of the LAT.

In addition to the finely-controlled pixel size, HEALPix facilitates sparse binning. Due

---

[2]HEALPix also describes a particular projection used in demarcating the pixels. We do not make explicit use of this feature.

[3]8192 is the largest power of two for which $12 \times N_{\mathrm{side}}^2 < 2^{31}$, the largest value of a signed integer available on 32-bit architecture.

to the falling spectra of γ-ray sources and the flat effective area of *Fermi*-LAT , there are relatively few high energy photons and most pixels are empty. By assigning each pixel a unique index, the data can take the form of a mapping of index to counts, and we need only store those indices with non-zero counts.

Data are initially binned into front-converting (those that pair produce in the forward section of the TKR with the thin tungsten foils) and back-converting (producing the first pair on a thick tungsten foil deep in the TKR) events. This distinction is maintained throughout the *pointlike* package, and a distinct IRF is used for each class. Data are next divided into uniform bins in logarithmic observed energy. (Logarithmic energy is the natural choice for broadband spectra, which often look like power laws.) By default, we use either 4 or 8 bins per decade, with the left edge of a bin at 100 MeV. Such energy bins align with that used in CALDB for the PSF parameters and allow the use of the same PSF parameters over an entire bin width.

For each energy bin (and conversion type), the N_{side} parameter is chosen so that the pixels are somewhat smaller than the core of the PSF at that energy. Precisely,

$$N_{\text{side}} = \frac{8192}{1 + 8192/N_0 \times (E/100\,\text{MeV})^{-0.8} \times \exp{-(E/2000\,\text{MeV})^2}}, \qquad (4.1)$$

with $N_0 = 86$ ($N_0 = 52$) for front (back) events. This binning puts roughly 5 pixel widths within the 68% containment radius of the PSF at the given energy, up to about 1 GeV. At 1 GeV, this binning rapidly asymptotes to $N_{\text{side}} = 8192$. At 1 GeV, the PSF-based pixels are sufficiently small and the data sufficiently sparse that the mean pixel occupancy drops to 1. We gain no computational savings by using PSF-based pixels, and so use the smallest pixels possible. Above 1 GeV, the *pointlike* binning becomes essentially unbinned in position. (The binning in energy persists, of course, as this binning allows us to define a coarse-grained PSF.) In Figure 4.1.2, we show the mean pixel occupancy for 18 months of data using the scheme in Eq. 4.1. The compression ratio at low energies is large, especially for back-converting events, and will grow linearly with time. (Essentially every pixel at low energies is occupied.) For all energies, the total compression ratio is about 2.8.

The "unbinned limit" is relatively robust against loss of efficiency with increasingly large data sets. The primary worry is if, by selecting very small pixel sizes, we prevent the
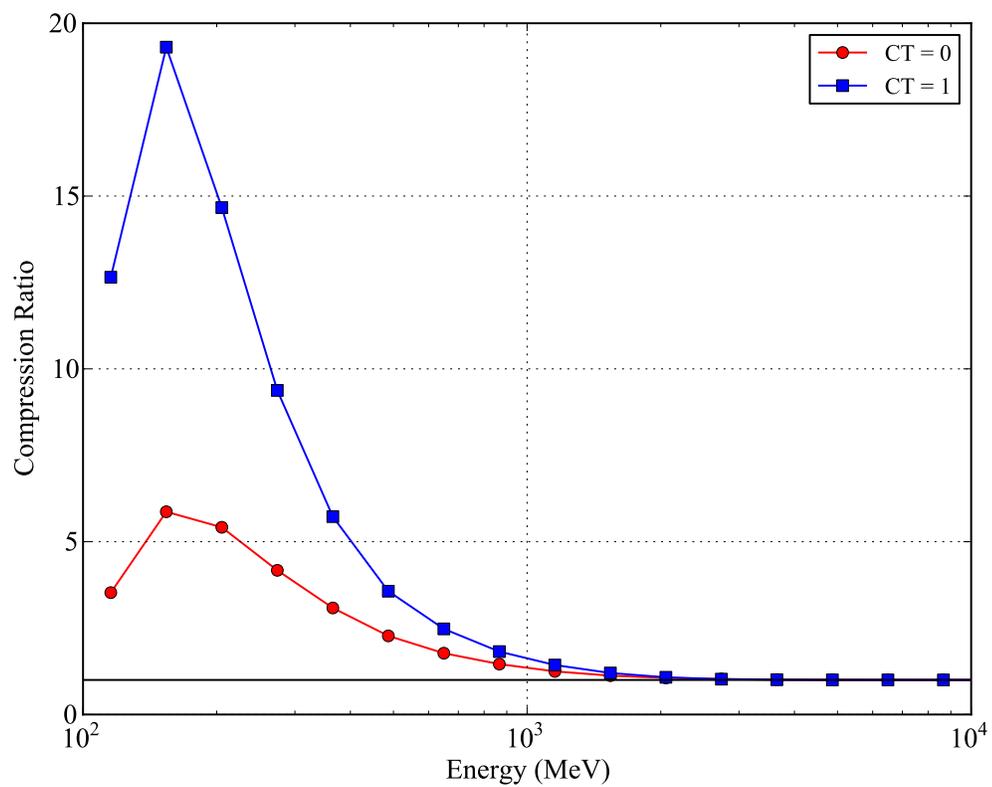
Figure 4.1: The compression ratio of the data, i.e., the mean photons per pixel, for a given energy band. The occupancy drops rapidly as the energy approaches 1 GeV due to the improving PSF and falling spectra of astrophysical sources.

mean occupancy of a particular band from increasing above 1, thus gaining performance. However, the generally steep spectra of GeV sources and the sharpening PSF conspire to prevent this. At 1 GeV, for instance, the scheme of Eq. 4.1 gives $N_{\text{side}} = 644$ (401) for front (back). These are approximately PSF-sized pixels, and there are 4.98 and 1.93 million pixels, respectively, for these N_{side} settings, while for an 18-month data set, only 0.36 and 0.33 million of these pixels, respectively, are occupied. Even in the extremely conservative limit that the number of occupied pixels grows with time, we must collect 9 years of data to reach a non-sparse pixelization for 1 GeV. *A fortiori* for higher energies with fewer photons and smaller PSF-based pixels.

During the binning process, we make additional cuts. An important facet of the binned analysis is that the PSF not depend too much on incidence angle, so that a "collective" PSF for a bin makes sense. This is generally true (see Figure 4.1.2); the variation of the PSF with incidence angle is on the order of 10%[4]. We typically reduce the variation even more by discarding events with a reconstructed incidence angle $> 66.4°$. Thus, a conservative choice of a pixel size appropriate to the on-axis PSF will remain efficient for all incidence angle.

Another cut is made to remove most of the γ rays produced by cosmic ray interactions in Earth's upper atmosphere[4]. Ultrarelativistic cosmic rays lead to a γ-ray cross-section in the Earth frame strongly peaked in the forward direction, and so from the point-of-view of the S/C, the intensity is approximately proportional to the total grammage of the atmosphere along the line-of-sight, i.e, the emission is dominated by the limb of the earth. The polar angle of the limb as measured from the zenith is given by

$$180° - \arcsin\left(\frac{R_{\text{earth}}}{R_{\text{earth}} + R_{\text{Fermi}}}\right),\tag{4.2}$$

about $113°$ from the zenith for Fermi's altitude of 550km. This number varies slightly (linear in the altitude) during the orbit, which is slightly elliptical, and decreases as the S/C slowly loses altitude to atmospheric drag. Although the position of the limb is independent of the S/C orientation, more albedo photons are successfully reconstructed when more of the limb

---

[4]Although the PSF containment radius increases to up to 50% of the on-axis value, the effective area is approximately proportional to $\cos(\theta)$ and very few events are collected with this large PSF.
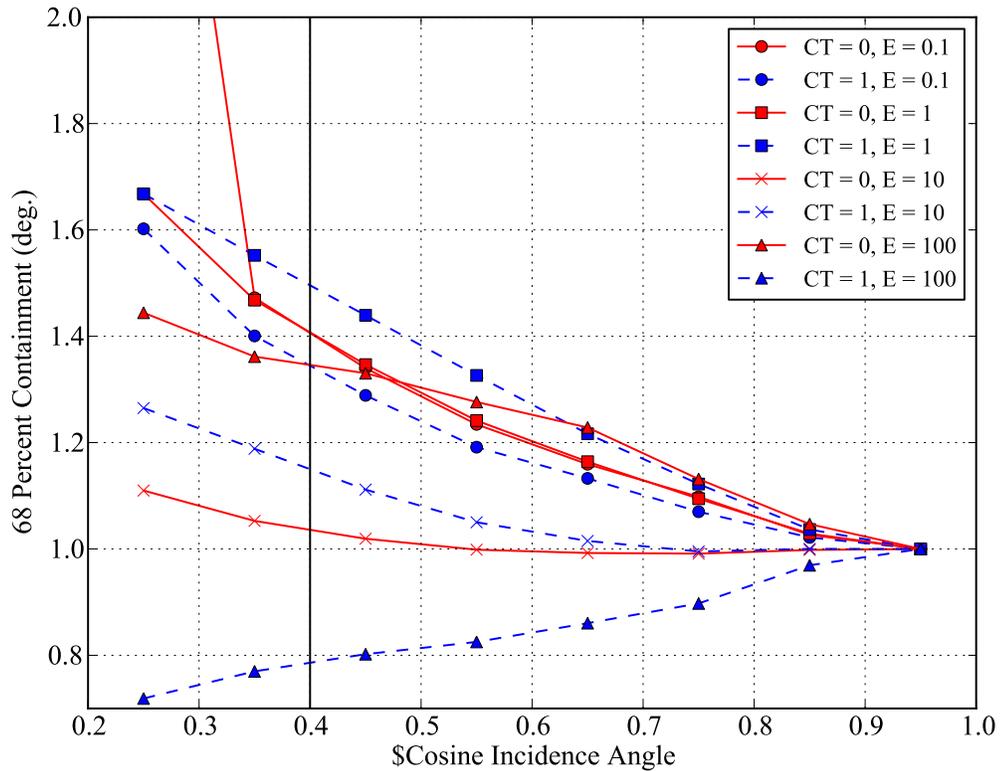
Figure 4.2: The PSF—characterized by the radius in which 68% events are enclosed—as a function of incidence angle, energy (in GeV), and conversion type ("CT" in legend). Generally, the angular resolution is best on-axis and degrades as the edge of the field-of-view is approached. Using a larger containment radius, e.g., 95%, as a metric yields nearly the same dependence on incidence angle and magnitude of shift. The standard cut on reconstructed incidence angle for *pointlike* of $\cos(\theta) = 0.4$, or $> 66.4°$, is indicated by the vertical black line.

is in the field of view, i.e., when the S/C is at larger rocking angles. At energies > 100 MeV, the reconstructed zenith angle of albedo photons is generally > 105° (see Figure 4.1.2); only events in the tails of the PSF at the lowest energies "disperse" more than 8° from their true direction[5] We therefore apply a cut on the reconstructed zenith angle > 105°.

Since the cuts on incidence and zenith angles use the reconstructed photon position, some events that would be cut if the true position were known are included, and others that would pass a cut in the limit of perfect reconstruction are discarded. In addition to such contamination, a small bias results. For instance, since the effective area increases approximately linearly with the cosine of the incidence angle, a cut on incidence angle will necessarily discard more events whose true incidence angle is less than the cut value. Since the limb tends also to be on the edge of the FOV, similar remarks apply. These effects are small, and we will ignore them.

Finally, we only allow events that lie within a *Good Time Interval* (GTI). These GTI form an additional table in the *FT1* file and are simply a list of tuples with the start and stop time of an interval during which we accept events. By default, the GTI are determined by SAA passages. (This cut is trivial since no data is taken during SAA passage.) However, additional temporal cuts may be made based on the S/C position and orientation. For instance, as the S/C rocks, it may overshoot its target rock angle and spend time at a larger rock angle with higher albedo background. Similar excursions are caused when the S/C makes a pointed observation. Such times can be excluded from the GTI. Gamma-ray bursts (background for other sources) may also be so excluded. Finally, when studying a particular source, in order to further reduce contamination from albedo events, times when the ROI is too close to the limb can be excluded.

4.2 Point Sources Rates

The data, whose treatment we have just discussed, and the source models (discussed in Chapter 3) form the "bookends" of the likelihood analysis, and we have now to link the

---

[5]In fact, a nontrivial—relative to *astrophysical* source rates—number of limb photons with energies of order 100 MeV survive this cut. The remaining photons contribute a non-isotropic foreground to astrophysical sources, and should be accounted for in background modeling. [The contamination is increased with a larger rock angle.]
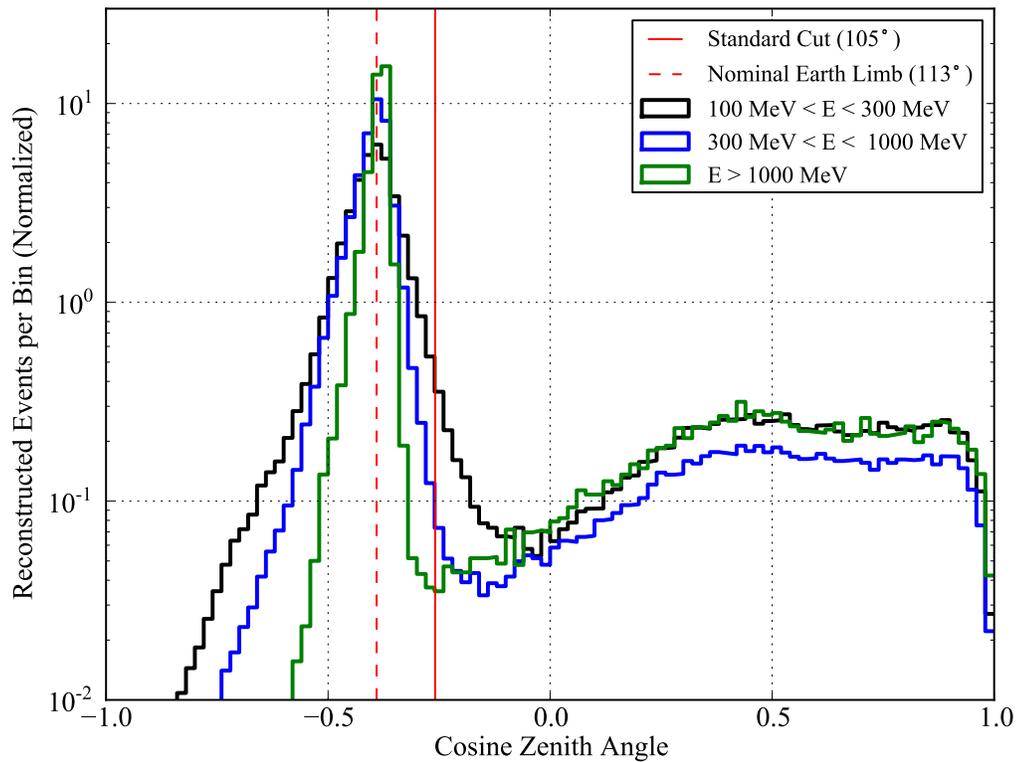
Figure 4.3: Reconstructed events (P6 "diffuse" class) with 100 MeV < E < 100 GeV in a 10° aperture centered on (l,b) = (0,-45), integrated from Aug. 4 2008 to Jul. 18 2010. The binning is in the reconstructed zenith angle. The effect of the PSF is clearly seen in the broadening of the distribution of limb photons about $\approx 113°$.

two by performing the integrals in Eq. 3.13 and folding the intrinsic source rates through the instrument response function. These integrals represent the primary computational (and infrastructural) difficulty in likelihood spectroscopy, and we devote some effort making time-saving approximations when possible while maintaining good overall accuracy.

There are two fundamental source classes: those that can be spatially resolved (diffuse sources) and those that cannot (point sources). The former includes the Galactic diffuse background due to cosmic rays, the extragalactic background due to distant sources below detection threshold, nearby galaxies such as the Magellanic clouds, supernova remnants, and some pulsar wind nebulae. The class of point sources includes the active nuclei of galaxies, pulsars, and small or distant remnants and nebulae. Since diffuse sources require convolution with the PSF, the two classes are treated independently, and we discuss point sources and diffuse sources in turn.

We begin with Eq. 3.11, which we recall is the differential (in phase space) expected rate from a point source, neglecting energy dispersion. To compare this to observations, we need to integrate it over each of the bins we have selected for the data. We begin with integration over observed energy and time, i.e., we calculate the quantity

$$r_i(\vec{\Omega}') = \int dE' \int dt' \mathcal{F}(E', t'; \vec{\lambda}) A[E', \cos \theta(t)] \, e(E', t') \, f_{\text{psf}}[\vec{\Omega}'; \vec{\Omega}_0, \cos \theta(t), E'], \qquad (4.3)$$

the expected rate per unit solid angle in the ith bin. (We leave the particular energy and time abstract; the bounds of the integrals over energy and time are implicit and extend over the ith bin.) For brevity, we suppress the prime notation in the sequel.

4.2.1 Integral Over Time

In order to evaluate this expression quickly, we make approximations and assumptions that allow us to pre-compute some quantities and simplify the integral. First, we assume that $\mathcal{F}(E, t; \vec{\lambda}) = \mathcal{F}(E; \vec{\lambda})$, i.e., that the source is stationary during the time range specified by the bin. Alternatively, we can view this assumption as measuring the time-averaged flux from the source[6]. If we are interested in capturing the time dependence of the source, we simply use multiple time bins and proceed as before.

---

[6]In fact, due to the time-dependent IRF, these statements are only equivalent to first order.

Next, we write $e(E,t) = e_0(t) + e_1(E,t)$. e_0 is the dominant contribution to the efficiency, determined by the deadtime (trigger rate) and is energy independent. e_1 is the second-order correction in which the reconstruction effectiveness changes based on the charged-particle background. To a very good approximation, this correction depends linearly on the trigger rate, e_0, i.e., $e_1(E,t) = f_1(E)e_0(t) + f_2(E)$, and we can reconstruct the full quantity from the simple-to-measure e_0.

Now, we break up the integral over time into sub-integrals,

$$\int dt \to \sum_{i=1}^{N} \int_i dt, \tag{4.4}$$

where each integral is over one of the (typically) 30-s intervals laid out in the FT2 file, and there are N such intervals in the parent bin. The sampling rate of the FT2 file is designed to be sufficiently high that the S/C orientation[7] does not change much over the course of an interval. Therefore, we evelute each sub-integral with a central-value approximation, and the time integral becomes

$$\sum_{i=1}^{N} \Delta t_i \, A[E, \cos\theta(t_i)] \, e(E,t_i) \, f_{\text{psf}}[\vec{\Omega}; \vec{\Omega}_0, \cos\theta(t_i), E], \tag{4.5}$$

with t_i the center of the ith FT2 bin and Δt is the bin width, typically 30 seconds.

Next, we discretize $\cos\theta$, typically into 40 bins spanning from 0.2 to 1.0[8]. The time integral becomes

$$\sum_{j=1}^{N_{\cos}} \sum_{i=1}^{N} \Delta t_i \, \mathcal{I}_{\cos\theta_j}[\cos\theta(t_i)] A[E, \cos\theta(t_i)] \, e(E,t_i) \, f_{\text{psf}}[\vec{\Omega}; \vec{\Omega}_0, \cos\theta(t_i), E], \tag{4.6}$$

with $\mathcal{I}_{\cos\theta_j}[\cos\theta(t_i)]$ the indicator function that evaluates as 1 if $\cos\theta(t_i)$ is in the jth $\cos\theta$ bin and 0 otherwise. We define the livetime, $\tau_j(E) \equiv \sum_{i=1}^{N} \Delta t_i \, \mathcal{I}_{\cos\theta_j} e(E,t_i)$ and the exposure, $\epsilon_j(E) = \tau_j(E) \times A[E, \cos\theta_j]$, in terms of which the time integral becomes

$$\sum_{j=1}^{N_{\cos}} \epsilon_j(E) \, f_{\text{psf}}[\vec{\Omega}; \vec{\Omega}_0, \cos\theta_j, E], \tag{4.7}$$

---

[7]This assumption is only marginally accurate when the S/C slews from its north-pointing attitude to its south-pointing attitude, but such a slew only occurs once per orbit, and $< 10\%$ of the livetime is accumulated during a slew.

[8]The LAT field of view is $\cos\theta \geq 0.2$.

or in terms of the summed exposure,

$$\epsilon(E) \sum_{j=1}^{N_{\cos}} \epsilon_j(E)/\epsilon(E) f_{\mathrm{psf}}[\vec{\Omega}; \vec{\Omega}_0, \cos\theta_j, E]. \tag{4.8}$$

We define the exposure-weighted PSF, $f_{\mathrm{wpsf}}(\vec{\Omega}; \vec{\Omega}_0, E) = \sum_{j=1}^{N_{\cos}} \epsilon_j(E)/\epsilon(E) f_{\mathrm{psf}}[\vec{\Omega}; \vec{\Omega}_0, \cos\theta_j, E]$, and the time integral becomes, finally,

$$\epsilon(E, \vec{\Omega}_0) f_{\mathrm{wpsf}}[\vec{\Omega}; \vec{\Omega}_0, E]. \tag{4.9}$$

The immense benefit of all of this definition legerdemain is that the livetime calculation is entirely independent of the source. We can pre-compute the livetime for some pixelization on the sky[9] and have it available for all subsequent analyses using the same data. The time integral is thus entirely eliminated in favor of a sum over ≈ 40 bins in incidence angle, a great improvement over the original sum over the S/C pointing history containing potentially millions of terms!

4.2.2 Integral over Energy

The bin rate is now

$$r_j(\vec{\Omega}) = \int dE \, \mathcal{F}(E, \vec{\lambda}) \, \epsilon(E, \vec{\Omega}_0) \, f_{\mathrm{wpsf}}(\vec{\Omega}; \vec{\Omega}_0, E). \tag{4.10}$$

As it stands, the expression is still computationally burdensome, as we would need to evaluate the PSF at each data pixel over many energies. To further simplify, we appeal to the Mean Value Theorem, according to which $\int dx \, f(x) \, g(x) = f(x') \int dx \, g(x)$, for some x' within the bounds of integration. Thus, if we determine the appropriate energy, E_{opt}, at which to evaluate f_{wpsf}, we can remove the PSF from the energy integral, obtaining

$$r_j(\vec{\Omega}) = f_{\mathrm{wpsf}}(\vec{\Omega}; \vec{\Omega}_0, E_{\mathrm{opt}}) \int dE \, \mathcal{F}(E, \vec{\lambda}) \, \epsilon(E, \vec{\Omega}_0). \tag{4.11}$$

Unfortunately, it is not possible to find a unique E_{opt}. Although within a given energy bin, the energy dependence of the PSF is entirely given by Eq. 3.8, the energy dependence

---

[9] $1°$ bins are typical

of the King function depends on position, i.e., E_{opt} is a function of $\theta = \arccos(\vec{\Omega} \cdot \vec{\Omega}_0)$. Indeed, for small θ^2, $f \propto \sigma^{-2} \propto E^{-1.6}$, while for large θ^2, $f \propto \sigma^{-2(\gamma-1)} \approx \sigma^2 \propto E^{1.6}$.

Rather than introducing an additional angular dependence through E_{opt}, we seek E_{opt} that gives the best representation of the PSF over a wide range of source spectra and positions. The natural metric for this is the log likelihood, Eq. 3.13. We concentrate on a single source centered in the ROI and write the expected counts for a given band as

$$\int d\Omega \int dE \, \mathcal{F}(E, \vec{\lambda}) \, \epsilon(E, \vec{\Omega}_0) \, f_{\mathrm{wpsf}}(\vec{\Omega}; \vec{\Omega}_0, E) \tag{4.12}$$

$$\approx \pi \int_0^{\theta^2_{max}} d\theta^2 \int dE \, \mathcal{F}(E, \vec{\lambda}) \, \epsilon(E, \vec{\Omega}_0) \, f_{\mathrm{wpsf}}(\theta^2, E) \tag{4.13}$$

$$\equiv \pi \int_0^{\theta^2_{max}} d\theta^2 \int dE \, N(E) f_{\mathrm{wpsf}}(\theta^2, E) \tag{4.14}$$

$$\equiv \int_0^{\theta^2_{max}} d\theta^2 \, g(\theta^2), \tag{4.15}$$

where the symmetry of the centered source allows the trivial azimuthal integral and we have made a small angle approximation. We now define an approximate rate using the "mean" Mean Value Theorem,

$$\pi \int_0^{\theta^2_{max}} d\theta^2 \int dE \, N(E) f_{\mathrm{wpsf}}(\theta^2, E) \tag{4.16}$$

$$= \pi \int_0^{\theta^2_{max}} d\theta^2 f_{\mathrm{wpsf}}(\theta^2, E_{\mathrm{opt}}(\theta^2)) \times \int dE \, N(E) \tag{4.17}$$

$$\approx \pi \int_0^{\theta^2_{max}} d\theta^2 f_{\mathrm{wpsf}}(\theta^2, E_{\mathrm{opt}}) \times \int dE \, N(E) \tag{4.18}$$

$$\equiv \int_0^{\theta^2_{max}} d\theta^2 f(\theta^2, E_{\mathrm{opt}}) \tag{4.19}$$

Now, the log likelihood for the energy band becomes

$$\log \mathcal{L}_1 = \int_0^{\theta^2_{max}} d\theta^2 \, g(\theta^2) - \sum_{i=1}^{N} n_i \log \int_{\theta^2_{i1}}^{\theta^2_{i2}} d\theta^2 g(\theta^2) \tag{4.20}$$

$$\rightarrow \int_0^{\theta^2_{max}} d\theta^2 \, g(\theta^2) \times [1 - \log g(\theta^2)], \tag{4.21}$$

where in the second line we have replaced the observed counts with the expected rate and

gone to the unbinned limit[10]. In terms of the approximate rate, the log likelihood is

$$\log \mathcal{L}_2(E_{\mathrm{opt}}) = \int_0^{\theta_{max}^2} d\theta^2\, N\, f(\theta^2, E_{\mathrm{opt}}) \times [1 - \log N\, f(\theta^2, E_{\mathrm{opt}})], \qquad (4.22)$$

and now we simply solve the integral equation $\log \mathcal{L}_1 = \log \mathcal{L}_2(E_{\mathrm{opt}})$ for E_{opt}.

In practice, E_{opt} is within a few percent of the geometric mean, $E_{\mathrm{geo}} \equiv \sqrt{E_{\max} \times E_{\min}}$, of the energy band, and we obtain a sufficiently good solution for E_{opt} with a single Newton-Raphson iteration,

$$E_{\mathrm{opt}} \approx E_{\mathrm{geo}} + [\log \mathcal{L}_1 - \log \mathcal{L}_2(E_{\mathrm{geo}})] \times \left[\frac{d\, \log \mathcal{L}_2(E_{\mathrm{opt}})}{d_{E_{\mathrm{opt}}}}(E_{\mathrm{geo}})\right]^{-1}. \qquad (4.23)$$

We apply this procedure to obtain an estimate of E_{opt} for each energy band (and event conversion type). The results for a typical ROI (centered on the position of the Vela pulsar) are shown in Figure 4.2.2. We see that, for a variety of source spectra, E_{opt} is within a few percent of E_{geo}. The largest discrepancy is at energies of a few hundred MeV. Here, the effective area rises rapidly (Fig. 3.4), meaning counts in the band are shifted towards higher energies, yielding a better effective PSF than E_{geo} implies. The same effect causes the dependence on spectral index: softer sources emphasize lower energies, shifting E_{opt} lower. The overall trend for an E_{opt} lower than E_{geo} can then be understood to have its origin in the $1/E$ (uniform in logarithmic energy) spectrum used to determine the PSF. In any case, at all energies, the E_{opt} for the different spectra are within a few percent of each other. Therefore, we adopt the omnibus prescription of calculating E_{opt} for a photon index (Γ) of 2.0. We present empirical results showing the improvement of the PSF representation using the "E_{opt}" prescription in §5.1.1.

We now begin the final step. We have the expected rate per unit solid angle, and we see from Eqs. 3.13 and 3.15 that we need to integrate this rate over both the entire ROI and over each data pixel (the terms on the right- and lefthand side of Eq. 3.15 respsectively). Recalling from the discussion on binning (§4.1.2) that we choose N_{side} such that our pixels are small compared to the PSF, we use a simple central value approximation to evaluate the righthand side terms. The lefthand side terms, the integrals over the ROI for a given

[10]The differential in the log term is not problematic, as it is multiplied by a second differential, and $\lim_{d\theta^2 \to 0} d\theta^2 \log d\theta^2 = 0$.
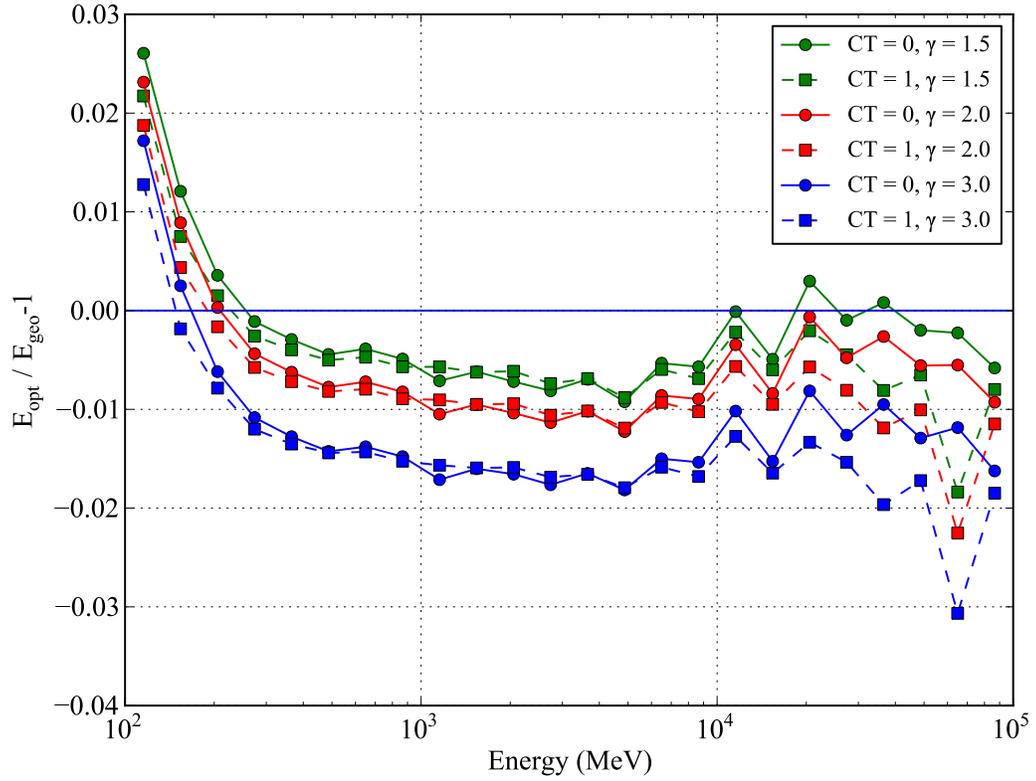
Figure 4.4: Comparison of the geometric mean ($E_{\mathrm{geo}} \equiv \sqrt{E_{\max} \times E_{\min}}$) of an energy band with the optimal energy E_{opt} as estimated by the log likelihood method described in the text. The binning is uniform in logarithmic energy with 8 bins per decade and separate bands for front- and back-converting events are shown. The relative difference of E_{opt} from E_{geo} for three different power law source spectra with photon indices $\gamma = 1.5, 2.0, 3.0$ is shown.

energy band, become

$$r(\vec{\lambda}) = \left[\int d\Omega f_{\text{wpsf}}(\vec{\Omega}; \vec{\Omega}_0, E_{\text{opt}}) \right] \left[\int dE \, \mathcal{F}(E, \vec{\lambda}) \, \epsilon(E, \vec{\Omega}_0) \right] \tag{4.24}$$

$$\equiv \left[\int d\Omega f_{\text{wpsf}}(\vec{\Omega}; \vec{\Omega}_0, E_{\text{opt}}) \right] N(\vec{\lambda}) \tag{4.25}$$

$$\equiv O(E_{\text{opt}}, \vec{\Omega}_0) \, N(\vec{\lambda}). \tag{4.26}$$

Here, we have defined $O(E_{\text{opt}}, \vec{\Omega}_0)$, the *overlap integral*. The overlap integral is simply the integral of the PSF over the ROI for a particular source. It must be between 0 and 1, and it depends only on the source position, the PSF for the energy band as obtained through the method discussed above, and the radius of the ROI. The payoff of the decoupling of the energy and position integrals we achieved by selecting E_{opt} is evident when we consider the log likelihood for the energy band,

$$\mathcal{L}(\vec{\lambda}) = -\sum_{j}^{N_s} O(E_{\text{opt}}, \vec{\Omega}_j) \, N_j(\vec{\lambda}) + \sum_{i}^{N_{bins}} N_i \log \sum_{j}^{N_s} f_{\text{wpsf}}(\vec{\Omega}_i; \vec{\Omega}_j, E_{\text{opt}}) N_j(\vec{\lambda}), \tag{4.27}$$

where we see that the dependence on $\vec{\lambda}$ comes only through the $N(\vec{\lambda})$ terms. Since only the $N(\vec{\lambda})$ terms change during the likelihood maximization, we are free to pre-compute once and for all the overlap integrals *and* the data-pixel contributions for each source. We have already discussed the latter calculation; we now proceed with a description of the terms on the lefthand side of Eq. 4.27.

Evaluating the Overlap Integrals

While we make the central-value approximation to evaluate the PSF integral over the data pixels, the overlap integrals are more complex. First, we assume that the ROI is circular. Through this assumption and the azimuthal symmetry (for sufficiently long integrations) of the PSF, we can reduce the 2-d integral to a quadrature. Let the ROI center be denoted $\vec{\Omega}_0$ and the source position be $\vec{\Omega}_s$. By the symmetries of the ROI and PSF, the only relevant quantities are the arclength separating the source and the ROI center, denoted

$\delta \equiv \arccos \vec{\Omega}_0 \cdot \vec{\Omega}_s$, and the radius of the ROI, R. In terms of these quantities,

$$O(E_{\text{opt}}, \delta, R) = \int_{ROI} d\Omega \, f_{\text{wpsf}}(\vec{\Omega}; \vec{\Omega}_0, E_{\text{opt}}) = \iint_{ROI} d\phi \, \frac{d\theta^2}{2} \frac{1}{2\pi\sigma^2} f_{wk}\left(\frac{\theta^2}{2\pi\sigma^2}\right) \qquad (4.28)$$

$$= \int_{ROI} \frac{d\phi}{\pi} \int_{u_1(\phi)}^{u_2(\phi)} du \, f_{wk}(u) = \int_{ROI} \frac{d\phi}{\pi} F_{wk}[u_2(\phi)] - F_{wk}[u_1(\phi)], \qquad (4.29)$$

where we have introduced the explicit weighted-sum of King functions for the PSF. (As discussed in §3.4.1, some versions of the LAT IRF express the PSF as the sum of two King functions. The overlap integral is linear in the King function, so the derivation presented here carries over directly. Likewise, F_{wk} is an exposure-weighted sum, subject to the same linear arguments.) F_{wk} denotes the (analytic) cumulative distribution of the King function, and we have achieved reduction to quadrature. All that remains is to determine the bounds of integration, i.e., $u_1(\phi)$, $u_2(\phi)$, and the bounds of the azimuthal integral.

For the case of a point source located within the ROI boundary, $\delta < R$, $u_1(\phi) = 0$. We place the x-axis on the diameter extending through the $\vec{\Omega}$ and the ROI center. The integrand is clearly symmetric about the x-axis, so

$$O(E_{\text{opt}}, \delta < R, R) = \frac{2}{\pi} \int_0^\pi d\phi \, F_{wk}[u_2(\phi)]. \qquad (4.30)$$

$u_2(\phi)$ can be determined from plane geometry[11]:

$$\sigma\sqrt{2u_2(\phi)} = \sqrt{R^2 - [\delta\sin(\phi)]^2} - \delta\cos(\phi), \qquad (4.31)$$

where σ stands in for the σ parameter for each relevant King function.

For an exterior source, $\delta > R$, both bounds in u are non-trivial. We again place the x-axis on the diameter joining $\vec{\Omega}_-$ and the ROI center and again note the reflection symmetry. The ϕ integral extends from the x-axis to the tangent joining $\vec{\Omega}_0$ and the boundary of the ROI. That is,

$$O(E_{\text{opt}}, \delta > R, R) = \frac{2}{\pi} \int_0^{\arccos(R/\delta)} d\phi \, F_{wk}[u_2(\phi)] - F_{wk}[u_1(\phi)]. \qquad (4.32)$$

---

[11]The assumption, of course, is that the ROI is sufficiently small that we may approximate the space as $\mathbb{R}^2$.

The bounds of integration are again determined from simple geometry:

$$\sigma\sqrt{2u_1(\phi)} = \delta\cos(\phi) - \sqrt{R^2 - [\delta\sin(\phi)]^2} \tag{4.33}$$

$$\sigma\sqrt{2u_2(\phi)} = \delta\cos(\phi) + \sqrt{R^2 - [\delta\sin(\phi)]^2} \tag{4.34}$$

Evaluating the Source Counts

Recall that we defined $N(\vec{\lambda}) = \int dE\, \mathcal{F}(E,\vec{\lambda})\, \epsilon(E,\vec{\Omega}_0) \equiv \int dE\, f(E)$. We evaluate this integral with a composite Simpson's rule quadrature over logarithmic energy. If the energy range of the band runs from e_a to e_b, then

$$N(\vec{\lambda}) \approx \frac{\log e_b/e_a}{3N_s} \left[e_a f(e_a) + e_b f(e_b) + \sum_{i=1}^{N_s-1} (2 + 2\mathcal{I}_{\in 2\mathbb{Z}+1}(i)) \times e_a \delta^{2i-1} f(e_a\delta^{2i-1}) \right], \tag{4.35}$$

where $\delta \equiv (e_b/e_a)^{1/N_s}$, $\mathcal{I}(i)$ is 1 (0) for odd (even) integers, and N_s (an even integer) is the number of intervals into which we divide the band. N_s moderates the tradeoff between speed and accuracy. In Figure 4.2.2, we explore the accuracy of the integral of a power law source over the (typically) lowest energy band in a *pointlike* analysis. The effective area here is rapidly rising and bumpy (see Figure 3.4), making this integral a good test case. We see $N_s = 8$ gives an error $< 1\%$. Since the evaluation of these integrals is not currently a serious bottleneck in the *pointlike* code, we adopt the conservative choice $N_s = 16$.

With a quadrature rule in hand, we can make one additional approximation for speed. We cache the specified energies E_i and factors (the 1s, 2s, and 4s in Eq. 4.35) for each band. Further, since the exposure varies relatively slowly in position, over a typical ($10°$) ROI we can approximate the exposure for a particular source by the exposure at the center, making a first order correction by normalizing to the exposure value at E_{geo}. We can then cache $w_{si} \equiv k_i\, E_i\, \epsilon(E_i, \vec{\Omega}_0)$, with k_i the Simpson's rule factor, and evalute the integral, for an arbitrary source at position $\vec{\Omega}$, as

$$\frac{\epsilon(E_{\text{geo}}, \vec{\Omega})}{\epsilon(E_{\text{geo}}, \vec{\Omega}_0)} \sum_{i=1}^{N_s+1} w_{si}\mathcal{F}(E_i, \vec{\lambda}). \tag{4.36}$$

Evaluating this expression is extremely fast, and the exposure approximation introduces negligible ($<< 1\%$) error (see Figure 4.2.2).
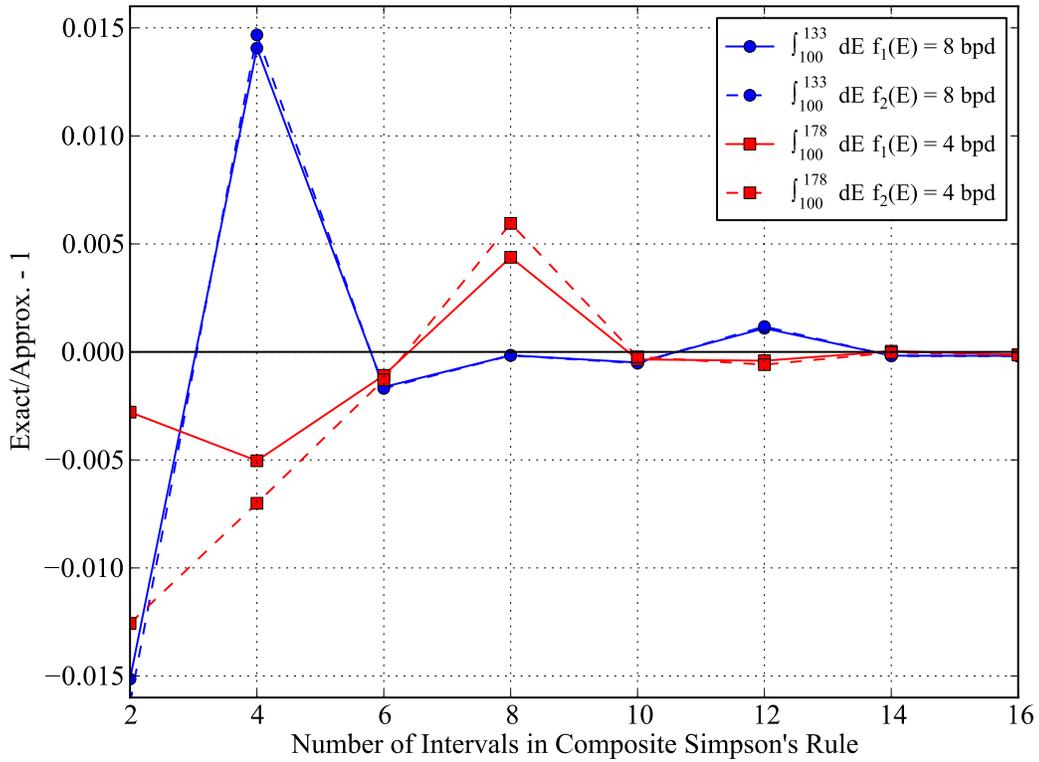
Figure 4.5: Two power law spectra ($f_1(E) \propto E^{-1.5}$ and $f_2(E) \propto E^{-3.0}$) integrated over the most problematic energy band in a typical analysis, 100-133 MeV in 8 bin-per-decade mode and 100-178 MeV in 4 bin-per-decade mode. The relative difference with the exact integral (as determined by numerical integration to machine precision) is shown as a function of the number of intervals used in the composite Simpson's rule.
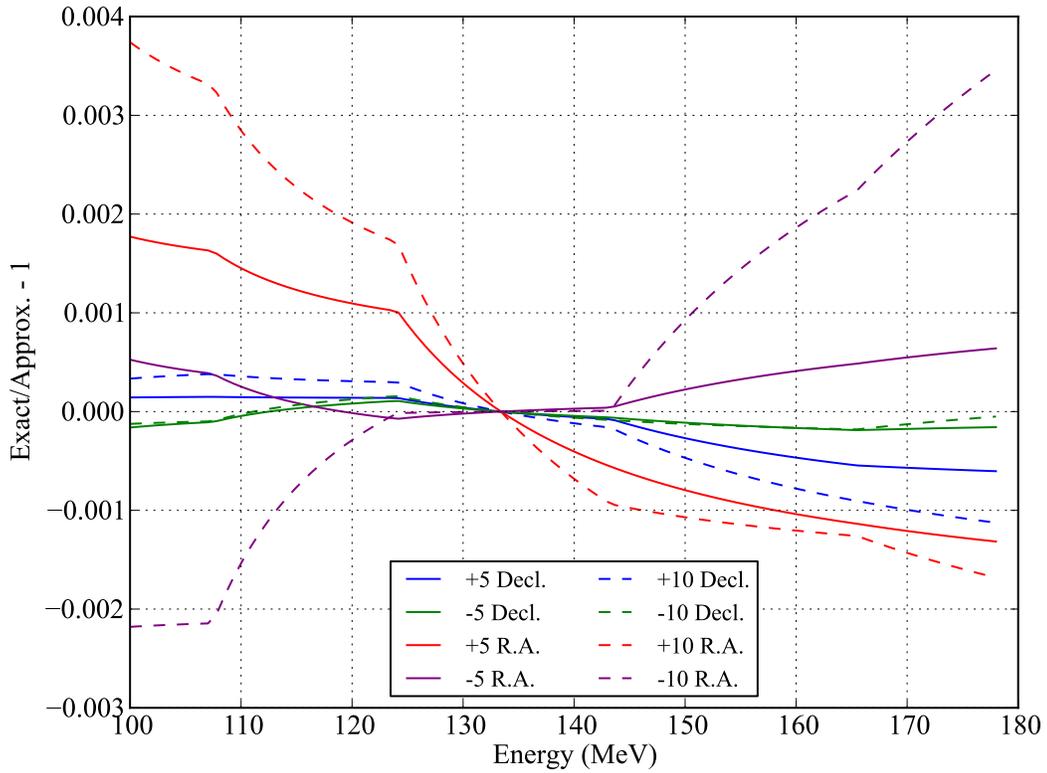
Figure 4.6: The exposure variation over a typical ROI (here, located at the position of the Vela PSR.) The approximated exposure assumes the same energy dependence as that at the ROI center and corrects only for the normalization at the geometric mean energy of the band. The exact value accounts for the spatial variation at all energies. Relative differences are shown for point sources at a modest (5°) and extreme (10°) separation from the ROI center. The band chosen, $100 < E/\mathrm{MeV} < 178$, has rapidly-varying effective area, making this a worst case. The error is $< 0.5\%$.

4.2.3 Point Source Summary

Our goal in this section was to calculate the contribution of a point source to a particular energy band. This task entailed evaluating the point source model (convolved with the IRF) both over the sparse data pixels and over an entire spatial aperture. We made approximations that allowed us to define a single PSF for the energy band, $f_{wpsf}(E_{\text{opt}})$, in turn allowing a factorization of the source rate integral into a spatial component and an energy component. We reduced the spatial integral over the aperture to a quadrature, and we outlined a rapid method for performing the energy integral for each source. Taken together, these approximations generally retain percent-level accuracy but greatly speed up computation of the contribution of point sources to the likelihood.

4.3 Diffuse Sources Rates

We now come to the calculation of the rates of diffuse sources, those that have resolvable angular extension in the LAT data. We return to Eqs. 3.10 and 3.13 which outline the integrations to be performed. The argument parallels that for point sources. We again begin by making the assumption that the diffuse sources are stationary[12]. As for point sources, we introduce the exposure and exposure-weighted PSF to arrive at the following expression for the expected rate per unit solid angle of a diffuse source for a given energy band:

$$r(\vec{\Omega}') = \int dE \int d\Omega \, \mathcal{F}(E, \vec{\Omega}; \vec{\lambda}) \, \epsilon(E, \vec{\Omega}) \, f_{\text{wpsf}}(\vec{\Omega}'; \vec{\Omega}, E). \qquad (4.37)$$

This expression is similar to that for point sources, Eq. 4.10 but with an additional complication, the convolution of the diffuse source count rate per unit solid angle with the PSF. In the following, we outline the evaluation of the integrals over energy and position.

The strategy differs slightly from point sources. Typically, we deal with only a couple of diffuse sources, the Galactic diffuse background from cosmic rays and the (approximately) isotropic background from unresolved sources at cosmological distances, irreducible charged

---

[12]This is of course perfectly true for celestial diffuse sources. Photons from the Earth's limb, and from irreducible charged particle background, have some time dependence based on the S/C precessional phase and the active rocking profile. These effects are certainly second order.

particle background, and contamination from the Earth's limb. With only two sources compared to tens or even hundreds of point sources, we can be more lavish with computational resources. Furthermore, all but the brightest point sources are utterly dominated by the diffuse background at low energy, particularly in the Galactic plane. To avoid bias, it is important to model the diffuse sources with as much accuracy as possible.

4.3.1 Energy Integral for Diffuse Sources

As with point sources, we adopt a composite Simpson's rule for the integral over energy. Adopting the same notation,

$$r(\vec{\Omega}') \approx \sum_{i=1}^{N_s+1} w_i \int d\Omega \, \mathcal{F}(E_i, \vec{\Omega}; \vec{\lambda}) \, \epsilon(E_i, \vec{\Omega}) \, f_{\mathrm{wpsf}}(\vec{\Omega}'; \vec{\Omega}, E_i) \equiv \sum_{i=1}^{N_s+1} w_i \, r(\vec{\Omega}', E_i), \qquad (4.38)$$

where in this case w_i is the product of the Simpson's rule factor and E_i. The expression $\int d\Omega \, \mathcal{F}(E_i, t', \vec{\Omega}; \vec{\lambda}) \, \epsilon(E_i, \vec{\Omega}) \, f_{\mathrm{wpsf}}(\vec{\Omega}'; \vec{\Omega}, E_i)$ is extremely costly to evaluate, so we are more cautious in selecting N_s. Indeed, since only the low-energy bands, with their bumpy and rapid rise in the effective area, require care, we adopt $N_s = 8$ for bands with $E < 200\mathrm{MeV}$ and $N_s = 4$ for all remaining bands.

Unlike for point sources, we do not attempt to pull the PSF out of the energy integral for two important reasons. First, in keeping with our remarks above, we want to evaluate the integral as accurately as is feasible. Second, unlike point sources, the spatial morphology of diffuse sources can and does change with energy. Such morphology shifts within an energy band are likely to be small but are in any case explicitly accounted for in our formulation here.

4.3.2 Spatial Integral (Convolution) for Diffuse Sources

We now have to evaluate

$$r(\vec{\Omega}', E) = \int d\Omega \, \mathcal{F}(E, \vec{\Omega}; \vec{\lambda}) \, \epsilon(E, \vec{\Omega}) \, f_{\mathrm{wpsf}}(\vec{\Omega}'; \vec{\Omega}, E). \qquad (4.39)$$

The primary source of interest is the Galactic diffuse background. The construction of the model is beyond the scope of this work, but uses methods similar to those employed in the

analysis of [14]. It is typically structured as a *mapcube*, a pixelized intensity map of the sky stored at a series of energy planes. We construct the continuous function $\mathcal{F}(E, \vec{\Omega}; \vec{\lambda})$ by first selecting the two image planes bounding E, determining the spatial intensity in the bounding image planes via bilinear interpolation of the four pixels closes to $\vec{\Omega}$, and finally logarithmically interpolating these two values in energy. The isotropic background is typically tabulated, and we simply interpolate the table. The isotropic model acquires spatial variation through multiplication with the exposure. We write a general model as $\mathcal{D}(E, \vec{\Omega}) \equiv \mathcal{F}(E, \vec{\Omega}; \vec{\lambda}) \, \epsilon(E, \vec{\Omega})$. Since the PSF only depends on the *difference* between the true and reconstructed positions, the spatial density is

$$r(\vec{\Omega}') = \int d\Omega \, \mathcal{D}(E, \vec{\Omega}) \, f_{\mathrm{wpsf}}(E, \vec{\Omega}' - \vec{\Omega}), \qquad (4.40)$$

readily seen to be a two-dimensional convolution.

Recall that a convolution of two functions, f and g, is defined in one dimension as

$$(f * g)(x') = \int\limits_{-\infty}^{+\infty} dx \, f(x) \, g(x' - x). \qquad (4.41)$$

The convolution arises in probability theory as the probability density function for the sum of two statistically independent variables with probability density functions f and g. (The integral is over all combinations of values drawn from f and g whose sum has the correct value.) Now, suppose f or g is concentrated near the origin. The convolution is then well approximated with some cutoff:

$$(f * g)(x') \approx \int\limits_{-X}^{+X} dx \, f(x) \, g(x' - x). \qquad (4.42)$$

We can evalute such an integral by sampling it at N uniform points between $-X$ and $+X$ and applying some quadrature rule. If we use N points in the quadrature, then all of the information in the convolution can also be contained with N points. Thus, a full evaluation of the convolution over its support requires $\mathcal{O}(N^2)$ operations.

Similar arguments apply to the evaluation of a discrete Fourier transform (DFT) at its independent frequencies. However, the use of the Fast Fourier Transform (FFT) algorithm[37]

allows evaluation of all independent frequencies of a DFT in only $\mathcal{O}(N \log N)$ operations. This development can be applied to convolutions via the Convolution Theorem:

$$\mathcal{F}(f * g) = \mathcal{F}(f) \times \mathcal{F}(g), \tag{4.43}$$

i.e., the Fourier transform of the convolution of f and g is equal to the product of the Fourier transforms of f and g. The N quadrature sampling points map to N independent frequencies in a DFT. Thus, by evaluating the FFTs of f and g, multipyling, and taking the inverse FFT, we can calculate the convolution of f and g with $\mathcal{O}(N \log N)$ complexity.

We now apply these arguments to the evaluation the convolution of a diffuse source with the PSF. To apply a 2D FFT, we need $\mathcal{D}(E, \vec{\Omega})$ and $f_{\mathrm{wpsf}}(E, \vec{\Omega})$ evaluated on a regular grid. The diffuse map, of course, is a curved space, so in making such a projection we make an approximation. However, since our grids are typically 10s of degrees on a side, a projection does not introduce too much distortion. We assume for now that we are working on the equator in the Galactic coordinate system. We use the plate carrée projection such that the Cartesian coordinates of the projection x and y are simply l and b. (The carrée projection is approximately equidistant near the equator; this is important, since the PSF depends only on the angular separation of two points.) We cut off the integral to a square grid with a side length of 2Δ. Then, the convolution becomes

$$r(l', b', E) = \iint d\Omega \, \mathcal{D}(E, l, b) \, f_{\mathrm{wpsf}}(E, l' - l, b' - b) \tag{4.44}$$

$$\approx \int_{-\Delta}^{+\Delta} dl \int_{-\Delta}^{+\Delta} db \, \mathcal{D}(E, l, b) \, f_{\mathrm{wpsf}}(E, (l' - l)^2 + (b' - b)^2) \tag{4.45}$$

$$\approx \sum_{i=-N}^{i=+N} \sum_{j=-N}^{j=+N} \mathcal{D}(E, i\Delta, j\Delta) \, f_{\mathrm{wpsf}}(E, (l' - i\Delta)^2 + (b' - j\Delta)^2) \tag{4.46}$$

The prescription is now clear: evaluate $\mathcal{D}$ and f_{wpsf} on the $N \times N$ grid, perform a 2D FFT on each, multiply, and invert, and store the final $N \times N$ image plane containing the convolved diffuse model. By interpolating the pixels of this image, we can evaluate the convolved diffuse model for any $\vec{\Omega}$ within the grid, i.e., we can compute the rate in Eq. 4.38 trivially. We note, importantly, that we normalize f_{wpsf} to the grid, with the net effect that "diffuse photons are conserved". This adjustment inevitably leads to some edge effects, as

at low energies some photons are scattered into/out of the ROI. We thus choose a grid size to minimize these effects on the model in the ROI itself. Currently, for each band, we calculate the 95% containment radius of the PSF and require the grid be $2 \times (R_{roi} + R_{95})$ on a side.

We note a few technical details. First, we are effectively rebinning the model of the diffuse source onto the uniform grid. Therefore, the pixel size of the uniform grid should be sufficiently small to avoid artifacts. The *gll_iem_v02* model[13] of the Galactic diffuse uses $0.5°$ pixels, so a sufficiently fine uniform grid size is $0.25°$. Second, the above derivation relies on a convolution at the equator, where the carrée projection is relatively distortion free and, more importantly, equidistant. In order to use this approach generally, we perform the following procedure for an ROI centered on (l, b):

1. Construct a uniform grid centered at $(l, 0)$.

2. Right-handedly rotate each point on the grid by $b°$ about the vector pointing to $(l - 90°, 0)$ to get (l_r, b_r).

3. Evaluate $\mathcal{D}(E, l_r, b_r)$ for each point on the rotated grid.

4. Perform the FFTs and convolution.

To evaluate the convolved background model at a particular point, say (l', b'), the process is reversed: rotate by $-b°$ about $(l - 90°, 0)$ and then linearly interpolate the grid values. (Another benefit of using the plate carée projection is the ease of indexing and interpolation, since the indices map directly to geographic coordinates with only arithmetic operations.)

4.4   Calculating and Maximizing the Likelihood

We have now described in some detail how the rates for point source and diffuse sources appearing in the likelihood, Eq. 3.13, are calculated in the *pointlike* framework. Particularly, we have demonstrated how to calculate rates for *bins in energy and event conversion*

---

[13]http://fermi.gsfc.nasa.gov/ssc/data/access/lat/BackgroundModels.html

type. We formalize this procedure in *pointlike* with the a *band,* an object (in the computer science sense) that associates data and source rates for a particularly energy range and conversion type and performs a series of tasks to link the two and facilitate computation of the likelihood.

4.4.1   The Band Object

In terms of data, each *Band* maintains an array storing the positions $\vec{\Omega}_i$ and counts n_i of the occupied pixels in the ROI that fall within its energy bounds.

Next, the *Band* is responsible for pre-computing some quantities. It causes E_{opt} (Eq. 4.23) for its energy range/conversion type to be determined and, from this, determines and stores the proper (approximate) PSF $f_{wpsf}(E_{\mathrm{opt}})$ introduced in Eq. 4.11. It determines and stores the Simpson's energies E_i and Simpson's weights, w_i, from Eq. 4.36. For each point source, referenced by index j, a *Band* computes and stores, the overlap integral, $\mathcal{O}_j$ (Eq. 4.24), and the exposure correction, $\alpha_j \equiv \epsilon(1\,\mathrm{GeV}, \vec{\Omega}_j)/\epsilon(1\,\mathrm{GeV}, \vec{\Omega}_0)$ appearing in Eq. 4.36. Using f_{wpsf}, it evaluates f_{ij}, the (normalized) contribution of the jth source to the ith pixel, i.e., $f_{ij} \equiv f_{\mathrm{wpsf}}(E_{\mathrm{opt}}, \vec{\Omega}_i; \vec{\Omega}_j)$, appearing in Eq. 4.27.

To actually evaluate the log likelihood, we provide a $\vec{\lambda}$. Let $\vec{\lambda}_j$ be the subspace of $\vec{\lambda}$ that parameterizes point source j. The *Band* then evaluates the source rate using its cached Simpson's rule information:

$$N_j(\vec{\lambda}_j) \approx \alpha_j \int dE\, \mathcal{F}_j(E, \vec{\lambda}_j)\, \epsilon(E, \vec{\Omega}_0) \approx \alpha_j \sum_{i=1}^{N_s+1} w_{si}\, \mathcal{F}_j(E_i, \vec{\lambda}_j) \tag{4.47}$$

$$\equiv \alpha_j \times \vec{w}_s \cdot \mathcal{F}_j(\vec{E}, \vec{\lambda}_j). \tag{4.48}$$

Ignoring diffuse sources for the moment, the log likelihood for the *Band* is then

$$\log \mathcal{L}(\vec{\lambda}) = -\left[\sum_{j=1}^{N_{\mathrm{sources}}} \mathcal{O}_j\, N_j(\vec{\lambda}_j)\right] + \left[\sum_{i=1}^{N_{\mathrm{pix}}} n_i \log \sum_{j=1}^{N_{\mathrm{sources}}} N_j(\vec{\lambda}_j)\, f_{ij}\right] \tag{4.49}$$

$$\equiv -\vec{\mathcal{O}} \cdot \vec{N}(\vec{\lambda}) + \vec{n} \cdot \log\left[\mathbf{f} \cdot \vec{N}(\vec{\lambda})\right] \tag{4.50}$$

We have written the above equations for the source rate and log likelihood both as sums and as vectors to make clear (a) what is being done operationally and (b) what is being

done internally. It is easy to appreciate, especially in the vector notation, the computational savings gained by factoring the spatial and energy terms. When we update $\vec{\lambda}$, all that is required to update the likelihood is a handful of vector arithmetic operations! The vector notation is also schematic for how the computation is carried out in the Python source code. Despite Python's interpreted nature, it is possible with $numpy$[14] to run vectorized code at nearly the same speed as compiled code[15]. We thus strive to formulate vectorized implementations whenever possible. (To avoid confusion, we note that, from left to right, the three inner products above are in "source space", in "pixel space", and in "source space", respectively.) A second benefit to this notation is it makes determination of the gradient of the log likelihood, $d\log\mathcal{L}(\vec{\lambda})/d\vec{\lambda}$, more tractable, as we shall see below.

Band objects also manage the likelihood calculation for diffuse sources in a manner similar to that described above. Recall that, after constructing an image of the convolved diffuse model, we can evaluate the quantity

$$r(\vec{\Omega}', E) = \int d\Omega\, \mathcal{F}(E, \vec{\Omega}; \vec{\lambda})\, \epsilon(E, \vec{\Omega})\, f_{\text{wpsf}}(\vec{\Omega}'; \vec{\Omega}, E) \qquad (4.51)$$

for a given diffuse source by interpolating from the convolved model values evaluated on our uniform grid. To integrate the rate over the ROI, we simply average all pixels on the uniform grid lying within the ROI, i.e.,

$$D(E, \vec{\lambda}) \equiv N(E, \vec{\lambda}) \int d\Omega'\, r(\vec{\Omega}', E) \approx \frac{\pi R_{ROI}}{\sum\limits_{i=0}^{N_{grid}} \mathcal{I}_{i\in ROI}} \sum_{i=0}^{N_{grid}} \mathcal{I}_{i\in ROI}\, d_i, \qquad (4.52)$$

where d_i is the the model value for the ith grid pixel and we use the indicator function $\mathcal{I}$ to select only grid pixels lying within a circle of radius R_{ROI}. $N(\vec{\lambda})$ is a scaling model which evaluates to 1 for all energies at the "default" parameter values. We integrate both $D(E)$ and $r(\vec{\Omega}', E)$ over the band using Simpson's rule analogously to the procedure for point

---

[14]http://numpy.scipy.org

[15]It is also possible to extend the Python language with functions written in C, but *numpy* typically suffices.

sources and define

$$D \equiv \int dE \, D(E), \tag{4.53}$$

$$d(\vec{\Omega}, \vec{\lambda}) \equiv \int dE \, r(\vec{\Omega}, E) N(E, \vec{\lambda}). \tag{4.54}$$

By using fixed Simpson's rule points, we are free to pre-compute and store the quantities $r(\vec{\Omega}, E)$ and $\int d\Omega' \, r(\vec{\Omega}', E)$, and updates from the likelihood will only require a sum with a few terms, viz. the Simpson's rule sum. With these quantities, we now have the full *Band* likelihood including point sources and diffuse sources:

$$\log \mathcal{L}(\vec{\lambda}) = - \left[\sum_{j=1}^{N_{\mathrm{ps}}} \mathcal{O}_j \, N_j(\vec{\lambda}_j) \right] - \left[\sum_{j=1}^{N_{\mathrm{ds}}} D_j(\vec{\lambda}) \right] + \left[\sum_{i=1}^{N_{\mathrm{pix}}} n_i \log \sum_{j=1}^{N_{\mathrm{ps}}} N_j(\vec{\lambda}_j) \, f_{ij} + \sum_{j=1}^{N_{\mathrm{ds}}} d_{ij}(\vec{\lambda}) \right] \tag{4.55}$$

$$\equiv -\vec{\mathcal{O}} \cdot \vec{N}(\vec{\lambda}) - \left[\sum_{j=1}^{N_{\mathrm{ds}}} D_j(\vec{\lambda}) \right] + \vec{n} \cdot \log \left[\mathbf{f} \cdot \vec{N}(\vec{\lambda}) + \sum_{j=1}^{N_{\mathrm{ds}}} \vec{d}(\vec{\lambda}) \right], \tag{4.56}$$

where N_{ps} (N_{ds}) indicates the number of point (diffuse) sources. We leave the sum over diffuse source explicit since there are typically only a few such sources modeled.

4.4.2 Calculating the Likelihood Gradient

We preface discussion on likelihood fitting with an important ingredient, an "analytic" calculation of the gradient—as opposed to finite differences—that provides significantly faster maximization of the log likelihood and determination of its curvature. The calculation of $\partial \log \mathcal{L}(\vec{\lambda})/\partial \lambda_i$ is complicated by the many-to-one mapping of parameters to sources. We introduce the auxiliary matrix $\mathbf{M}$ defined by $M_{ij} = \partial N_j(\vec{\lambda})/\partial \lambda_j$. This $N_{\mathrm{source}} \times N_{param}$ matrix will only have nonzero entries where the parameter in the denominator matches a parameter of the source in the numerator. Then, ignoring diffuse sources,

$$\vec{\nabla} \log \mathcal{L}(\vec{\lambda}) = -\vec{\mathcal{O}} \cdot \mathbf{M} + \left[\frac{\vec{n}}{\mathbf{f} \cdot \vec{N}(\vec{\lambda})} \right] \cdot [\mathbf{f} \cdot \mathbf{M}], \tag{4.57}$$

where by $\vec{n}/(\mathbf{f} \cdot \vec{N}(\vec{\lambda}))$ we mean elementwise division of the numerator by the denominator. We note that the denominator of this term has already been determined in the log likelihood

calculation. Thus, in practice, we first build up $\mathbf{M}$, requiring N_{param} Simpson's rule quadratures and some bookkeeping, and the remainder of the gradient can be computed with a few matrix operations. Terms arising from the diffuse sources can be incorporated analogously. This method should be compared to $\geq 2N_{param}$ likelihood evaluations required to estimate $\vec{\nabla} \log \mathcal{L}(\vec{\lambda})$ using finite differences; the "analytic" approach is also free from numerical errors resulting from inappropriate step size choice.

4.4.3   Spectral Models

Up to now, we have been fairly abstract about $\mathcal{F}(E, \vec{\lambda})$, the source flux. A variety of spectral models are available in *pointlike*, including those in most common use, a power law and a power law + exponential cutoff. Using the computer science principle of inheritance, particular spectral models inherit from a single base class. The base class manages all common tasks, while the subclasses implement the methods particular to the spectral model, e.g., the density, $dN/dE(E, \vec{\lambda})$, and the gradient, $\partial(dN/dE(E, \vec{\lambda}))/\partial\vec{\lambda}$.

Parameters for spectral models are almost universally positive definite, and negative values are undefined or unphysical. To prevent the fitting algorithm from attempting these disallowed values, we perform a logarithmic transformation of all parameters internally.

4.4.4   Maximizing the Likelihood and Estimating Errors

The default fitter for *pointlike* is the implementation of the BFGS[70] algorithm in *scipy*[16] via *fmin_bfgs*. This algorithm uses the computation scheme for the log likelihood and its gradient outlined above. If a spectral model is included for which it is difficult to implement $\partial \log \mathcal{L}(\vec{\lambda})/\partial\lambda_i$, then the downhill simplex algorithm—which makes use of the log likelihood only—implemented in *scipy* via *fmin* is also available.

We estimate the errors from the information matrix, the inverse of the hessian of the log

---

[16]http://www.scipy.org

likelihood function. That is, the covariance matrix is estimated by

$$\sigma^2_{\lambda_i \lambda_j} = \mathcal{H}^{-1}_{\lambda_i \lambda_j} \tag{4.58}$$

$$\mathcal{H}_{\lambda_i \lambda_j} = -\frac{\partial^2 \log \mathcal{L}}{\partial \lambda_i \partial \lambda_j} \tag{4.59}$$

If the analytic gradient is available, we employ a first-order finite difference scheme, while in the absence of a gradient we perform second-order finite differencing of the log likelihood function. In both cases, we use an iterative process that attempts to find an ideal step size for estimation of the curvature, i.e., one for which the change in the log likelihood is of order unity.

4.4.5 Source Localization

While we have concentrated on spectral parameters, the source position is also a model parameter and we can find its MLE in a similar process. Whereas the approximations we have made allow us to fit spectral parameters for point sources and diffuse sources without re-evaluating model values for pixels (e.g., $\mathbf{f}$ in Eq. 4.55), this is manifestly not the case for source localization since we must obviously re-evaluate the row of $\mathbf{f}$ corresponding to the source being localized as its position is updated during ML fitting. To speed up this process, we adopt an iterative process for source localization. An initial spectral fit is performed, and then the spectral parameters are taken as given while the position of the source in question is varied to maximize a two-dimensional likelihood function (i.e., two angular coordinates). The improved position can be used in a second spectral fit, and so forth until convergence. The only drawback to this approach is a potential slight underestimate of the positional uncertainty since correlation with spectral parameters is not taken into account. However, tests of the localization algorithm presented in Chapter 5 indicate such an effect is negligible.

Let $\vec{b} \equiv \mathbf{f} \cdot \vec{N}(\vec{\lambda}) + \sum_{j=1}^{N_{\mathrm{ds}}} \vec{d}(\vec{\lambda}) - \vec{f}_s \, N_s(\vec{\lambda})$ (see Eq. 4.55), i.e., the rate in each pixel for all sources under the best-fit spectral model with the rate for one point source, labeled "s", subtracted. This quantity is constant when the position of "s" changes. The log likelihood is then

$$\log \mathcal{L}(\vec{\Omega}) = -\mathcal{O}_s(\vec{\Omega}) \, N_s + \vec{n} \cdot \log \left[\vec{b} + \vec{f}_s(\vec{\Omega}) \, N_s \right] . \tag{4.60}$$

To maximize the function, we fit a quadratic form to the log likelihood surface via least squares and iterate until the position change is sufficiently small. The final quadratic form measures the likelihood surface and provides an estimate for the uncertainty in the two position parameters.

Finally, we note that this formulation is independent of the spectral model used for source "s", and while we typically employ one of the usual spectral models, e.g., a power law, we can also use a "model-independent" approach outlined below.

4.4.6   Model-independent Fits

In the previous sections, we have concentrated on the *Band*. Of course, when fitting a (broadband) spectral model, we accumulate the log likelihood contributions from each *Band* to compute the total likelihood. However, we may also be interested in the likelihood from a single band. The typical use case is assessing the shape of the spectral energy density of a source, e.g. to look for deviations from the broadband model. In this procedure, we model all sources initially with a broadband spectral model and maximize the likelihood to arrive at a good spectral model for the ROI. We then freeze the parameters for all sources save source "s". We assume that, *within the band*, the source has a power law spectrum with $\Gamma = 2.0$. (Since the bands are narrow, there is little dependence on the particular slope chosen.) Then, the log likelihood for the source flux *within the band* is given by

$$\mathcal{L}(\mathcal{F}_s) = -\mathcal{O}_s \frac{N_s}{\mathcal{F}_{0s}} \mathcal{F}_s \vec{n} \cdot \log \left[\vec{b} + \vec{f}_s(\vec{\Omega}) \frac{N_s}{\mathcal{F}_{0s}} \mathcal{F}_s \right], \tag{4.61}$$

with $\mathcal{F}_{0s}$ the flux calculated for source "s" in the initial spectral fit. This single-dimension function is easily extremized by differentiating and finding the root of the resulting equation. Thus, $\mathcal{F}_s$ can be rapidly estimated in each band, yielding a band-by-band estimate for the spectral energy density for the source. These estimates can be used for localizing sources as outlined in the previous section, or for generating plots of the spectral energy density. For the latter application, we typically use the joint likelihood for the *Band* objects for front-converting and back-converting events at the same energy.

In this prescription, the only sensitivity of the final band-by-band fits to the Ansatz broadband model comes from the other sources. Provided that model is not too far wrong—

or that source "s" is not too bright—we expect the initial model to have little effect on the band-by-band estimates.

4.5   Summary

In this chapter, we laid out the major design principles and detailed the implementation of *pointlike*, a full-featured maximum likelihood analysis package. We first outlined our method of binning the data with pixels that scale with the PSF to achieve good performance at low energy and good accuracy at high energy. We next began the "heavy lifting" of evaluating the source rates folded through the *Fermi*-LAT IRF by distinguishing point sources and diffuse sources. For point sources, we defined an effective PSF for each band by estimating an optimal energy at which to evaluate the King function parameters, and we used this PSF to determine point source contributions to data pixels and in the overlap integrals, which we reduced to quadrature. For diffuse sources, we described a method for evaluating the diffuse model on a locally flat grid and then applying the Convolution Theorem to quickly evaluate the convolution with a two-dimensional Fast Fourier Transform. Finally, we described how these ingredients are unified in a *Band* object to facilitate calculation of the likelihood and its gradient.

Along the way, we attempted to validate each piece of the calculation and verify that it met our goal of percent-level accuracy. In the next chapter, we test the entire *pointlike* package for actual science tasks, viz. estimating the spectral and positional parameters for sources under a variety of conditions.

Chapter 5

THE *POINTLIKE* PACKAGE: VALIDATION

Although in Chapter 4 the individual portions of *pointlike* have been considered and the efficacy of the various approximations tested, the most important check is a holistic one— the determination of spectral parameters from realistic data. Although some bright γ-ray sources can be considered standard candles, *Fermi*-LAT is the first HE γ-ray observatory with good sensitivity above 1 GeV, and comparison to past experiments can only serve as a sanity check[5]. Since we are concerned with percent-level effects, we instead use data generated with Monte Carlo simulations using a method we outline below.

As described in previous sections, the LAT (as modeled) is entirely characterized by its IRF. By combining a simulated or actual history of the S/C position and orientation (*FT2* file) with the IRF and a source model, a realization of data from the sources can be created. This task is implemented in the *gtobssim* tool developed by the LAT Collaboration. We make extensive use of it below. Except where otherwise specified we use the *P6_v3_diff* IRF for simulating and fitting validation data. Additionally, except where otherwise noted, we use the true energy of the simulated photon, i.e, we turn off the effects of energy dispersion.

The majority of LAT sources are modeled as power laws. While the brightest sources may require additional degrees of freedom for accurate representation, dimmer sources are adequately described, from a statistical standpoint, by power laws. For the dimmest sources, even two degrees of freedom may be too much. Thus, while we do test more complex models, especially power laws with exponential cutoffs given their strong connection with pulsars, we concentrate here on power law spectra.

We begin by simulating a diffuse background. We adopt the model used for the 1FGL catalog analysis[12], the *gll_iem_ v02* mapcube for the Galactic diffuse and the *isotropic_iem_v02* intensity tabulation[1] for the isotropic background. We simulate 1 year of data using the

---

[1]http://fermi.gsfc.nasa.gov/ssc/data/access/lat/BackgroundModels.html

actual *FT2* file describing the *Fermi*-LAT position and orientation during the calendar year 2009. Simulating diffuse sources is a slow process, and 1 year of data occupies about 1 GB of disk space. Therefore, for validation requiring multiple realizations of the same source configuration, we simulate photons in a $12°$ disk (radius) about the position of the Vela pulsar, (R.A.,Decl.) = (128.8463, -45.1735).

Next, we simulate a point source with a power law spectrum, an integral flux from 100 MeV to 200 GeV of 10^{-5} ph cm^{-2} s^{-1} and a photon index of either 1.5, 2.0, or 3.0, spanning the typical spectrum of *Fermi*-LAT source spectra. For convenience, we define $\mathcal{F}_X \equiv 10^{-X}$ ph cm^{-2} s^{-1} and let $\mathcal{F}$ indicate a general flux. A source with $\mathcal{F} = \mathcal{F}_5$ is comparable in brightness to the Vela pulsar, the brightest steady γ-ray source in the sky. We are interested in the performance of *pointlike* over a range of source fluxes or, more importantly, signal-to-background ratios. Strong sources expose inaccuracies when statistical fluctuations are unimportant, while weak sources test both statistical issues (e.g., error estimate) and bias from strong backgrounds. Rather than simulate different data sets for each flux, we resample a subset of photons from the original $\mathcal{F}_5$ file using the following prescription:

1. Draw a Poisson random variable N_{tar} from a distribution with rate $N_{tot}\mathcal{F}/\mathcal{F}_5$.

2. Select N_{tar} photons uniformly from the original file set.

Here, N_{tot} is the number of photons in the original data set and $\mathcal{F}$ is the target flux. Note that we make no cut on incidence or zenith angle.

We simulate 20 base sets with $\mathcal{F} = \mathcal{F}_5$ each with an integration time of 1 year, and from these we can generate subsets for flux-dependent study of ensembles of sources. But before we proceed to ensemble testing, we take advantage of the extreme precision allowed by summing data from many MC realizations to verify various aspects of the analysis outlined above.

5.1   *Validation with Summed Data*

While the 20 independent realizations are valuable for ensemble testing, by combining them into a single data set we can test pieces of the *pointlike* framework to high precision. When

combined, since we use the exposure for only a year, we are essentially using a year of data for a source with a flux of $2\mathcal{F}_4$. Likewise, the diffuse background is 20 times brighter than in actuality. The resulting quantity of photons, a few million, allows measurements of deviations of the model and data to a few percent. In the following sections, we examine our implementation of the point-spread function and convolution of diffuse sources. We conclude with a comparison of the maximum likelihood parameters estimated from the summed data with their Monte Carlo truth values.

5.1.1   Checking the Band PSF Shape

We wish to verify two aspects of the approximate exposure-weighted PSF we defined for each energy band (see Eq. 4.11). First, do we correctly characterize the *shape* of the distribution of photons for a point source, and second, do we accurately calculate the overlap integral, Eq. 4.24? We begin with the PSF shape.

First, recalling the azimuthal symmetry of the PSF, in each band we bin the photons in θ, the angular separation of the reconstructed photon position and the point source position. (Since we are using binned data, the reconstructed photon position is actually associated with a HEALPix pixel center, but this position differs negligibly from the actual position by design.) To achieve an approximately equal number of photons in each band, we invert the cumulative distribution function furnished by the on-axis PSF and use the resulting quantiles such that, e.g., $\approx 5\%$ of the photons are in each θ bin. (We choose bins with equal counts, rather than bins with equal size in θ or θ^2, so that the statistical weight of the bin on the results is easy to infer.)

In calculating the expected number of counts in a particular bin, we must account for the "ragged binning" of the HEALPixels. To do this, we simply integrate the PSF over each θ bin numerically by evaluating the PSF at each HEALPixel, occupied or not, in the θ bin, and take the mean. Since we are interested in the *shape*, we normalize the sum of the predicted counts over all θ bins to the total observed counts. Importantly, then, these results are *independent* of how accurate the source model and estimate of the exposure are.

The results for a power law source with $\Gamma = 2.0$ are shown in Figs. 5.1.1, 5.1.1, and
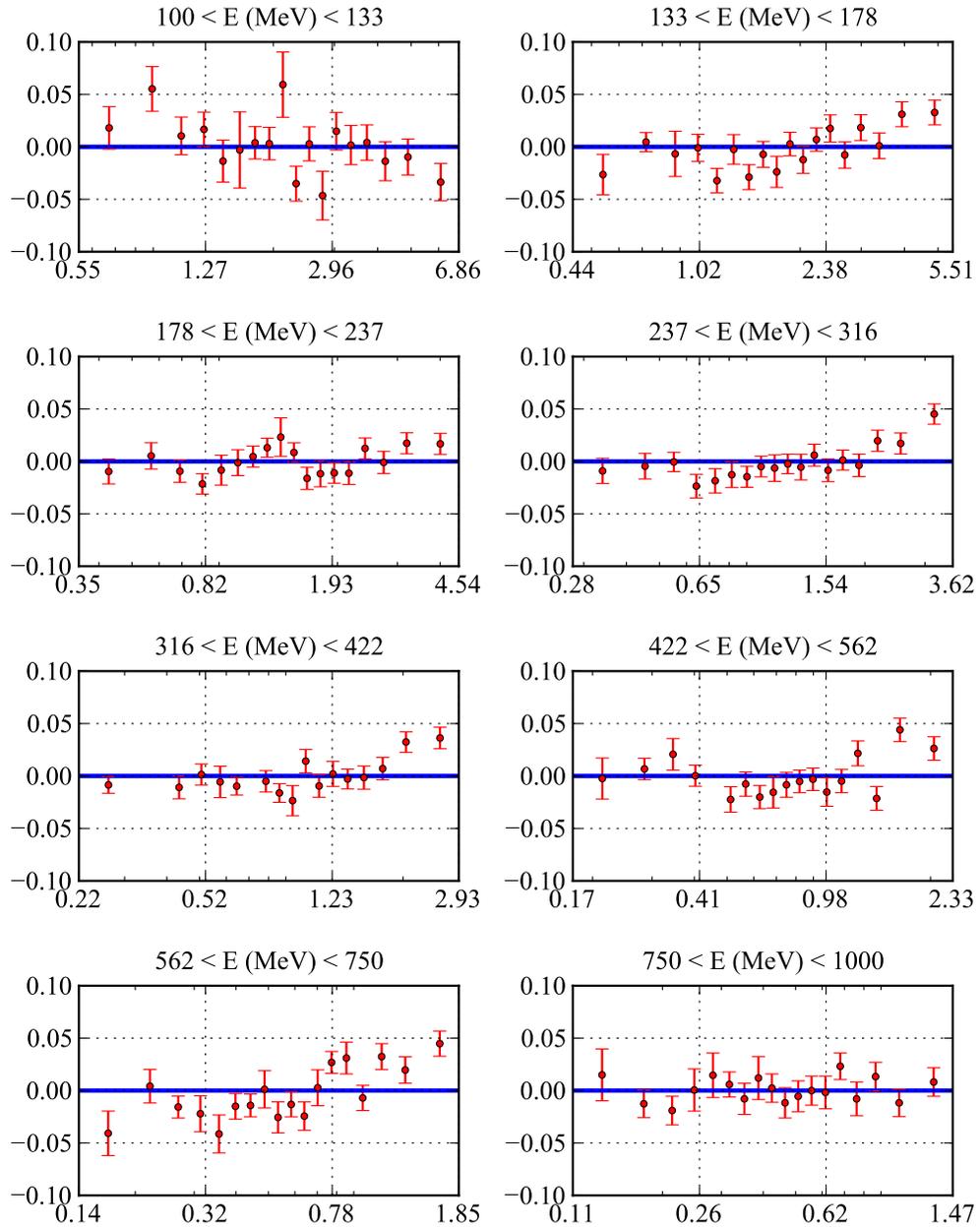
Figure 5.1: The PSF profile for bands with front-converting events between 100 and 1000 MeV. The x-axis gives the angular separation of the bin center from the point source position in degrees and is on a logarithmic scale. The y-axis indicates the relative difference, observed less predicted, in percent. These results are for a source with a power law spectrum of photon index 2.0.
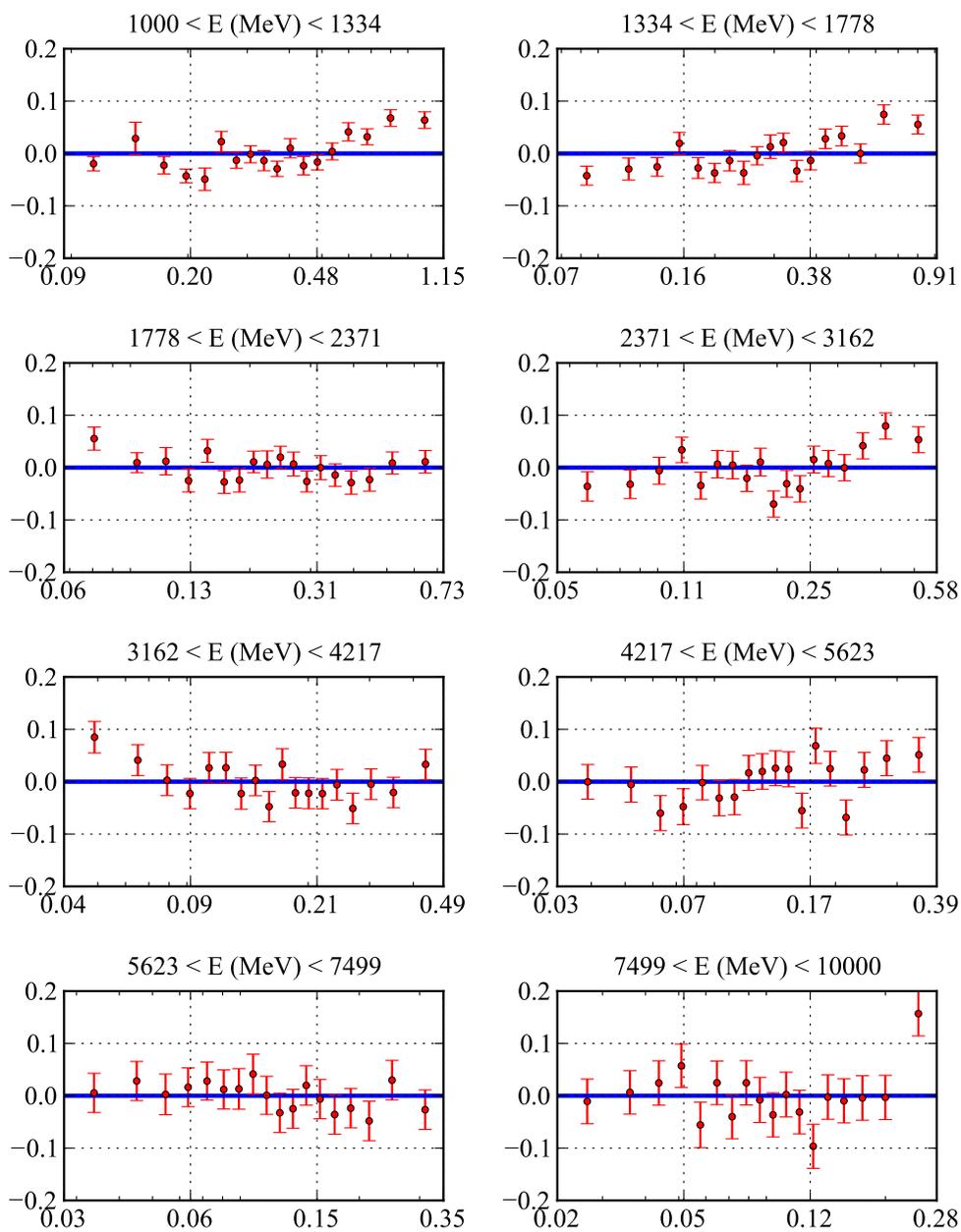
Figure 5.2: As Figure 5.1.1, for bands with front-converting events between 1000 and 10000 MeV.
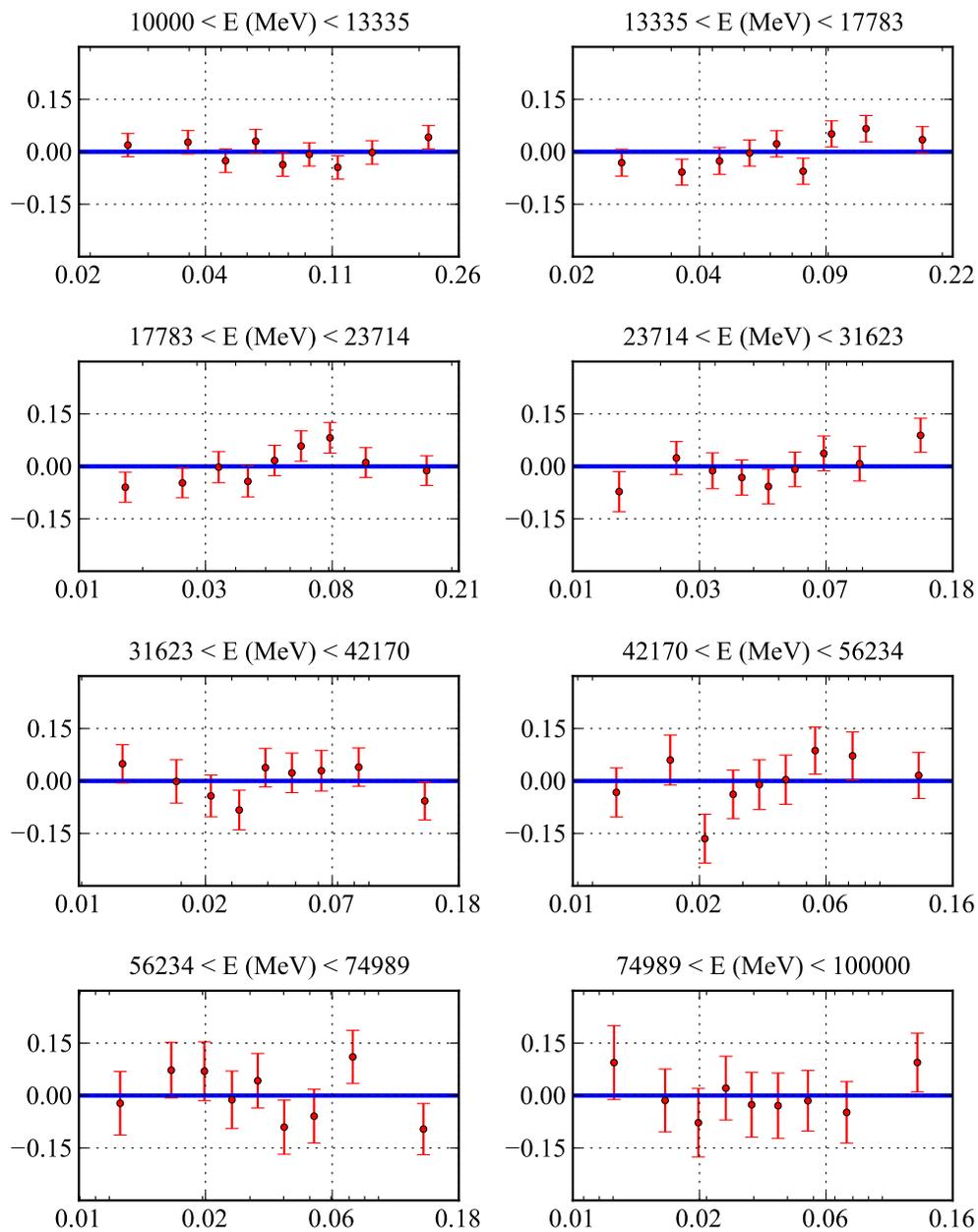
Figure 5.3: As Figure 5.1.1, for bands with front-converting events between 10000 and 100000 MeV.

	$\Gamma = 1.5$	$\Gamma = 2.0$	$\Gamma = 3.0$
$\Delta \log \mathcal{L}$ (Front)	287	206	82
$\Delta \log \mathcal{L}$ (Back)	539	378	156

Table 5.1: The change in the log likelihood for 20 combined years of Monte Carlo data for three different spectral shapes when comparing band PSFs based on E_{geo}, the geometric mean energy for the band, and E_{opt}, the optimal energy estimated by Eq. 4.23.

5.1.1. These results, for front-converting events, are representative of results for $\Gamma = 1.5$ and $\Gamma = 3.0$ and for back-converting events. Generally, the outcome is quite good. For many bands, the residuals are perfectly flat to within the statistical error bars. Others show a slight trend, i.e., the PSF is too narrow (broad), resulting in negative (positive) residuals in the core and positive (negative) residuals in the tail. This is to be expected, as the band PSF is only approximate.

We can also investigate what gains we made by optimizing the energy at which we evaluate the scaled σ parameters for our band PSF E_{opt} (see §4.2.2) for our band PSF. We compare in Table 5.1 the log likelihoods for the three spectral shapes using both E_{opt} as estimated by Eq. 4.23 and the simple prescription of $E_{\mathrm{opt}} = E_{geo}$, i.e., choosing the geometric mean energy as the optimal energy. There is an appreciable increase in the log likelihood with E_{opt} as estimated from Eq. 4.23.

5.1.2 Checking the Overlap Integrals

Next, we compare the total number of counts predicted by the model to that actually observed in the entire ROI. The accuracy of the predicted counts depends on the value of the exposure, the proper integration of the spectral model over the energy band, and the correct calculation of the PSF overlap. The exposure is typically binned with $1°$ pixels, and varies sufficiently slowly that the resulting error is $< 0.5\%$. Likewise, we have shown in Figure 4.2.2 that the integral of the spectral model introduces negligible error. The primary source of error is then the calculation of the overlap integral, which in turn comprises two

error sources. Numerically, there is neglibible error, but since we use a single, approximate PSF for the band, the overlap integral will invariably depart from the true overlap. Second, while we calculate the overlap integral assuming a circle aperture, the use of binned data results in an actual data set with a "ragged" edge.

We test these effects with the production code by again summing all 20 years of Monte Carlo data and comparing the observed and actual counts in a 10° ROI. Since we are interested in the absolute accuracy of the model predictions, we do not renormalize the data or perform an initial spectral fit. The results appear in Figures 5.1.2 and 5.1.2 and indicate overall accuracy of about 1%[2].

5.1.3 Checking the Convolution

The accuracy of the diffuse convolution is extremely important, particularly in the Galactic plane, because the the diffuse counts dwarf counts from all but the brightest point sources. Small relative inaccuracies can become extremely statistically significant and bias point source fits.

We first validate the spatial distribution, Eq. 4.37 as follows. For each data pixel in the ROI, we evaluate Eq. 4.37, as determined from our convolution scheme (including integration over energy) at the pixel center. Since the pixels are small compared to the PSF, this is an accurate estimate of the pixel rate. We then examine the weighted residuals, $(N_{obs} - N_{mod})/\sqrt{N_{mod}}$, as a function of position. Representative results for two energy bands are presented in Figures 5.1.3 and 5.1.3 and are quite featureless, i.e., the model is correct in both shape and magnitude.

At energies above 1 GeV, the pixelization becomes too small for this scheme to work. However, above 1 GeV, the convolution is less important as the PSF scale is small compared with the model scale (0.5° for *gll_iem_v02*). For the remainder of the validation, we simply present the integral of Eq 4.37 over the aperture in Figure 5.1.3. The model predictions agree with the Monte Carlo data to about 1% at all energies, although the poor statistics (from the relatively soft spectrum of the Galactic diffuse) at high energy preclude validation

[2]The trend to overpredict the counts from 0.1 to 1 GeV is under investigation.
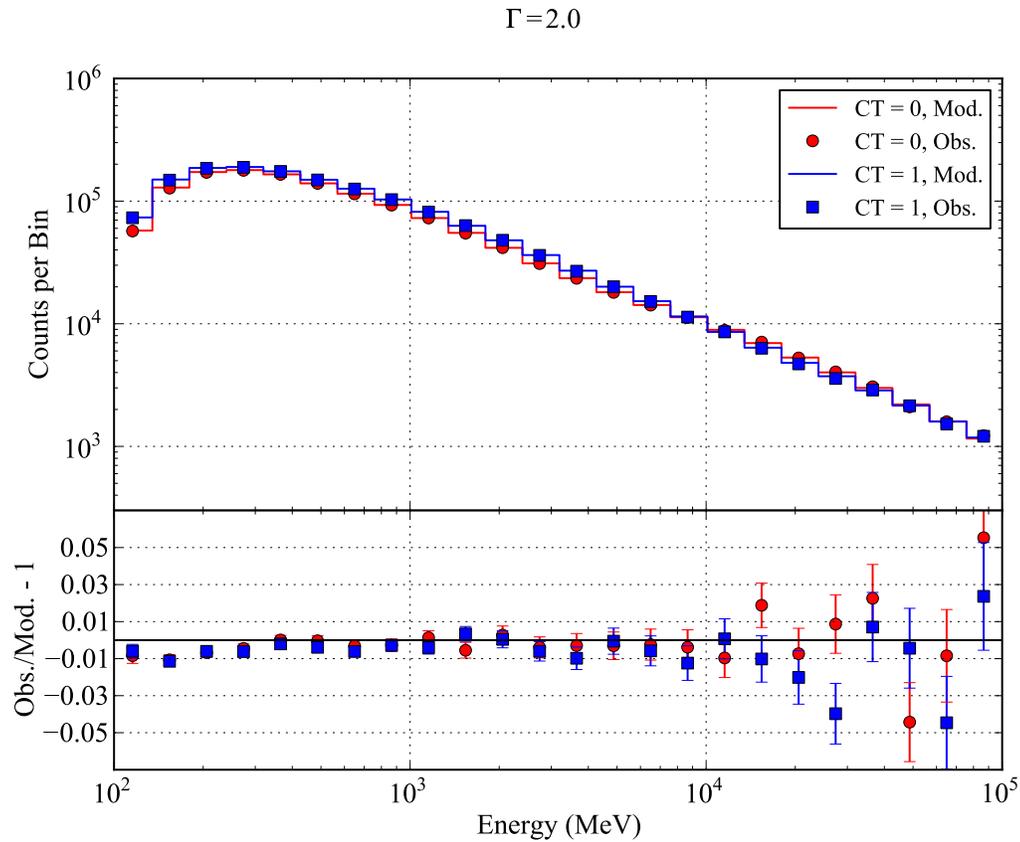
Figure 5.4: Comparison of modeled and observed counts for a point source with a $\Gamma = 2.0$ spectrum. The histogram gives the model expectation and the data points the observed counts. The residuals indicate that the accuracy is better than 1% at all energies. The minor but definite trend for an excess in the expected counts below 1 GeV can lead to a very small bias in the estimation of the photon index—see Table 5.2. It is in this energy range the effects of PSF approximation and "ragged" edge effects are most pronounced.
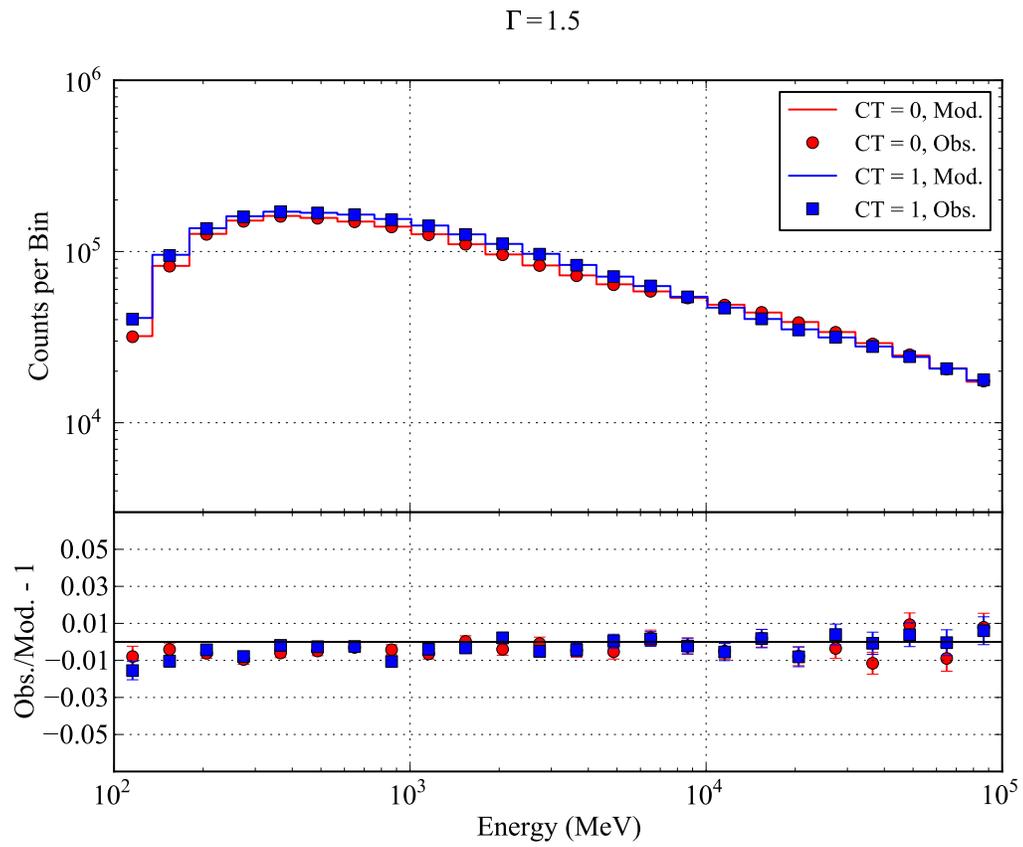
Figure 5.5: Comparison of modeled and observed counts for a point source with a $\Gamma = 1.5$ spectrum. We include this model for its superior statistics at high energy, where it is clear the accuracy remains better than 1%.

$$100 < E \ (\mathrm{MeV}) < 133$$
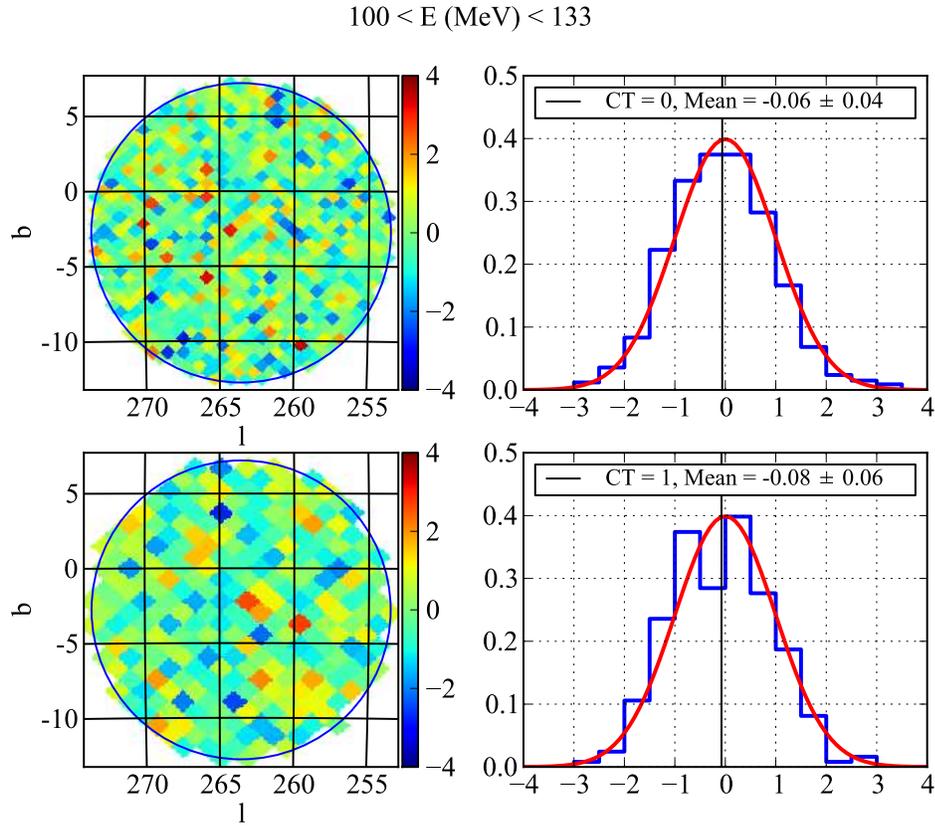
Figure 5.6: The lefthand panels show the weighted residuals, $(N_{obs} - N_{mod})/\sqrt{N_{mod}}$, for the diffuse background (*gll_iem_v02 + isotropic_iem_v02*) at low energy in zenithal equal-area (ZEA) projection. The top (bottom) row shows front-converting (back-converting) events. The "ragged edge" of the HEALPix data compared to the ROI boundary (blue circle) is clear at these low energies, particularly for the back events. The righthand panels show a histogram of the weighted residuals. There are many counts per pixel for these low energies (mean 102 (360) for front (back)), and the residuals should (and do) follow a normal distribution (shown in red). The means of the distribution are consistent with 0, indicating no statistically significant bias.
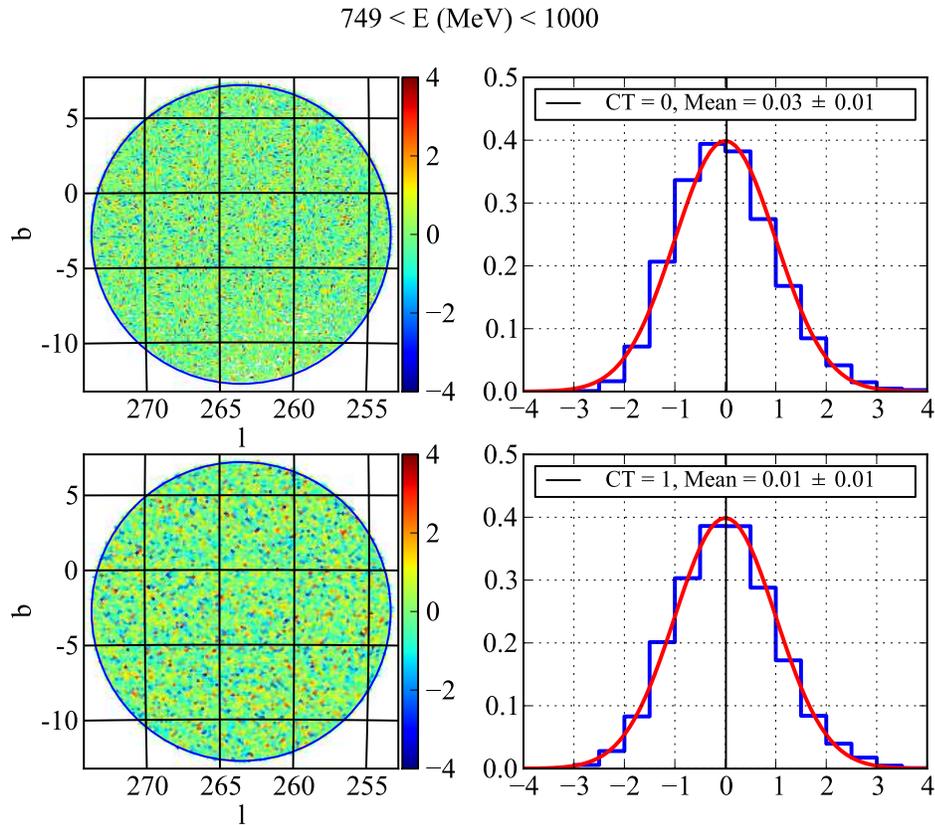
Figure 5.7: Weighted diffuse residuals at moderate energies. The spatial residuals are again featureless. The count rate per pixel is at the limit of normal approximation (8 (24) for front (back)), and the residuals for the front-converting events clearly show a long right tail in keeping with the asymmetric Poisson distribution. (The mean is accordingly slightly biased.) The "ragged" edge is essentially eliminated by 1 GeV for nearly any reasonable ROI size.
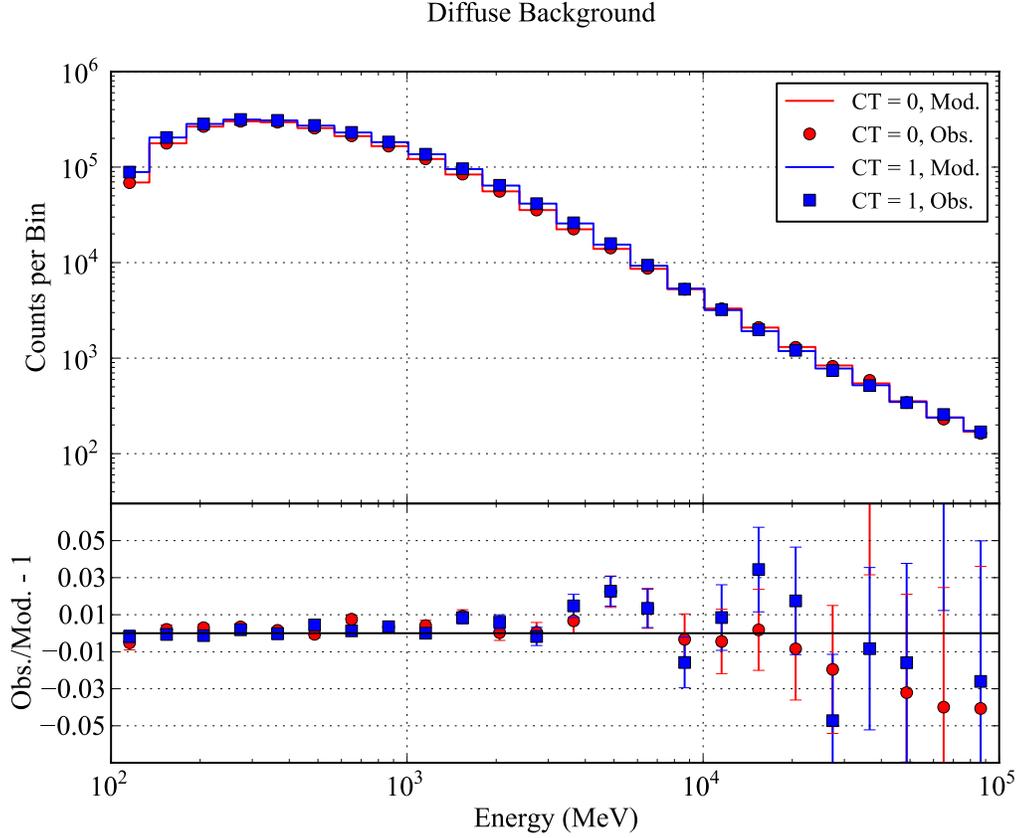
Figure 5.8: The predicted and observed counts for the diffuse background (*gll_iem_v02* + *isotropic_iem_v02*) for all energies and conversion types.

beyond about 5%.

5.1.4 Checking the Maximum Likelihood Fit

We test the absolute level of accuracy of the maximum likelihood estimates of the spectral parameters by performing maximum likelihood fits to the *summed* data sets, i.e., by measuring the spectrum of a single point source with $\mathcal{F} = 2\mathcal{F}_4$. Such an exercise indicates the magnitude of any bias for a point source in the limit of very small statistical errors. Table 5.2 shows the results for a point source alone, indicating better than percent-level accuracy. Table 5.3 shows results for the same sources in the presence of the diffuse background. The

	$\Gamma = 1.5$	$\Gamma = 2.0$	$\Gamma = 3.0$
$100 \times (\mathcal{F}_{sim}/\mathcal{F}_{fit} - 1)$ (CT=0)	-0.54	0.54	0.18
$100 \times (\mathcal{F}_{sim}/\mathcal{F}_{fit} - 1)$ (CT=1)	0.53	0.62	0.74
$100 \times (\Gamma_{sim}/\Gamma_{fit} - 1)$ (CT=0)	0.07	0.11	0.02
$100 \times (\Gamma_{sim}/\Gamma_{fit} - 1)$ (CT=1)	0.10	0.04	-0.01

Table 5.2: The accuracy of maximum-likelihood estimated parameters for three point sources with a flux of 2×10^{-4}ph cm^{-2} s^{-1} and three different power law photon indices. The agreement is much better than 1% for all three sources, and there is additionally good agreement between the estimates for front-converting events and back-converting events. The general trend is for the source-as-fit to have a harder spectrum than simulated; this is in keeping with the slight excess of modeled events at low energy, as a hardened spectrum ameliorates this excess.

additional degrees of freedom lead to a modestly increased bias in the flux measurement, but the accuracy is still about 1%; the accuracy of the photon index remains much better than 1%.

5.1.5   Checking the Maximum Likelihood Fit: With Energy Dispersion

Previously, we have ignored the effects of energy dispersion by using data binned using the *simulated* energy for each photon. To illustrate and gauge the effect of energy dispersion, we repeat two of the above exercises with data binned using the *reconstructed* energy. In Figures 5.1.5 and 5.1.5, the residuals—in which the model neglects energy dispersion but the data contains it—for two point sources with power law spectra appear. It is clear that energy dispersion is not entirely negligible! The residuals can be of order 10%, and the dispersion introduces slight curvature in the spectrum[3]. However, it is clear that much of the curvature can be eliminated by restricting analysis to photons with reconstructed energies

---

[3]The dominant mechanism is as follows: the effective area increases rapidly from 0.1 to about 0.5 GeV, while the energy dispersion is approximately symmetric in logarithmic energy, and there is a net "migration" of photons from high to low energy, causing the model to underpredict at low energy and overpredict at high energy.

	$\Gamma = 1.5$	$\Gamma = 2.0$	$\Gamma = 3.0$
$100 \times (\mathcal{F}_{sim}/\mathcal{F}_{fit} - 1)$ (CT=0)	0.84	1.01	0.87
$100 \times (\mathcal{F}_{sim}/\mathcal{F}_{fit} - 1)$ (CT=1)	0.92	1.26	1.62
$100 \times (\Gamma_{sim}/\Gamma_{fit} - 1)$ (CT=0)	0.14	0.15	-0.03
$100 \times (\Gamma_{sim}/\Gamma_{fit} - 1)$ (CT=1)	0.18	0.12	-0.07

Table 5.3: As Figure 5.2, but with the Galactic and isotropic diffuse backgrounds added. The agreement remains at the 1% level.

	$\Gamma = 1.5$	$\Gamma = 2.0$	$\Gamma = 3.0$
$100 \times (\mathcal{F}_{sim}/\mathcal{F}_{fit} - 1)$ (CT=0)	0.31	0.89	3.68
$100 \times (\mathcal{F}_{sim}/\mathcal{F}_{fit} - 1)$ (CT=1)	0.55	1.40	4.86
$100 \times (\Gamma_{sim}/\Gamma_{fit} - 1)$ (CT=0)	-0.53	-0.64	-0.11
$100 \times (\Gamma_{sim}/\Gamma_{fit} - 1)$ (CT=1)	-0.31	-0.41	0.25

Table 5.4: As Table 5.2, but now including the effects of energy dispersion in the data. No diffuse background is included.

above 200 MeV. Then—for power law sources, at least—the effect of energy dispersion is primarily a renormalization of the spectrum, decreasing the measured flux below its true value by a few percent.

While the residuals look somewhat dire, the parameter values for bright sources are not significantly affected. Tables 5.4 and 5.5 show the absolute deviation of the measured parameters from the Monte Carlo truth without and with, respectively, the presence of a diffuse background. The bias from neglect of energy dispersion is $< 5\%$.

5.2 Spectral Analysis of Ensembles: Verification of Central Values and Error Estimates

The most important capability of *pointlike* is reliable and rapid extraction of spectral parameters and estimates for their statistical errors. To test this, we perform a spectral analysis
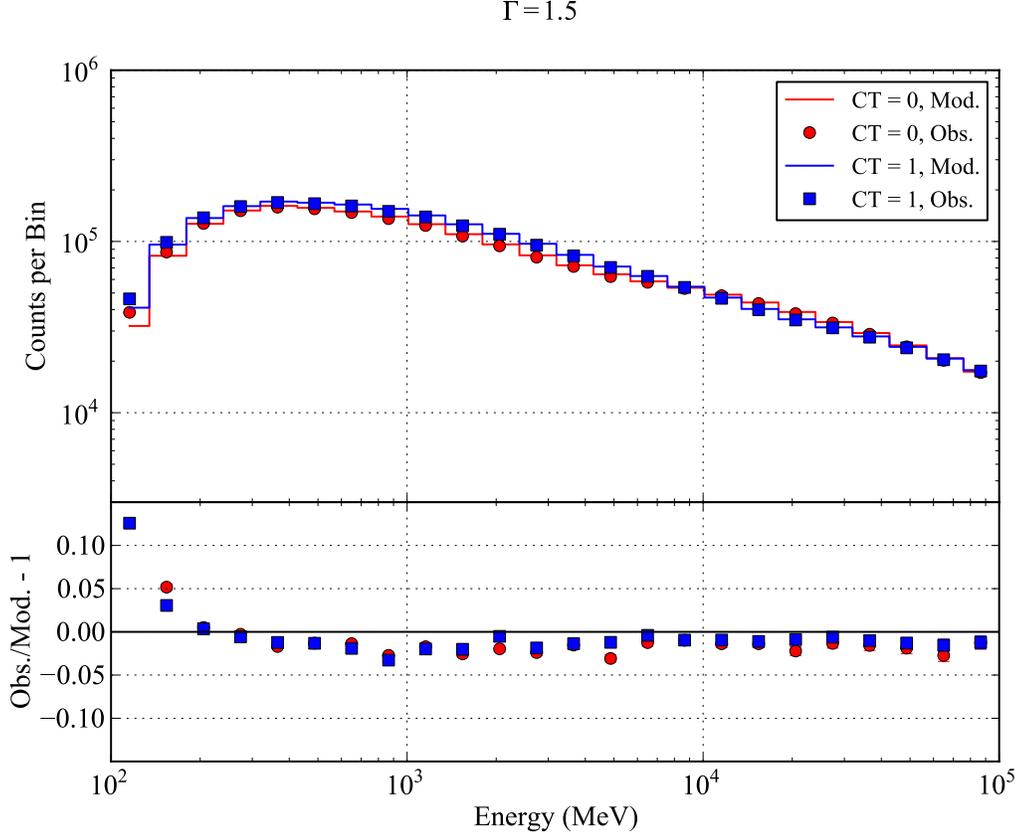
Figure 5.9: As Figure 5.1.2, but for a power law source with photon index of 1.5 and allowing for energy dispersion in the data. The model value is fixed at the Monte Carlo truth, so the residuals indicate how the photons have been redistributed from the true values. The intrinsic spectrum of hard sources enhances the tendency for high energy photons to redistribute to low energies.

	$\Gamma = 1.5$	$\Gamma = 2.0$	$\Gamma = 3.0$
$100 \times (\mathcal{F}_{sim}/\mathcal{F}_{fit} - 1)$ (CT=0)	0.34	-0.14	-1.86
$100 \times (\mathcal{F}_{sim}/\mathcal{F}_{fit} - 1)$ (CT=1)	0.06	0.56	-1.31
$100 \times (\Gamma_{sim}/\Gamma_{fit} - 1)$ (CT=0)	-0.06	-0.23	-0.96
$100 \times (\Gamma_{sim}/\Gamma_{fit} - 1)$ (CT=1)	0.08	-0.27	-1.30

Table 5.5: As Table 5.4, but now including a diffuse background.

$\Gamma = 3.0$
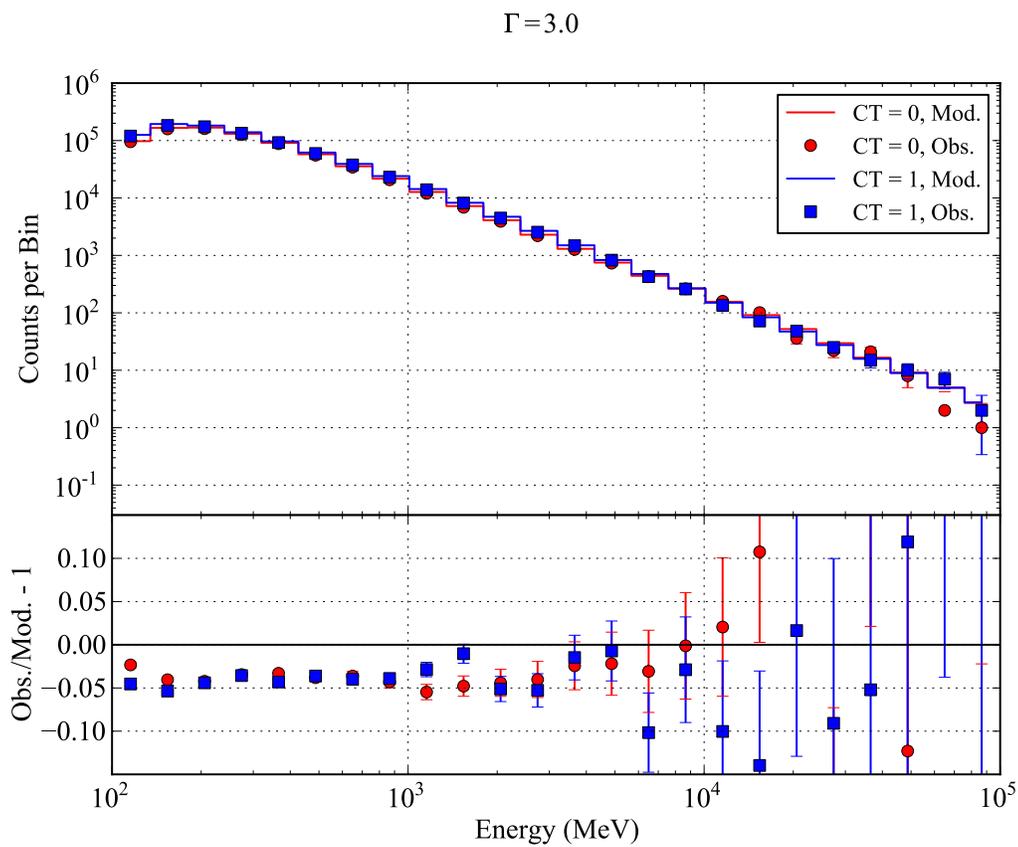
Figure 5.10: As Figure 5.1.5, for a power law source with photon index of 3.0. Soft sources show a decreased curvature at low energies, but the overall loss of photons from the passband is increased.

of 20 Monte Carlo realizations of the diffuse background and a point source at a given flux. We plot the results for a series of decreasing point source fluxes, $\mathcal{F}_6$, $\mathcal{F}_7$, $\mathcal{F}_8$, and $5\mathcal{F}_9$. The latter flux is below the detection threshold of the LAT for modest (1-2 years) integration baselines for sources near the Galactic plane, while the penultimate ($\mathcal{F}_8$) is marginal. $\mathcal{F}_7$ is the flux of a typical Galactic source. We show the results both in absolute units and in units in which the parameters become approximately bivariately normal.

5.2.1 Spectral Parameters

We consider first Fig. 5.2.1 which shows a scatter plot of the estimated integral flux and photon index for a power law source with $\Gamma = 2.0$ in error-weighted units: $x'_i = (x_i - \hat{\mu}_x)/\sigma_i$, i.e., we have subtracted from each estimated parameter the sample mean and then divided by the estimated error for the parameter. If the error estimates are accurate (implying the likelihood surface is approximately Gaussian), then the error-weighted estimates should be normally distributed. Thus, we draw the confidence contours of a two-dimensional Gaussian (using the sample covariance matrix) on each figure to determine if the appropriate sample fraction lies within the appropriate contour. In general, this is the case, indicating (a) the Gaussian approximation of the likelihood surface is a good one and (b) the likelihood surface is being accurately measured by the error estimation algorithms.

The sample mean has an error that scales as the square root of the sample size, say $1/\sqrt{N}$. With sufficiently accurate estimates of the parameter errors (bright sources) and/or sufficiently numerous Monte Carlo realizations, we can begin to resolve systematic bias in the parameter estimates. We adopted the former approach in the previous section and explore the latter approach here. The magnitude of the bias, in units of the statistical error of a single source, can be read off of the figure by the position of the black cross with error bars. For the bright sources ($\mathcal{F} = \mathcal{F}_6$), systematic bias is detectable but is at most comparable to the statistical error. For dimmer sources, the statistical errors dominate.

To gauge the absolute magnitude of both the statistical and systematic errors, we display the absolute parameter values in Fig. 5.2.1. (Note we have still scaled the flux by the simulated value.) Here, the position of the parameters relative to the contours are not

meaningful and are only shown for reference. In the case of $\Gamma = 2.0$ and $\mathcal{F} = \mathcal{F}_6$, we see the absolute error on the flux is $< 2\%$.

5.2.2 Position Parameters

To check the accuracy of the MLE for the source position and position uncertainty, we first performed a ML fit for the spectral parameters and followed with a ML fit for the position. We did not iterate further. The deviations in relative units and error-weighted units for Right Ascension and Declination are shown in Figures 5.2.2 and 5.2.2, indicating that the mean of the position estimate is consistent with the simulated value and the the position uncertainty estimates are not too small, indicating only a small correlation between the position parameters and the spectral parameters. Some of the low-flux ($\mathcal{F} \leq \mathcal{F}_8$) sources have a signal-to-noise ratio too low for robust localization.

By symmetry and the apparent insensitivity of the best-fit position to the precise value of the spectrum, we conclude that position estimates will be unaffected by energy dispersion.

5.3 Summary

By combining 20 years of simulated data for bright sources, we demonstrated that machinery used by *pointlike* to calculate expected rates for point sources and for diffuse sources is generally accurate to within a few percent of the expectation. By performing spectral (and position) ML fits on ensembles of sources, we assessed the accuracy of the central values and uncertainty estimates reported by *pointlike* over 2.5 decades of source brightness, finding generally good agreement with systematic bias $< 10\%$.
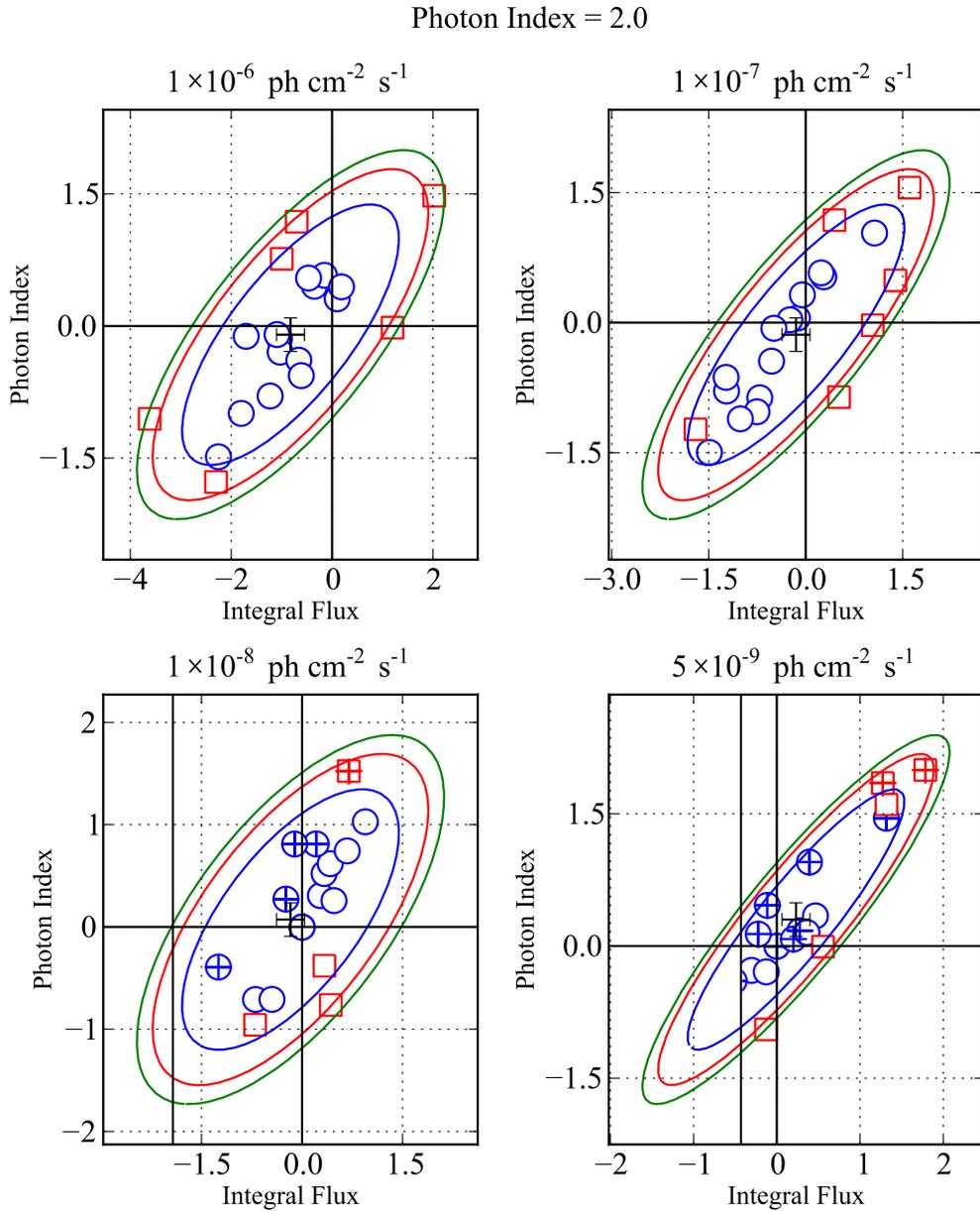
Figure 5.11: Results for $\Gamma = 2.0$ in error-weighted units. The contours give the 68%, 95%, and 99% confidence intervals as estimated from a Gaussian distribution using the sample covariance matrix. Blue circles (red squares) are the 68% (32%) of the sample closest to (farthest from) the mean. The single, black errorbar at the center of the contours gives the sample mean, and the length of the error bars gives the 1-d sample standard deviations divided by $\sqrt{N}$. The black axis marks crossing at the origin represent the simulated ("true") values. Thus, a failure of the cross to overlap the origin indicates a systematic bias in the parameter estimate. The position of the cross relative to the origin gives the magnitude of the bias in "sigma" units. Sources with a test statistic < 10 are indicated with a cross-filled symbol.
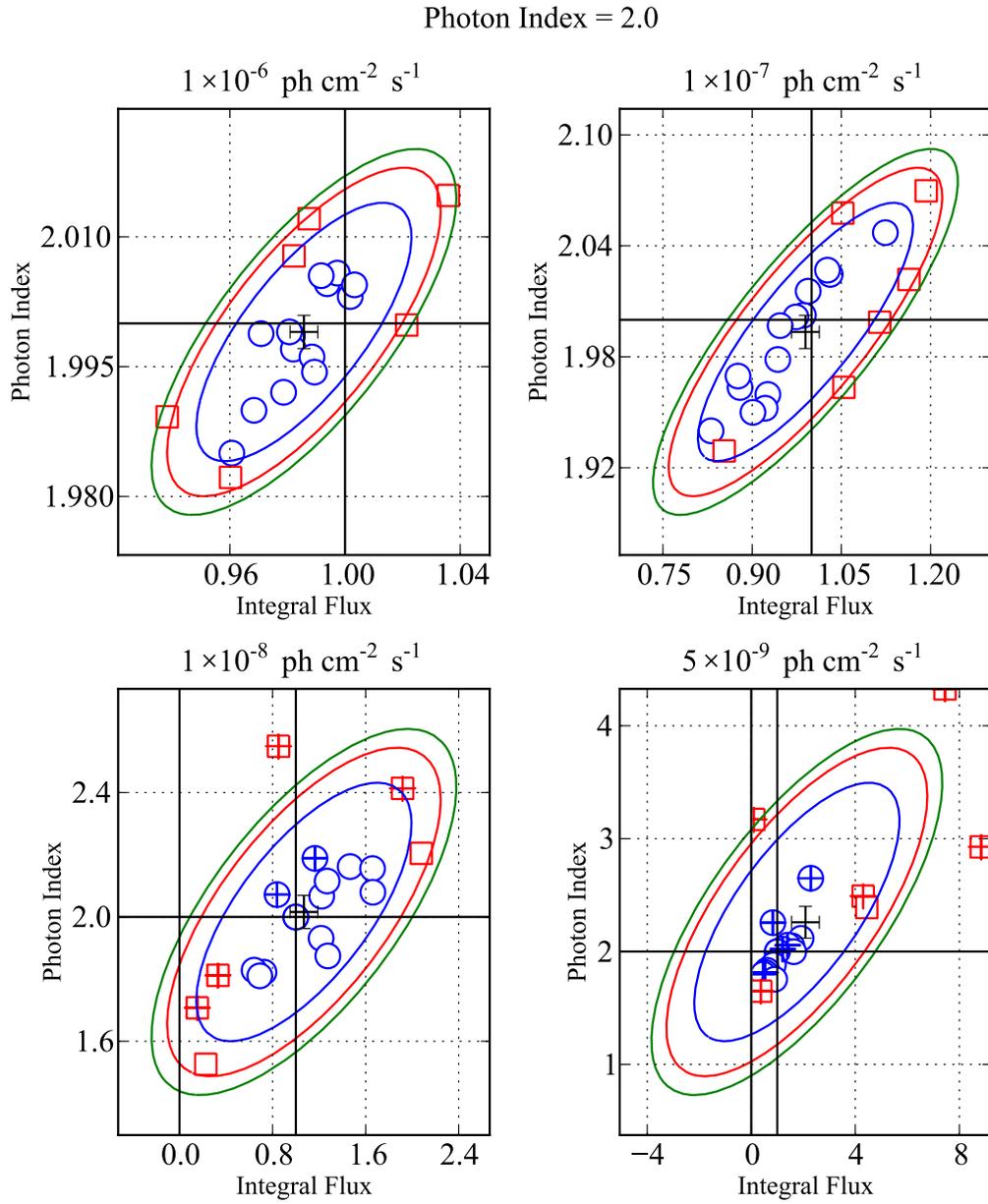
Figure 5.12: As Figure 5.2.1 in relative units. The thick dashed line indicates zero flux. Absolute outlying values are more apparent at low fluxes.
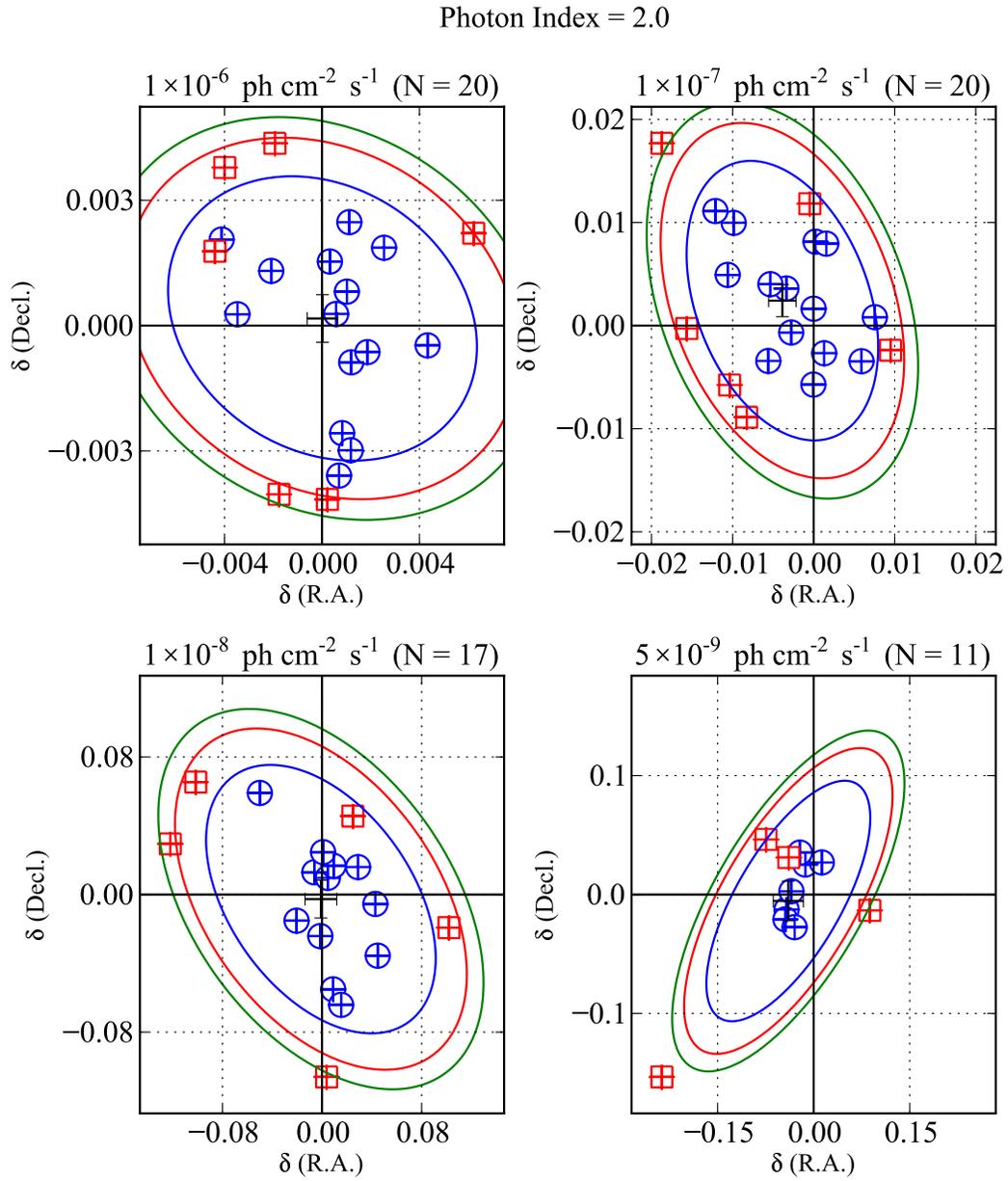
Figure 5.13: The angular deviation in absolute units (°) for the position in the R.A. and Decl. directions. The number of sources with successful position fits is indicated by N.
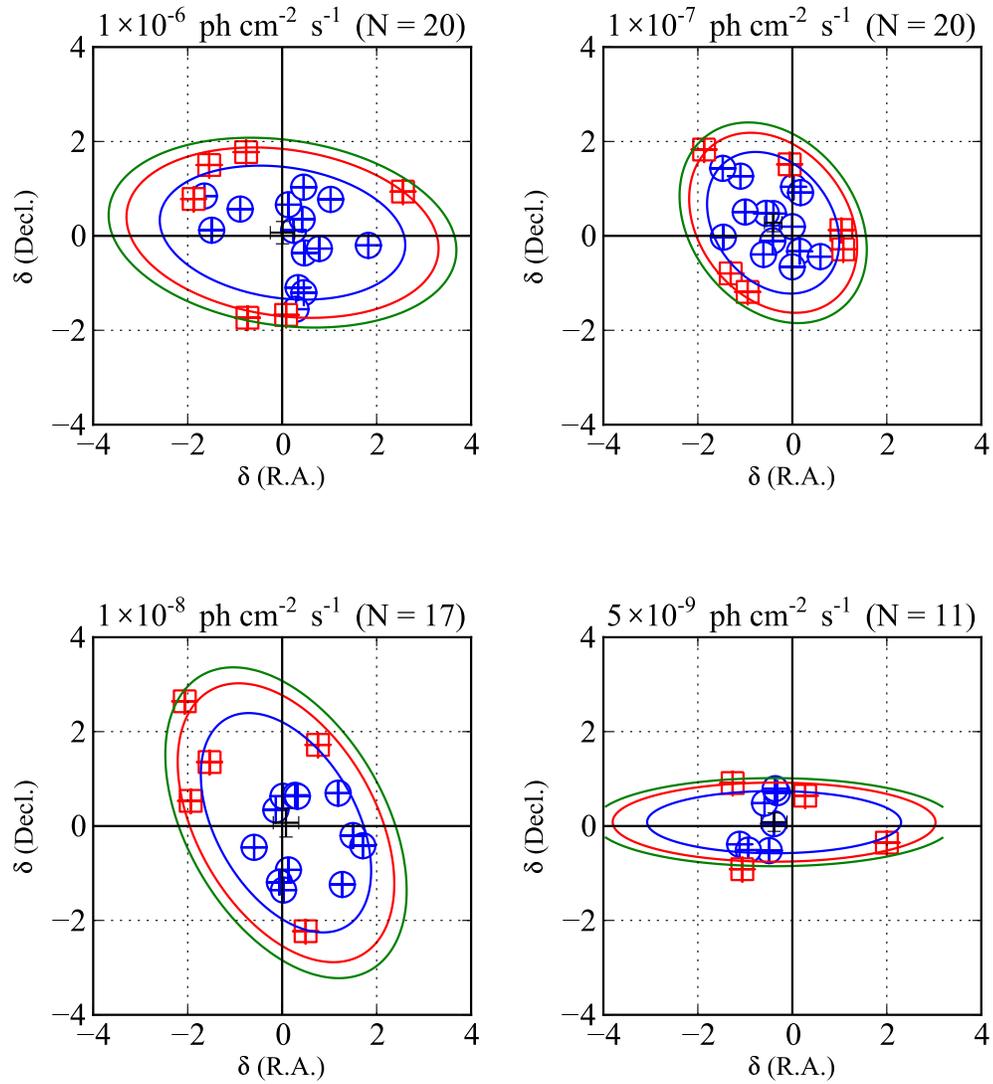
Figure 5.14: As Figure 5.2.2 but with error-weighted units.

Chapter 6

APPLICATIONS OF *POINTLIKE*

Having detailed the inner workings of *pointlike* and validated the tool on a variety of test cases, we now outline some of the science results *pointlike* is capable of delivering. We start with "single-source" analysis—by which we mean analysis of a relatively small region of the sky—to demonstrate the flexibility of *pointlike*. To showcase the speed of *pointlike*, we then develop an all-sky analysis in which we determine spectra for every known GeV source. Finally, we make use of this "optimal" model for other analysis tasks.

6.1 Single-source Analysis

To illustrate the use of *pointlike* in its "single-source" mode, we provide an example analysis of sources in the Cygnus region of the sky. This region, rich in sources, has already delivered multiple discoveries, including the LAT's first detection of a new radio-loud pulsar[8], a particularly bright radio-quiet pulsar in the γ-Cygni supernova remnant[3], the discovery of orbitally-modulated emission from Cygnus X-3[48], and one of the first pulsars discovered in radio *after* detection by *Fermi*-LAT [33]. This is clearly a region that rewards accurate analysis!

6.1.1 Data Prepaparation

The *FT1* and *FT2* files are obtained from a database maintained at SLAC[1]. Similar data products are available from the FSSC[2]. For this exercise, we select data obtained between 4 August 2008 and 8 August 2010, processed using "Pass 6" reconstruction algorithms. We restrict data to events with a reconstructed position lying with 15° of PSR J2021+3651 and

---

[1]http://glast-ground.slac.stanford.edu/DataPortalAstroServer/, SLAC credentials required

[2]http://fermi.gsfc.nasa.gov/ssc/data/access/lat/

with reconstructed energies between 100 MeV[3] and 100 Gev. Finally, we use "diffuse" class events, a subset of events with a low background contamination.

Before ingestion into *pointlike*, we apply the Science Tool *gtmktime*[4] to the data. The tool applies a general filter based on the S/C orientation and position. We apply a filter that modifies the GTI in the *FT1* file to exclude times when the S/C exceeds a $52°$ rocking angle and when the $15°$ ROI defined in the preceding paragraph approaches the horizon. These excisions are designed to decrease contamination from Earth's limb and, since they use the *known* S/C telemetry are essentially bias free.

Additional preparation, binning, and generation of secondary data products are performed within *pointlike*. These procedures are discussed in Chapter 4, and we only provide some of the specifics here: data are binned with eight-bin-per-decade resolution, the livetime as a function of position is tabulated with $1°$ resolution, and we remove events with reconstructed zenith angles $> 105°$ and reconstructed incidence angles $> 66.4°$.

6.1.2 Source Model

After preparing the data, we must construct a spectral model for the sources in the region. Models for both the diffuse background and point sources are in a state of continual improvement as more data is acquired. E.g., additional photons allow dimmer point sources to be detected, which in turn allow a better diffuse model to be generated, in turn allowing more point source detections, and so forth. Systematic refinements independent of integration time also allow for model improvement.

pointlike is capable accepting a wide variety of input spectral models. For instance, it can parse all published *Fermi*-LAT point source catalogs as well as many source lists internal to the LAT Collaboration. It can ingest mapcubes representing a variety of diffuse emission. However, rather than using the most recent internal models, we choose to make contact with the published literature and adopt models used in and resulting from the 1FGL[12] catalog analysis. For diffuse models, we use the *gll_iem_v02* mapcube for the Galactic diffuse and

---

[3]To ameliorate effects of neglecting energy dispersion, we restrict *analysis* to energies above 200 MeV.

[4]http://fermi.gsfc.nasa.gov/ssc/data/analysis/scitools/help/gtmktime.txt

the *isotropic_iem_v02* intensity tabulation[5] for the isotropic background. For point sources, we use the spectral models presented in the 1FGL catalog.

In general, there is a tradeoff between a statistically-optimal fit—a joint fit in which the parameters for all sources are allowed to vary—and the computational intractability of minimizing a high-dimensioned function with a large data set. The typical approach is to select an ROI centered on the source of interest and allow sources within a fixed angular separation—say 5° or 10°—to vary along with the source of interest. Sources at larger radii remain fixed. Due to the large tails of the PSF at low energy, sources lying outside the ROI must be included in the source model, and it is clearly impossible to fit these sources. To estimate parameters for these sources, similar analyses must be performed for other regions of interest in what is essentially a bootstrap procedure. We shall have more to say on this in devising an iterative, all-sky fit below.

In the following analysis, we model the 53 1FGL sources within 20° of the ROI center, and we maximize the likelihood with respect to the 17 1FGL sources within 8° of the ROI center as well as three parameters for the diffuse background, a power law scaling for the Galactic diffuse and a normalization for the isotropic diffuse.

6.1.3 Broadband Spectroscopy

The first step in any analysis is to maximize the likelihood over the free sources. This is done using the gradient fitter as described in Chapter 4. During this process, spectral parameters and estimates are determined for all 17 free sources. The process takes $1-2$ minutes. As an example of the output, and to begin an example on the importance of correct modeling, we report the broadband spectral parameters for two sources, PSR J2021+3651[8] and 1FGL J2015.7+3708, thought to be a blazar[54].

```
0 -- 1FGL J2021.0+3651 fitted with PowerLaw, e0=798
Norm     : (1 + 0.013 - 0.012) (avg = 0.013) 1.34e-010
Index    : (1 + 0.005 - 0.005) (avg = 0.005) 2.34
Ph. Flux : (1 + 0.021 - 0.021) (avg = 0.021) 1.29e-006  (DERIVED)
```

---

[5]http://fermi.gsfc.nasa.gov/ssc/data/access/lat/BackgroundModels.html

```
En. Flux  : (1 + 0.012 - 0.012) (avg = 0.012) 7.62e-010  (DERIVED)


1 -- 1FGL J2015.7+3708 fitted with PowerLaw, e0=1051

Norm      : (1 + 0.043 - 0.041) (avg = 0.042) 1.47e-011

Index     : (1 + 0.014 - 0.014) (avg = 0.014) 2.46

Ph. Flux  : (1 + 0.073 - 0.068) (avg = 0.070) 3.27e-007  (DERIVED)

En. Flux  : (1 + 0.043 - 0.041) (avg = 0.042) 1.62e-010  (DERIVED)
```

In this output, "e0" indicates the pivot energy, an estimate for the energy at which the flux density ("norm") and photon index ("index") are least correlated. The integral photon flux (> 100MeV,cm^{-2} s^{-1}) and integral energy flux (erg cm^{-2} s^{-1}) round out the parameters. The relative error is reported, and since the parameters are transformed to log space internally, they are naturally two-sided when transformed back, but for bright sources become approximately symmetric.

The 1FGL models only provide power law parameters for the point sources. Yet the two brightest sources in the region, PSRs J2021+4026[3] and J2021+3651, are known to have exponentially-suppressed spectra. Further, 1FGL J2015.7+3708 is only 1.1° from PSR J2021+3651. To correct this flaw in the model while simultaneously demonstrating the flexibility of *pointlike*, we modify the sources interactively. We first add a cutoff energy to J2021+4026 and re-maximize the likelihood, increasing its logarithm by 480. We repeat the process for J2021+3651, improving the log likelihood by 296. We now have:

```
0 -- 1FGL J2021.0+3651 fitted with ExpCutoff

Norm      : (1 + 0.026 - 0.026) (avg = 0.026) 2.01e-010

Index     : (1 + 0.020 - 0.020) (avg = 0.020) 1.74

Cutoff    : (1 + 0.065 - 0.061) (avg = 0.063) 2.93e+003

Ph. Flux  : (1 + 0.034 - 0.033) (avg = 0.033) 8.18e-007  (DERIVED)

En. Flux  : (1 + 0.018 - 0.017) (avg = 0.018) 5.44e-010  (DERIVED)


1 -- 1FGL J2015.7+3708 fitted with PowerLaw, e0=1051

Norm      : (1 + 0.040 - 0.038) (avg = 0.039) 1.6e-011
```

Method	R.A.	Decl.	r95 (°)	δ/r95
Timing	305.2728	36.858	—	—
BB, PL	305.2533	36.850	0.010	1.5
BB, PL+EC	305.2543	36.848	0.010	1.5
BF	305.2542	36.848	0.010	1.5

Table 6.1: A comparison of the DC position obtained for PSR J2021+3651 with three methods to the (well-constrained) timing position[8]. The "BB" entries corresponding to using a broadband spectral model (respectively a simple power law and a power law with exponential cutoff), while "BF" refers to the model-independent method discussed in Chapter 4. The three methods are self-consistent, but inconsistent with the radio position at high confidence.

```
Index    : (1 + 0.015 - 0.014) (avg = 0.014) 2.62
Ph. Flux : (1 + 0.071 - 0.067) (avg = 0.069) 4.67e-007  (DERIVED)
En. Flux : (1 + 0.044 - 0.042) (avg = 0.043) 1.94e-010  (DERIVED)
```

We note the reported spectrum for 1FGL J2015.7+3708 has softened significantly, $\Gamma = 2.46 \to \Gamma = 2.62$, well outside of the statistical errors of ± 0.04, while the integral flux has increased by $> 40\%$. We take up again the (cautionary) tale of this source when we discuss model-independent spectroscopy, but first we give a brief example of localization with *pointlike*.

6.1.4 Source Localization

To demonstrate the localization capabilities of *pointlike*, we again focus on the region near PSR J2021+3651. Since the position of J2021+3651 is well-known from radio timing, its brightness in γ-rays allows a validation to high statistical precision of the ML best-fit position. The results of localizing J2021+3651 (having first made the spectral modifications described in §6.1.3, viz. adding cutoff energies to the two bright pulsars) appear in Table 6.1.

Method	R.A.	Decl.	r95 (°)	δ/r95
Timing	305.2728	36.858	—	—
BB, PL+EC	305.2685	36.859	0.010	0.83

Table 6.2: PSR J2021+3651, comparison of the DC position to the (well-constrained) timing position. This fit includes a new, highly significant source, and the *Fermi*-LAT position now agrees with the timing position to a precision of $< 1'$.

In this iteration, it is clear that either the position is off or the uncertainty estimate is too small, as the best-fit position lies well outside of r95, the radius inside which we expect the ML position to lie in 95% of realizations of the experiment. In the time since the 1FGL catalog was collated, we have detected a new point source at (R.A., Decl.) $\approx (304.46, 36.5)$[6]. (We use this source as an example of the source detection algorithm in §6.4.) This source is $< 0.8°$ from PSR J2021+3651, and we expect that failing to account for its emission may bias—albeit slightly—the position of the pulsar. We therefore add a new point source to the model—a task which can be done interactively—at a trial position estimated from the TS map appearing in Figure 6.4.2. We adopt a simple power law with Ansatz parameters ($\Gamma = 2.0$ and an integral flux of 10^{-7}ph cm^{-2} s^{-1}) for the spectum, and we (a) re-maximize the likelihood, (b) re-localize the new source, and (c) re-maximize the likelihood. The new source improves the log likelihood by 361, and after incorporating its emission in the spectral model, localization of J2021+3651 yields the results in Table 6.2. With the new model, the LAT position is consistent with the radio position, and with a precision of better than one arcminute! As in the previous section, we see both the importance of correct modeling in likelihood techniques and the consequent benefit of interactive/exploratory analysis.

6.1.5   *Model-independent Spectroscopy*

We introduced a method of generating (nearly) model-independent spectra in §4.4.6. By model-independent, we mean that we do not assume a broadband functional form for the

---

[6]In fact, it was this exercise—and subsequent failure to arrive at a good localization—that led to the detection!

source in question and we maximize a series of likelihoods to estimate the flux in each band. However, we do assume an Ansatz model in the initial fit in order to obtain parameters for the other sources in the ROI in a consistent fashion.

We recall that the broadband spectral fit for 1FGL J2015.7+3708 was affected by the addition of a cutoff parameter to the nearby, bright PSR J2021+3651. We can also examine what impact it has on the model-independent spectrum. In Figure 6.1.5, we show a series of spectra for 1FGL J2015.7+3708 as we progressively refine the model for the nearby sources as described above. In the first panel, we see a strong suppression of the spectrum below ≈ 500 MeV. This is easily understood as the power law model for PSR J2021+3651 overpredicts the emission at low energy (the spectrum is concave down), and through the poor PSF at these energies the overprediction suppresses the flux estimates for neighboring sources. When we model the pulsars correctly, with a cutoff, the low-energy behavior is much improved, as seen in the second panel of the figure. Note that there is essentially no change above 1 GeV; the PSF is sufficiently good that the sources are resolved.

Looking closely at this new fit, we see there is tension between the simple power law model and the model-independent spectra. Motivated by this evidence for spectral curvature, we can switch the model for 1FGL J2015.7+3708 to a log parabolic form:

$$\mathcal{F}(E) = N_0 \left(\frac{E_b}{E} \right)^{\alpha + \beta \log(E_b/E)}, \qquad (6.1)$$

which reduces to a power law for $\beta = 0$. After re-maximizing the likelihod, we see the new fit is in much better agreement with the model-independent points.

Finally, we note that the new source we introduced in the previous section is only $\approx 0.8°$ from 1FGL J2015.7+3708 and, like J2021+3651 but to a much smaller extent, is likely to affect the low-energy spectrum of 1FGL J2015.7+3708. In the final panel of Figure 6.1.5, we show the spectrum after this source has been included in the model. As expected, it "soaks" up some of the low-energy emission and the lowest-energy point of the 1FGL J2015.7+3708 spectrum drops accordingly.

For completeness, we also include the spectra of the two bright pulsars in Figure 6.1.5.
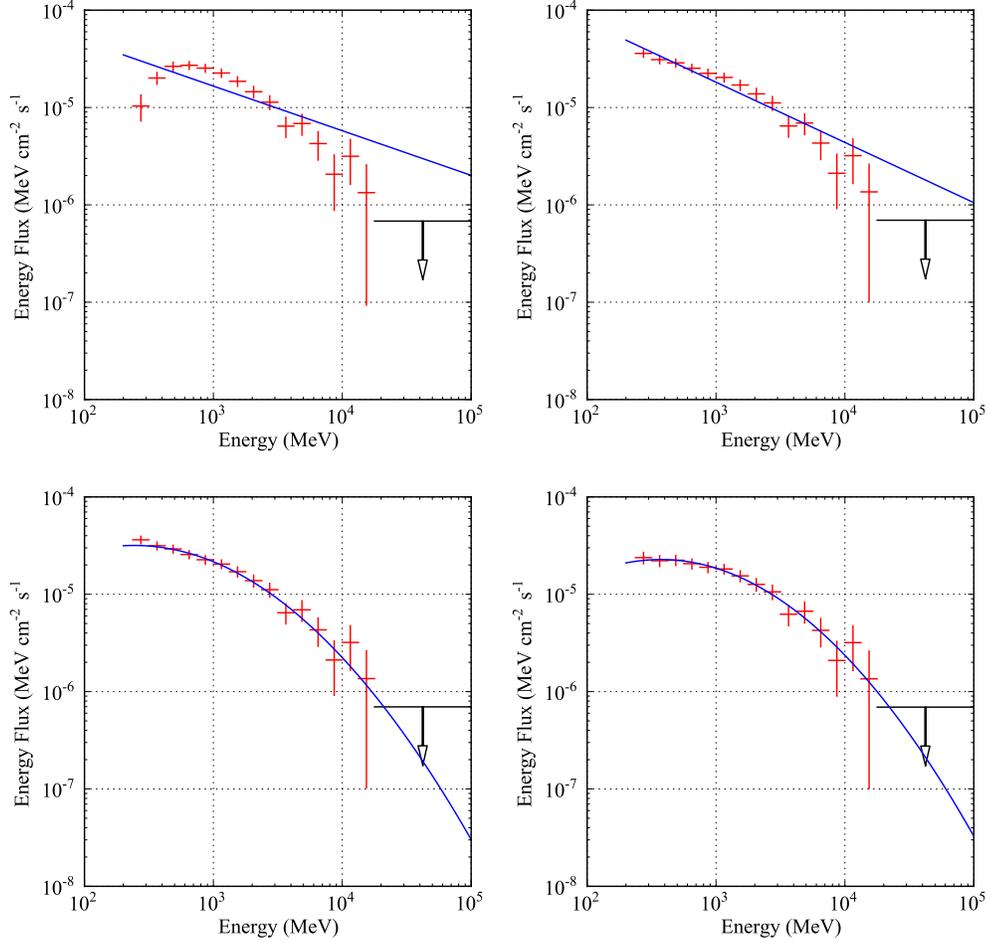
Figure 6.1: The model-independent spectrum for 1FGL J2015.7+3708. At upper left, the initial spectrum estimated with all sources fixed at 1FGL values. At upper right, the spectrum after adding cutoff parameters to PSRs J2021+3651 and J2021+4026. At lower left, no external sources have changed but we model 1FGL J2015.7+3708 with a log parabola (see text). Finally, at lower right, we have added the new source discussed in the previous section.
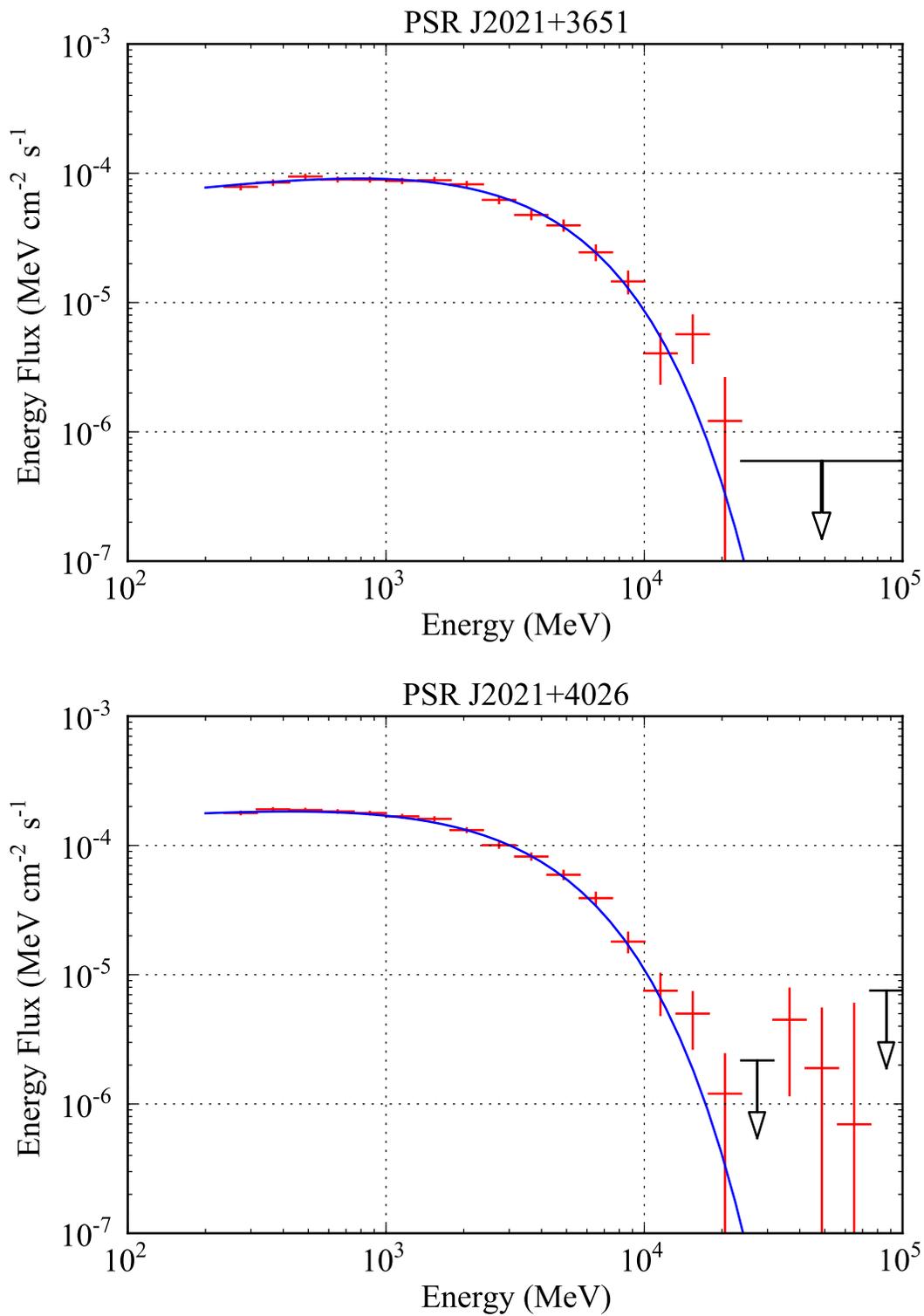
Figure 6.2: Spectra for PSRs J2021+3651 and J2021+4026.

6.2 All-sky Analysis

In discussion of single source (rather, single region) analysis, we mentioned a weakness of this approach: by selecting only a subset of the data, we cannot constrain all of the sources that contribute to the model for the data. That is, since photons from sources from up to $10°$ outside the ROI can disperse into it, we must account for these sources in our model. However, we cannot actually fit these sources since only low-energy photons contribute to the data. We must have some *a priori* model. On the other hand, these same remarks apply to an ROI containing these external sources, and we have the makings of a "chicken and egg" scenario.

Obviously, the larger an ROI becomes, the less important are these "edge effects", particularly for a source in the center. One extreme solution then is to include all data in the fit and attempt to jointly maximize the likelihood for *all* sources in the sky. At this juncture, such an approach is computationally infeasible. A less extreme version of this solution would employ an ROI sufficiently large that these "edge effects" become less important. However, using too large of an ROI brings its own drawbacks:

- the algorithms for convolution, likelihood evaluation, and other tasks have a complexity $\mathcal{O}(R_{ROI}^2)$ as well as additional overhead for large ROIs;

- the plate carée projection begins to show distortion for large ROI radius;

- scaling parameters for the diffuse model become less effective for correcting local deficiencies.

Another approach is an iterative "bootstrap" analysis in which a set of ROIs are fit sequentially with the hope of convergence after a few iterations. On the initial iteration(s), only the brightest sources—the most statistically independent and influential[7]—are fit, along

---

[7]*Fermi*-LAT sources span about four orders of magnitude in integral photon flux, from a handful of sources like the Vela pulsar with fluxes of order 10^{-5}ph cm^{-2} s^{-1} to myriad high-Galactic-latitude sources on the detection threshold with fluxes of order 10^{-9}ph cm^{-2} s^{-1} . Sources in the top decade or so of flux utterly dominate the dimmer sources from a photon count standpoint, and (a) can be fit sufficiently without accounting for dimmer sources (b) must be sufficiently fit in order to fit—or even detect—dimmer sources.

with parameters for the diffuse background. In later iterations, the flux threshold is relaxed. With each iteration, the "external" background for an ROI in principle becomes better modeled, and the fits for sources within an ROI improve accordingly.

By combining these approaches—using an iterative, all-sky analysis (ASA) with the largest practical ROI radius—we can hope to obtain parameter estimates close to those that could be obtained from a true joint maximum likelihood fit. Such an all-sky model, beyond the intrinsic interest of characterizing the GeV sky—is an ingredient *sine qua non* for more complex analyses to be discussed in the sequel. Below, we outline a particular implementation using *pointlike*.

6.2.1   ROI Prescription

In such an all-sky analysis, there are three initial decisions to make:

1. the geometry of the ROI (circular, square, other). The infrastructure in *pointlike* is designed around a circular aperture. This geometry maximizes symmetry (useful for calculating, e.g., PSF overlap integrals) and makes data extraction particularly simple. However, it is impossible to tessellate the sky with circles, i.e., our ROIs must overlap, and we must then adopt a prescription for dealing with the same point source appearing in several ROIs as well as address a similar problem with the degrees of freedom of the diffuse background. On the other hand, if we use a scheme like HEALPix to select data, we *can* tessellate the sky, achieving statistically independent ROIs.

2. the size of the ROI. As discussed above, the size of the ROI determines the magnitude of "edge effects".

3. the (possibly proper) subset of the ROI space within which point sources will be fit, i.e., the method for dealing with ROI overlaps.

In light of these considerations, we adopt the following prescription:

N_{side}	R_{ROI} (°)	Fraction ROI within HEALPix
3	15	0.54
4	12	0.48
6	10	0.30

Table 6.3: The fraction of the photons in a circular ROI with radius R_{ROI} contained within a HEALPix (sharing the same center as the ROI) of size $4\pi/(12N_{\text{side}}^2)$ radians.

1. Tessellate the sky using HEALPix pixels with $N_{\text{side}} = 4$. These pixels have an area of $(14.7°)^2$. Although they are not regular in shape, in discussion we can approximate them as squares about 15° on a side. There are 192 such pixels.

2. About the center of each pixel, extract data using a circular aperture with a radius of $R_{ROI} \equiv 12°$. The radius must satisfy the constraint that the pixel be entirely encircled, roughly that $R_{ROI} > 30\sqrt{6/\pi}/N_{\text{side}} \approx 41°/N_{\text{side}}$.

3. Perform a maximum likelihood analysis in which point sources within the boundary of the HEALPix are allowed to vary while all sources lying in other HEALPix are kept fixed.

These choices of N_{side} and ROI radius represent a compromise between overall ROI size and the ratio of photons within the HEALPix pixel to the total number of photons, in this case 0.48. (A ratio of one gives statistically independent ROI. Since each HEALPix has 7 or 8 neighbors, to first order, a given pair of pixels have $< 10\%$ of their photons in common.)

Other choices are presented in Table 6.3. Although an even larger (15°) ROI allows reduced redundancy and overlap, the background convolution begins to become problematic with an ROI of thise size.

6.2.2 Iteration Prescription

In order to allow the array of ROIs to "relax" to an approximately global ML solution, we proceed via iteration. In each iteration, we allow all diffuse source parameters to vary, while
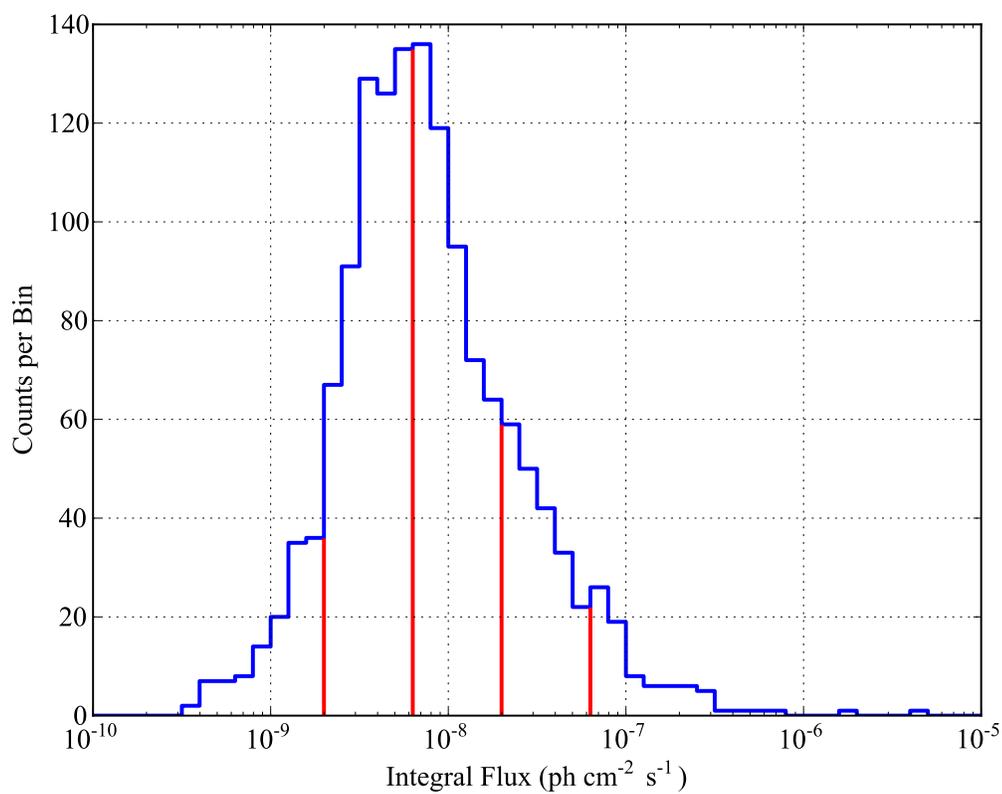
Figure 6.3: The integral photon flux of 1FGL sources above 300 MeV. The vertical lines indicate the flux thresholds in the iterative all-sky fit.

only a fraction of point sources above a certain flux threshold are allowed to vary. We find that the integral photon flux above 300 MeV is a good indicator of source significance and we define the quantity $F_{300} \equiv \int\limits_{300}^{\infty} dE\, dN/dE$ for applying flux cuts. This quantity for all sources in the 1FGL catalog is shown in Figure 6.2.2. Concretely, the iteration strategy is:

1. Perform two fits in which only point sources with $F_{300} \geq 10^{-7.2}$ are allowed to vary. In Figure 6.2.2, this cut includes all sources to the right of the rightmost red horizontal line, essentially the hundred brightest point sources. These sources are sufficiently strong to alter the diffuse background model parameters but are essentially independent of weaker sources. These two iterations provide a "baseline" model with which to fit additional point sources.

2. Perform three iterations in which the flux threshold is lowered by one half of a decade each time. These cuts are also indicated in Figure 6.2.2.

3. Perform three iterations in which the flux threshold is set to 0, i.e., all point source parameters are allowed to vary.

Additionally, at each iteration, after performing an initial ML fit for the ROI, we check the free point source parameters for unphysical values, e.g., a photon flux consistent with zero or a photon index lower than 0.5 or higher than 4.0. If we find such sources, we reset their parameters to a nominal value and refit the ROI.

Finally, we note that performing all 8 iterations for all ROIs requires on the order of 100 CPU hours on modern machines, a small cost on a modest computing clusters.

6.2.3 Tests for Convergence

To test the convergence of the above iteration scheme, we track the change in the log likelihood for each ROI from iteration to iteration. As an example, we consider an ASA performed according to the standard iteration prescription using 18 months of data to which standard cuts[8] were applied. The source model was taken to be that of the 1FGL

---

[8]Incidence angle $\leq 66.4°$, zenith angle $\leq 105°$.

analysis, i.e., for point sources the 1FGL catalog and for diffuse sources *gll_iem_v02* and *isotropic_iem_v02*. The log likelihood changes for the eight iterations are shown in Figure 6.2.3 and indicate the final iteration is well-converged.

Particularly for ROIs with relatively large overlaps with neighbors (Table 6.3 e.g.), total convergence is rather difficult to achieve. With each iteration, a pair of coupled ROIs will update their parameters to reflect the previous iterations background model, and this oscilllation can proceed for many iterations. The magnitude of the log likelihood shift is typically of order unity, and we regard these shifts as negligible. An approach that will in principle eliminate this issue uses the same HEALPix both for organizing free point sources *and* extracting data, i.e., it abandons circular ROIs.

6.2.4   Results

As an example, we show a comparison of the parameters obtained with the *pointlike* ASA and those obtained by *gtlike* as presented in the 1FGL catalog[12]. We apply the iteration prescription to the same data set outlined in [12] but performing our own standard cuts (see previous section) and calculating the livetime with the *pointlike* package.[9] In some sense, this exercise is only a sanity check, but we emphasize these results are obtained with an entirely independent toolchain. We compare the central values obtained for the power law parameters for all 1451 1FGL sources in Figure 6.2.4 and find good agreement. A detailed analysis in Figure 6.2.4 find good agreement between the error-weighted parameter estimate differences[10] and the uncertainty estimates.

6.2.5   Diffuse Source Studies

As an additional validation of the treatment of diffuse sources presented in Chapter 4, we take advantage of the ASA infrastructure to obtain fits to diffuse scaling models and verify they are consistent with the input. We used 18 months of data generated by *gtobssim* for the *gll_iem_v02* Galactic diffuse and *isotropic_iem_v02* isotropic diffuse sources and ran a

---

[9]The livetime as calculated by the Science Tool application *gtltcube* differs slightly from the *pointlike* application, as the livetime calculation in *pointlike* corrects for the cuts on zenith and incidence angle.

[10]The curious offset of the *pointlike* fluxes of about 1/2 an error unit is under investigation.
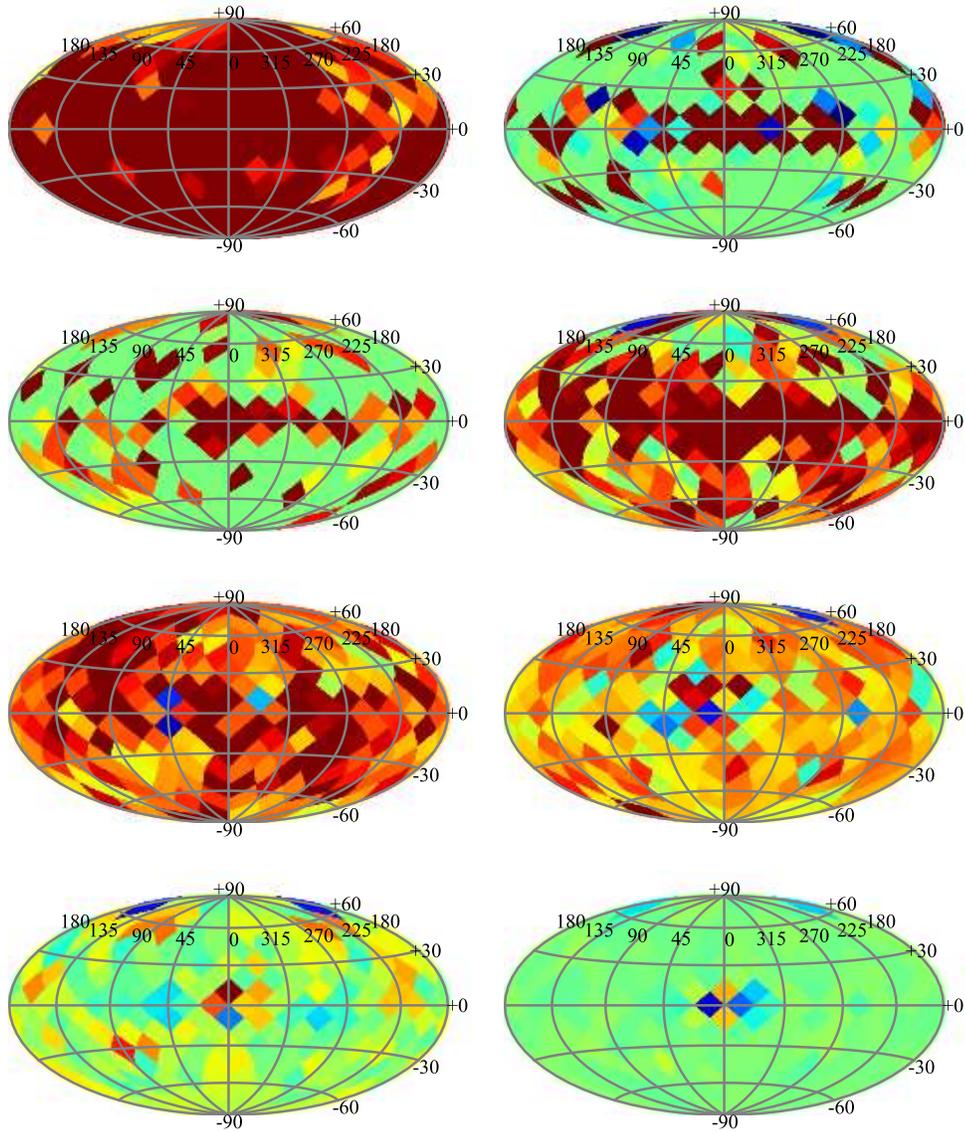
Figure 6.4: The log likelihood changes in each ROI for each iteration. The iterations proceed from left to right and top to bottom. The precise quantity shown is $\text{sign}(\Delta \log \mathcal{L}) \times \sqrt{(\Delta \log \mathcal{L})}$, and the scale extends from -5 (dark blue) to 5 (dark red). For these large ROIs, a change in log likelihood of order a few is negligible. In the upper lefthand corner—the initial iteration—the log likelihood shift is dominated by the initial fit of the parameters for the diffuse models. The third through sixth panels show the log likelihood improving as more sources are allowed to vary. The brightest sources are concentrated in the Galactic plane (selection and projection effects), so only by the fifth panel are a significant fraction of sources at all latitudes free. The final three panels show the refinement with all point sources varying. In the final panel, essentially all ROIs have ceased to improve. Four pixels at the Galactic center continue to shift by order a few in an "oscillation" described in the main text.
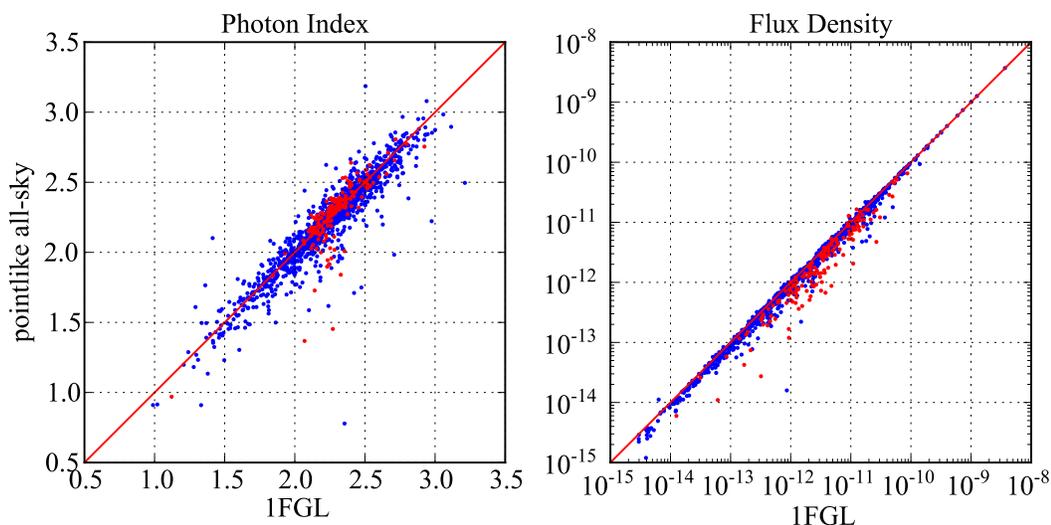
Figure 6.5: A comparison of the ML estimates obtained by the *pointlike* ASA and the unbinned likelihood pipeline analysis of the 1FGL catalog. All sources are fit with a power law. The flux density is evaluated at the "pivot energy", an estimate of the energy at which the covariance of the photon index and flux density is minimized. Sources in red correspond to 1FGL sources with a "c" appended, indicating they lie along the Galactic ridge and their spectral values should be taken with caution. Many of these sources may be spurious. A subset of these sources clearly departs from the main population for which *pointlike* and the 1FGL values are in excellent agreement.
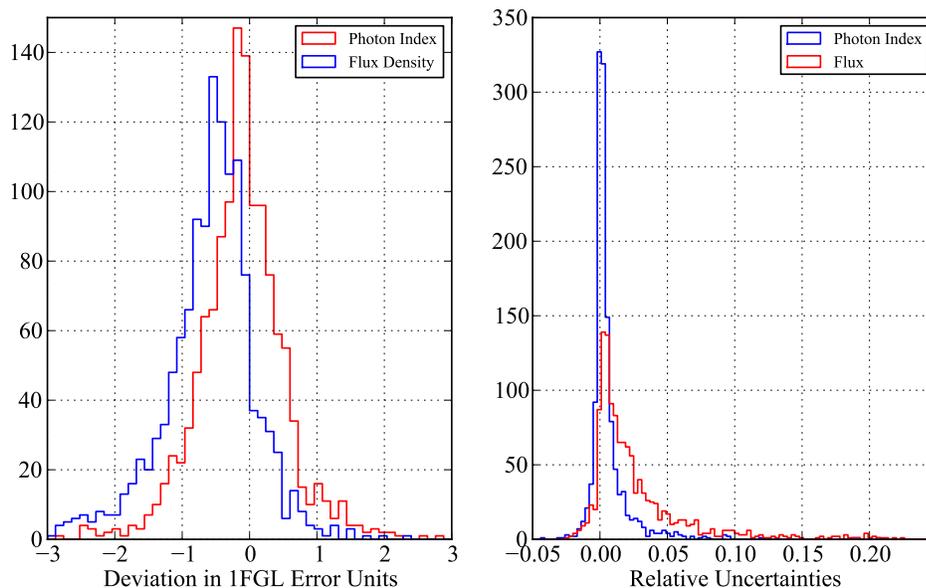
Figure 6.6: In the lefthand panel, the difference in the ML parameter estimates scaled by the parameter error estimated by the 1FGL analysis. The photon indices are in excellent agreement, while the flux densities shows a small trend toward underestimating the flux. In the righthand panel, we compare the uncertainty estimated by each analysis. This comparison is made by estimating the relative error for each method and then taking the difference (1FGL - *pointlike* ASA). The error estimated by *pointlike* is generally in excellent agreement, with some flux error estimates exceeding the 1FGL estimate by up to 20%, an unimportant difference. It is in any case not surprising that an unbinned method may deliver slightly smaller error estimates.

single iteration of the ASA procedure. The *gll_iem_v02* was scaled by a power law with free normalization and photon index, while the *isotropic_iem_v02* source was scaled only by a constant. During a ML fit, these parameters vary due to statistical fluctuations, but the distribution means should be consistent with the input model. The distribution of the best-fit parameters for each ROI is shown in Figure 6.2.5, from which it is clear that the fit parameters indeed agree with the input model.

6.3   Sky Maps

In this section, we deviate from likelihood analysis (but not from all-sky analysis!) to discuss depiction of *Fermi*-LAT data. Constructing maps that give a good visual representation of the sky is challenging because (a) low energy events are significantly smeared by the PSF and (b) the emission is highly anisotropic, peaking strongly in the Galactic plane. Constructing a simple counts map—a histogram—with a fixed pixel size is not optimal since we will either lose resolution in high-count regions or suffer from "shot noise" in low-count regions.

A well-known alternative to a histogram for visual representation of random data is *kernel density estimation*. In one dimension, e.g., the a kernel density estimator $\hat{f}$ for the probability density function f given a set of data $\{x_i\}$, is

$$\hat{f}(x) = \frac{1}{N} \sum_{i=1}^{N} g[(x - x_i)/b] \tag{6.2}$$

where g is the *kernel*, some normalized density function, typically a gaussian, and b is the *bandwidth*, essentially a smoothing parameter.

An obvious extension to the current use case is adopting the PSF for the kernel, with a bandwidth then naturally parameterized by the reconstructed energy and event class. As part of the all-sky analysis, we have already constructed for each ROI, in each energy band, an approximate, exposure-weighted PSF, as well as binned the data in a sparse format that eases the computation. We can thus calculate a kernel density estimate for the photon density on the sky as

$$f(\vec{\Omega}) = \sum_{i=1}^{12N_{\text{side}}^2} \sum_{j=1}^{N_{bands}} \sum_{k=1}^{N_{pix}} n_k \, f_{wpsf}(i, j, |\vec{\Omega} - \vec{\Omega}_k|), \tag{6.3}$$
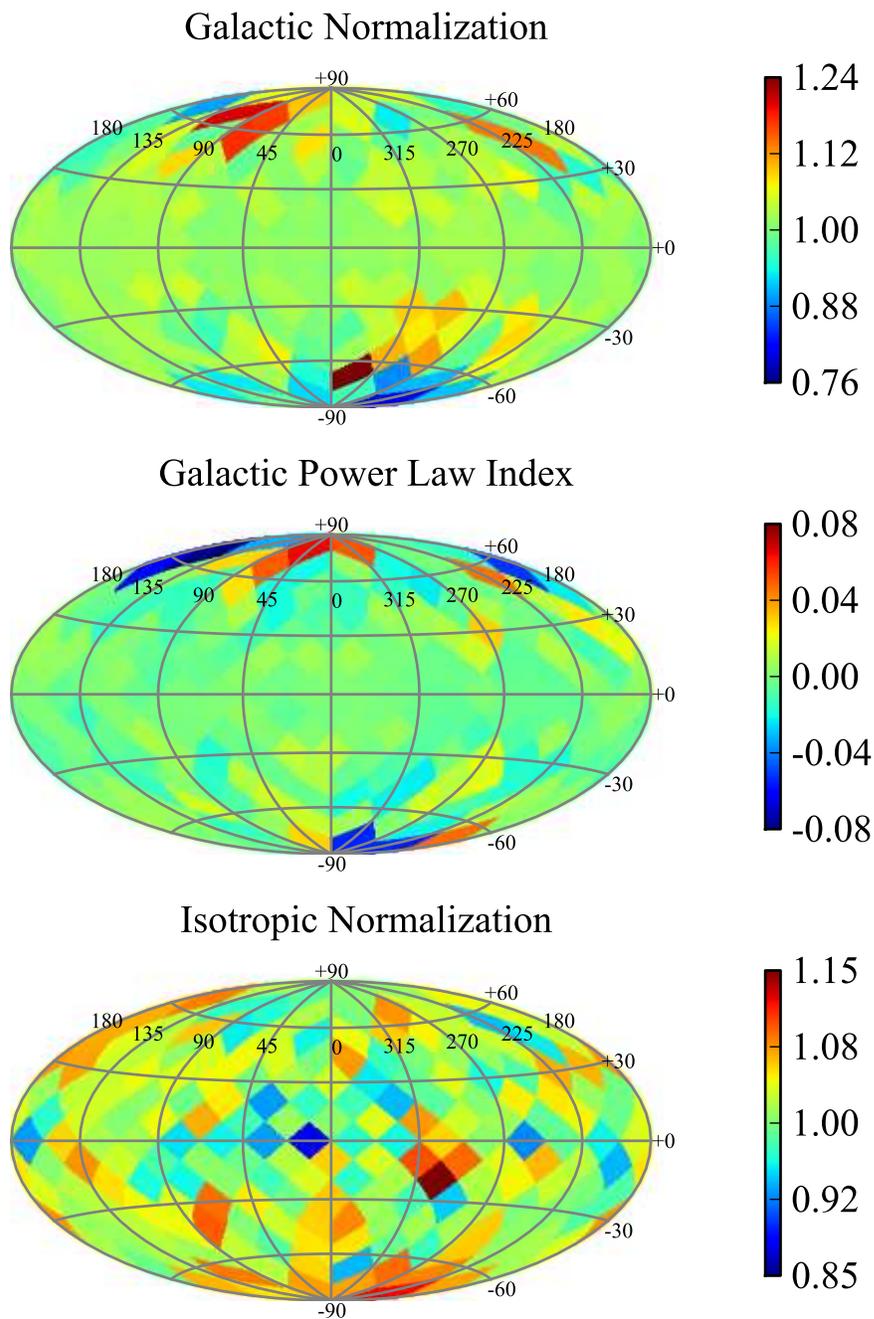
Figure 6.7: The best-fit values of the parameters for the two scaling models discussed in the text. The null value for the constants is 1, while for the photon index of the power law scaling the Galactic diffuse, 0. The abundant photons in the Galactic plane strongly constrain the Galactic diffuse model, while at high latitudes the sparse photons allow large fluctuations. The isotropic background is generally constrained to within 10% independent
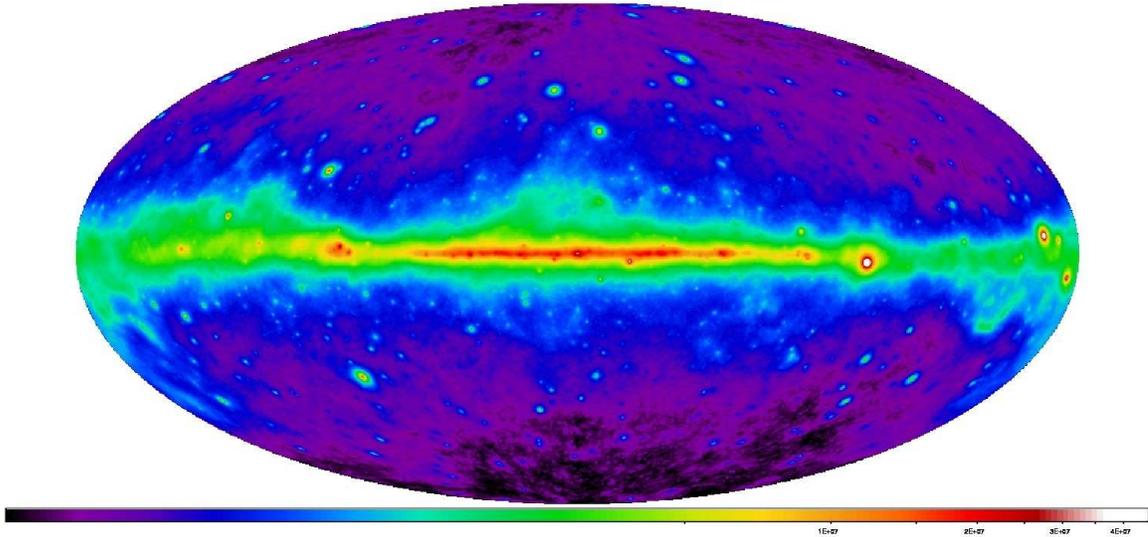
Figure 6.8: An image (in Hammer-Aitoff projection of Galactic coordinates) of the GeV sky produced by kernel density estimation with the PSF with 18 months of data. The units are somewhat arbitrary but can be roughly interpreted as counts per steradian. All energies > 100 MeV are included.

where we symbolically indicate the PSF dependence on energy and conversion type through the band subscript j and where n_k gives the observed counts in the kth pixel. In practice, we would evaluate $f(\vec{\Omega})$ over some grid corresponding to an image projection.

We show such an image for the entire γ-ray sky in Figure 6.3. An image restricted to the Cygnus region of the sky, and to relatively high-energy photons, is displayed in Figure 6.3. It is useful to compare this image, particularly the evidence for the new source mentioned in the discussion of the spectral analysis of the Cygns region, with the series of TS maps in Figure 6.4.2.

6.4   Source Finding with TS Maps

An important application of likelihood is searching for new sources. By examining how the likelihood improves by expanding the model to include new sources, we can determine (in a statistically calibrated fashion, as we show below) whether the sources are required by the data.
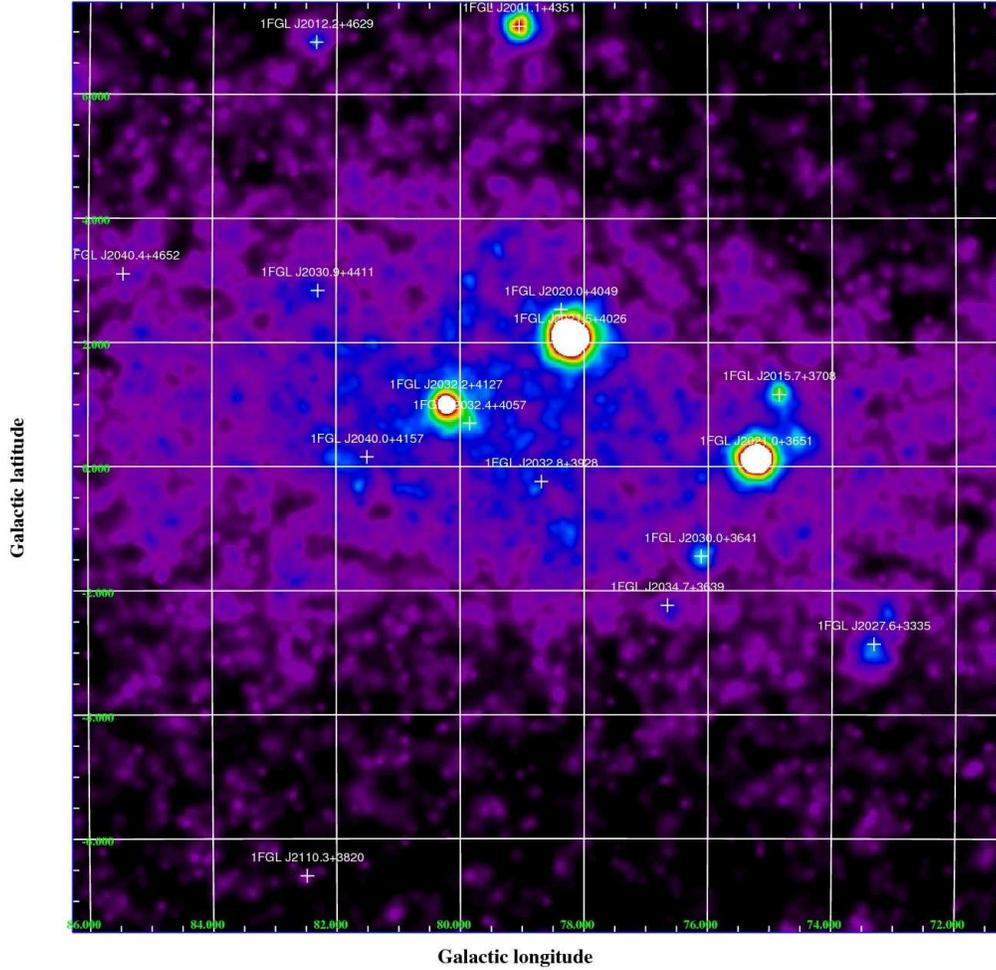
Figure 6.9: An image (in zenithal equal-area projection of Galactic coordinates) of the GeV sky produced by kernel density estimation with the PSF with 18 months of data. The image shows the Cygnus region and has overlaid the locations of sources in the 1FGL catalog. In constructing this image, photon energies were restricted to $2200 < E/MeV < 10000$ ($4400 < E/MeV < 20000$) for events converting in the front (back) of the detector. Due to the $\approx 1/E$ dependence of multiple scattering and the factor of ≈ 2 difference in the radiation length of the front and back radiations, the two samples of events have approximately the same angular resolution.

6.4.1   A Test Statistic for Source Detection

The typical technique is a likelihood ratio test in which the ratio of the likelihood under the null hypothesis (no new source) is compared to the likelihood under the alternative hypothesis (there is a new source with some position/spectrum). The likelihood ratio is a *test statistic* (TS) and we reject the null hypothesis (claim detection of a new source) if the TS exceeds some threshold.

To ensure a small probability of Type I error (false positive), it is useful to know the distribution of the TS in the event the null hypothesis is true. Then we immediately know our probability of Type I error is the tail probability of the distribution integrated from the threshold. (If we test multiple, independent data sets, or check the same set for multiple sources, we must of course multiply this probability by the number of total trials.) A celebrated result from Wilks[84] gives the asymptotic (i.e., large sample) distribution for a certain class of likelihood ratios. If the models for the null and alternative hypotheses are nested and have, respectively, n_0 and n parameters, and if the null values do not fall on the boundary of the parameter space, and if the likelihood satisfies certain regularity conditions, then the likelihood ratio is distributed as $\chi^2_{n-n_0}$, i.e., chi-squared with degrees of freedom equal to the difference in the dimensions of the parameter spaces. This result is extremely useful for testing the statistical significance of additional components in models *provided the new component is not on the boundary of the parameter space in the null hypothesis.*

Unfortunately, the likelihood ratio test for source detection does not satisfy the criteria for Wilks' Theorem. For example, suppose the likelihood for a data set in the presence of an additional point source at a fixed position with a fixed spectral shape but free flux parameter is calculated. The models (with and without the new point source) are clearly nested, but the new parameter—the flux—is 0 in the null hypothesis. Since flux is positive, this value lies on the boundary of the parameter space.

Fortunately, Chernoff[35] was able to extend the results of Wilks to show that, in cases where the alternative (null) model has a n $(n-1)$ dimensional parameter space, and the additional parameter is resticted to one side of an $(n-1)$ dimensional plane—as is the case for the scenario outlined above—then the null distribution is an equal mixture of $\delta(0)$ and

χ_1^2. It is easy to show then that the chance probability of observing a value in excess of TS is given by the tail integral of a normal distribution from $\sqrt{TS}$ to ∞, and the confidence level can be quoted as "$\sqrt{TS}\sigma$". Analysis of EGRET data, applying the same test for new sources, observed this distribution in Monte Carlo simulations[66].

6.4.2 Generating a TS Map

To generate a TS map, we begin with a single region, viz. one of the HEALPix/circular aperture pieces of the ASA. After the ASA, we have a consistent model that should—if we have represented every source correctly—account for every photon in the GeV sky. Next, we subdivide the HEALPix pixel into sub-pixels small enough that we can resolve new sources. For $N_{\text{side}} = 4$, we might choose to divide each side of the base HEALPix pixel into 150 segments, for 150^2 sub-pixels approximatly $0.1°$ on a side. (It will be recalled that this procedure is the typical method for producing a HEALPix pixel of finer resolution; this example grid has $N_{\text{side}} = 600$.)

We regard the background model as fixed and we assume a power law model with $\Gamma = 2.0$ for the putative point source. (Other indices can be adopted.) For the pixel at position $\vec{\Omega}_i$, the log likelihood for a given band (using the notation of Chapter 4) in the presence of a new source with flux $\mathcal{F}_s$ is

$$\mathcal{L}(\vec{\Omega}_i, \mathcal{F}_s) = -\mathcal{O}_s(\vec{\Omega}_i)\, N_s(\mathcal{F}_s) + \vec{n} \cdot \log\left[1 + N_s(\mathcal{F}_s)\frac{\vec{f}_s(\vec{\Omega}_i)}{\vec{b}_s}\right], \qquad (6.4)$$

where $\vec{b}$ represents the contribution of the background to each data pixel and $\vec{f}$ is the value of the PSF for a source centered at $\vec{\Omega}_i$ at each data pixel. The likelihood is single-dimensional (by virtue of the fixed photon index) and can be easily maximized to determine the best-fit $\mathcal{F}_s$ (denoted $\hat{\mathcal{F}}_s$ for each $\vec{\Omega}_i$. The TS is twice the quantity obtained by subtracting the null log likelihood ($\mathcal{F}_s = 0$) from the log likelihood evaluated at $\hat{\mathcal{F}}_s$. As defined above, $\mathcal{L}(\vec{\Omega}_i, 0) = 0$, so

$$TS(\vec{\Omega}_i) = 2 \times \mathcal{L}(\vec{\Omega}_i, \hat{\mathcal{F}}_s). \qquad (6.5)$$

We determine this quantity for each sub-pixel in each base ($N_{\text{side}} = 4$) HEALPix pixel. Since the background model is assumed to be fixed, the base pixels can be processed in

parallel. The output is the TS evaluated on a very fine grid, producing a TS map for the entire sky. It is then a relatively simple matter to "convert" these TS values to new sources by selecting the highest, isolated TS excesses, incorporating them into the model, re-maximizing the likelihood via the ASA, crafting a new TS map, etc.

As an example of the efficacy of TS maps, we present a case example of searching for a source in the Cygnus region in Figure 6.4.2.

6.5   Source Classification

It has historically been the case that detection of γ-ray pulsations from neutron stars follows detection of radio pulsations[82, 81, 8] or even X-ray pulsations[55, 26]. This is due to the extreme paucity of photons collected from HE sources; long integration times are required, and even though pulsars are relatively stable rotators, the phase space in frequency, frequency derivative, etc. that must be searched has made direct pulsation detection infeasible. Instead, radio observations of a particular pulsar are used to build a timing solution (see Chapter 7) that can be used to map the time of arrival of a detected γ ray to the rotational phase of the neutron star. With a stable timing solution, photons collected over weeks, months, or even years can be accumulated, building up profiles in phase that can be tested against a uniform distribution (Chapter 7).

With the improved angular resolution and effective area of *Fermi*-LAT , GeV telescopes have finally reached some parity. Time-differencing techniques[20] have led to the discovery of over 24 new pulsars[3, 75], wrapping up two longstanding mysteries[13, 1] on the way. This number is comparable to the number of new γ-ray pulsars detected to date by *Fermi*-LAT using timing solutions[17].

Besides direct detection of pulsations, "γ-ray initiated" detection of pulsars is opened by the sensitivity to *unpulsed* sources. The high yield so far of detected pulsars suggests many of the > 1000 sources detected by *Fermi*-LAT so far[12] must harbor neutron stars emitting pulsed γ rays. There are many reasons why such sources may be difficult to detect with direct pulsations searches. Many young pulsars exhibit stochastic timing noise[19], limiting the maximum integration period over which a detection may be obtained. Even if the pulsar is stable, long integration times require fine-grained searches and the possible
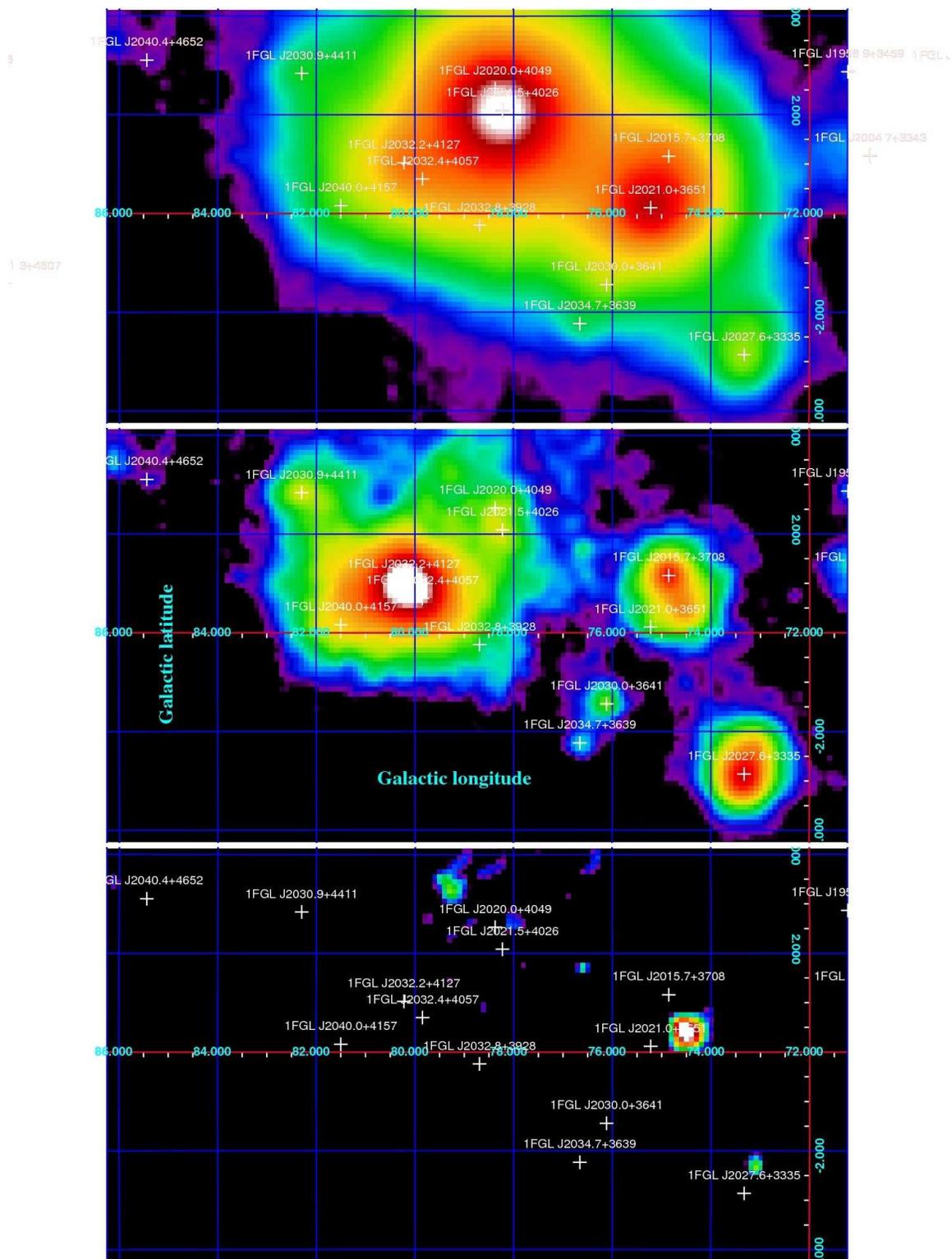
Figure 6.10: A TS map constructed of the Cygnus region to demonstrate the source finding principle discussed in the text. In the top panel, only the best-fit diffuse model (Galactic + isotropic) is used for a background model. All point sources then contribute to the calculated TS, and the map is dominated by the exceedingly bright pulsars J2021+4026 and J2021+3651. The range spanned by the color scale is 25–40,000. In the second panel, the two bright pulsars are included in the background, and the color scale is now 25–2200,

addition of additional search parameters (e.g., a second-order term in the Taylor expansion of phase-versus-time, i.e., a second derivative of frequency.) Integration times comparable to a year also require a precise position to correct for the changing light travel time (the Römer delay) as the earth revolves about the sun. Unfortunately, dim sources (requiring long integrations) are also difficult to localize and the problem becomes rapidly intractable. Finally, millisecond pulsars are difficult to detect because of the extreme sensitivity to precise values of period and period derivative, and millisecond pulsars in binary systems are essentially undetectable by *Fermi*-LAT .

One approach, then, is to search for *radio* pulsations from promising *Fermi*-LAT sources. The dimmest known pulsars can be detected in hours with modern radio telescopes, solving most of the problems outlined above. For this approach to fail (a) the source is not a pulsating neutron star (b) the position is not good (c) the pulsar radio beam is not aligned with the earth (d) the signal was temporarily suppressed by scintillation. If pulsations *are* detected, we are essentially guaranteed both a new radio pulsar *and* a new γ-ray pulsar, since the chance coincidence of radio pulsars and LAT sources is quite low.

What constitutes a promising *Fermi*-LAT source? A good candidate should

1. have a reliable position with an uncertainty comparable to or smaller than the main beam of the radiotelescope used to conduct the search. For the 100-meter Green Bank telescope, this constraint ranges from a few arcminutes at 2GHz frequencies to up to $0.3°$ (error radius) at 350MHz. Sources within the Galactic plane suffer some additional systematic errors (e.g. unmodeled neighboring point sources, mismodeled diffuse emission).

2. be unassociated with an extragalactic source. By comparing the positions of *Fermi*-LAT sources with multiwavelength catalogs, probabilistic associations of γ-ray sources with counterparts can be made[12, 16]. If a source has a good probability to be associated with another source class (particularly active galactic nuclei (AGN), the largest class of *Fermi*-LAT sources), then it is sensible to pursue better candidates.

3. show no variability.

4. have a spectrum (a) like known pulsars and (b) unlike known AGN. Essentially all pulsars detected to date have spectra consistent with a power law + exponential cutoff and have a constant flux when averaged over many periods[17]. In particular, the observed cutoff energy (photon index) is generally ≤ 5 GeV (≤ 2). On the other hand, AGN have spectra broadly consistent with power laws. Flat-spectrum radio quasars, a subclass of AGN, almost universally have $\Gamma > 2$, and thus a candidate can be spectrally vetoed by observation of a falling spectral energy density from $0.1 - 1.0$ GeV. BL Lac objects, a second subclass with harder spectra, can generally be excluded by observation of significant emission above 10 GeV.

Thus, we seek to analyze *all* known *Fermi*-LAT sources with the aim of (a) estimating the best possible positions and uncertainties (b) excluding known variable and known associated sources (c) fitting multiple spectra to assess the statistical strength of the exponential cutoff parameter (d) producing a spectral energy density plot to visually assess spectral features. The all-sky analysis discussed above can handily achieve (a), (c), and (d). We outline a specific analysis scheme along with radio surveys conducted to search the resulting candidates and present preliminary results.

6.5.1 All-sky Analysis for Pulsar Candidate Identification

We performed an all-sky analysis as described above (in a more preliminary version) using 11 months of *Fermi*-LAT data and an internal (to the LAT Collaboration) version of the source list that became the 1FGL catalog. (The 1FGL catalog was based on the same 11-month dataset employed here.) To supplement the analysis, we made use of the association and variability studies performed for the catalog preparation. For each source, we fit a power law spectrum, then added a cutoff, allowing a likelihood ratio test of the significance of the cutoff. We generated a spectral energy density plot for each source to assess its shape qualitatively. An example of these seds—which include also the other pieces of information save position—appears in Figure 6.5.1.
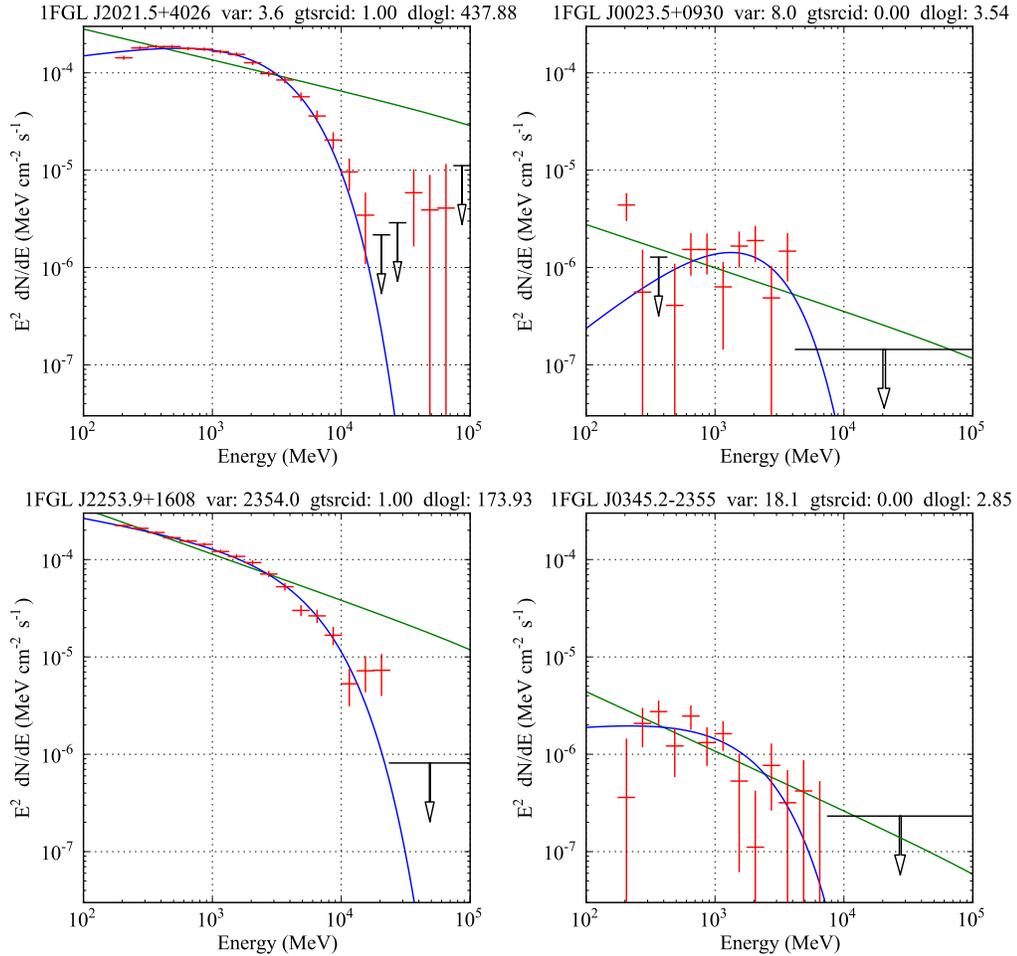
Figure 6.11: A gallery of typical plots used for classifying unidentified sources as good pulsar candidates. In the upper left panel, a known LAT PSR J2021.5+4026[3] gives an example of a typical pulsar spectrum: flat or rising in the $0.1 - 1$ GeV decade, a strong cutoff, and a low "variability index". In the upper right panel, a much dimmer pulsar candidate (at which location radio pulsations were subsequently detected), which shows similar features scaled down by two orders of magnitude. At bottom left is the quasar 3C 454.3[24] exemplifying the features of typical AGN emission: a falling spectrum in the $0.1 - 1$ decade and extreme variability. At bottom right, a comparable (unidentified) source showing similar features scaled down by two orders of magnitude in flux. It has modest variability and would not be considered a good candidate due to this and its falling spectrum at low energy.

6.5.2   Observations

Green Bank Telescope, 350 MHz

A Fermi Cycle 2 Guest Investigator[11] proposal (P.I. Mallory Roberts) to observe approximately 50 unidentified Fermi sources at high Galactic latitudes with the 350 MHz receiver of the 100-m Green Bank Telescope (GBT) was accepted in 2009. Since the spectrum of pulsars in radio is typically a falling power law, observations at low frequency are more sensitive provided there are not significant disperserive effects from the interstellar medium (ISM). By restricting candidates to high Galactic latitudes, ISM effects are minimized.

The ASA preparation and classification scheme was used to draw up a candidate list. An initial pass rejected any candidates at low ($< -5°$ or $> 5°$) Galactic latitude and with Decl. $< -40°$, essentially the lowest latitude visible from GBT. Since the full-width at half-maximum (FWHM) of the 350 MHz beam at GBT is about $0.6°$[12], no cut was made on the positional uncertainty candidate sources as virtually all sources in the 1FGL catalog have a 95% error radius $< 0.3°$. For sources satisfying the position consitraints, plots such as those presented in Figure 6.5.1 were generated for the remaining sources. By considering spectral shape, cutoff strength, association with known blazars, and variability (at time scales of ≈ 1 month), sources were classified as 0, 1, 2, or 3, viz. unacceptable, poor, good, and excellent.

Observations were carried out using the GUPPI[44] backend with the 350 MHz receiver. During the campaign, most sources classifed as "2" or "3" were visible and observed. In total, 47 sources were observed with a typical integration time of 32 minutes. The collected data requires considerable computation to fully search for pulsations[13] and have not yet been fully analyzed. However, the first 215 seconds of each observation have been fully searched, and five new MSPs have been detected. A brief summary of their properties is

---

[11]http://fermi.gsfc.nasa.gov/ssc/proposals/cycle2/

[12]To order of magnitude, the FWHM is given by the diffraction limit, FHWM $\approx \lambda/D \approx 0.5°$, with $D \equiv 100$m. The actual beam depends on the exact geometry of the receiver and the blazing of the dish surface.

[13]E.g., many millisecond pulsars will be found in binary systems, requiring an "acceleration" search in which a linear correction for gravitational acceleration of a given magnitude is made.

Name	NS Period (ms)	Orbital Period (hr)	Companion Mass (M_{sun})
J0023+09	3.05	3.3	0.016
J0340+41	3.30	Isolated	N/A
J1302-32	3.77	>24	N/A
J1810+17	1.66	4.9	0.089
J2215+51	2.61	4.2	0.22

Table 6.4: The properties of the 5 millisecond pulsars detected to date in the GBT 350 MHz survey. The binary J1302-32 awaits further timing to constrain the orbital parameters. PSRs J0023+09 and J1810+17 are likely of the "Black Widow" class[50], while PSR J2215+51 presents eclipses.

given in Table 6.4[14].

Parkes Observatory, 1400 MHz

A second serendipitous[15] survey was conducted with the 1.4 GHz receiver on the 64-m dish at the Parkes Observatory (P.I. Fernando Camilo). The source list was prepared as described in the previous section, *mutatis mutandis*, e.g. restricting the candidates to have Decl. $< -40°$. A total of 14 sources were observed with integration times ranging from 1 to 2 hours. Five new millisecond pulsars, listed in Table 6.5, were detected.

6.6 Summary

In this chapter, we demonstrated the applicability of *pointlike* to both a restricted analysis (in the sense of a small region with a subset of sources) and to all-sky analysis (ASA). For the former, we carried out a spectral analysis in the Cygnus region, obtaining estimates of the parameters of broadband spectral parameters (§6.1.3), source positions (§6.1.4), and model-independent estimates of source spectra (§6.1.5). An exploratory approach was crucial to the

---

[14]After a poster presented by P. Bangale at the 2010 meeting of the High-energy Astrophysics Division of the AAS.

[15]Observations were carried out during scheduled but unutilized maintenance time.

Name	Classification
J0100-64	Bi
J1514-49	Bi
J1658-53	Iso
J1747-40	Iso
J1902-51	Bi

Table 6.5: The five millisecond pulsars discovered in a serendipitous Parkes survey.

success of this analysis as we observed the interplay of the bright source PSR J2021+3651 with two dim, neighboring sources.

Next, we outlined an approach for ASA in which we can obtain a good estimate for the spectral parameters (and positions) of all (known) sources in the sky. The method adopted—organizing sources by HEALPix while extracting data by cone—was a compromise between manageable data sets and optimal estimates for the parameters. We demonstrated that the iterative approach converged, while comparison with the 1FGL catalog indicated good agreement between the two methods, i.e., a valid solution.

Finally, we made use of the model estimated from the all-sky analysis (ASA) for additional applications: we constructed TS maps for source finding using the ASA spectral model as background; we generated a kernel density estimator map of the sky; and we used the ASA pipeline to perform a custom analysis of known sources to identify good pulsar candidates. In the next chapter, we shall make use of ASA-estimated spectra in building more sensitive tests for periodicity detection.

Chapter 7

PROBABILITY-WEIGHTED STATISTICS AND PULSED SENSITIVITY

We discussed in §6.5 the difficulties involved in detecting pulsed signals with HE instruments. There, we focused on how the improved sensitivity of the LAT enabled detection of pulsation using γ rays only, or through an initial detection of a point source with subsequent pulsation detection in a targeted radio search. Here, we focus on the "other side of the coin"—since the LAT boasts much improved sensitivity, by using timing solutions, we should be able to detect periodic emission from many known sources.

And this has turned out to be the case, e.g., the first detection of orbitally modulated HE γ rays from LS I +61° 303 [7], the first detection of pulsed γ rays from a millisecond pulsar [9], and the first new young pulsar detected using a radio timing solution[8].

The pulsars detected so far have been sufficiently bright that relatively simple techniques have sufficed to discern pulsation. We discuss the details fully below, but in essence a set of photons is extracted from the data, the arrival times are converted to phase using the timing solution, and a statistical test for uniformity is performed. For the same reason that aperture photometry is inadequate for spectral analysis, so too is this procedure suboptimal for time-domain analysis: any given aperture that one selects will either be strongly contaminated by other sources or be selected so stringently that most of the signal is removed. In either case, the test results will depend strongly on the cut.

For spectroscopy, we were able to tackle the problem of the multi-scale PSF and source confusion by using likelihood to account for the IRF. However, while we could model nearly any spectrum well with a simple model with only a few parameters, this is not the case for time-domain analysis. Gamma-ray light curves tend to be complex, and more importantly, since we are searching for periodicity, we do not know *ab initio* what the light curve should look like. Therefore, we do not pursue a full likelihood analysis here but shall instead adopt

a hybrid approach in which we combine the results of spectral analysis with simple pulsation tests.

In further distinction to our approach to spectroscopy, we adopt an *unbinned* time-domain analysis. In this way we lose no information and—more importantly—incur no bias from an arbitrary choice of binning[1]. Since we will be using the results of a previous spectral analysis and don't have to worry about jointly fitting sources, we can use a relatively small ($\leq 2°$) ROI. This makes unbinned analysis practical from a computational standpoint.

Specifically, we shall study a class of unbinned statistics composed of sums of sines and cosines of the observed phases. Next, we show how these statistics can be altered to *weighted* sums of sines and cosines. By using the results of a ML spectral analysis, we can calculate the probability that each photon originated from the source we are testing for pulsation, and this probability is an excellent choice of weight. We go on to show that the probability-weighted versions of the statistics drastically reduce the probability of Type I and Type II error through tests on ensembles of simulated pulsars. The machinery that we develop in this section can be easily applied to determine the pulsed sensitivity of the LAT.

Before proceeding to periodicity searches, we discuss the necessary first step of converting the observables—photon times of arrival—into phase of the object of interest.

7.1 Time-to-phase Mapping

A preliminary step in any periodicity search is to map the event times to the appropriate phase, e.g., the rotational phase of a neutron star or the orbital phase of a binary. For instance, the reconstructed time associated with a *Fermi*-LAT event is recorded in *Mission Elapsed Time*, essentially TAI (International Atomic Time) with an origin of January 1, 2001. In order to connect with sources at astronomical distances, the proper reference is time measured at the Solar System barycenter, which eliminates dependence on the (time-dependent) configuration of the Solar System. Using either an exact ephemeris for some astronomical source, or even simply a guess for the period, the barycentered times can be

---

[1]Pulsar light curves can be quite sharply peaked, e.g. P2 of PSR J0030+0451[9], or relatively broad, e.g. PSR J1836+5925[13]. Choosing the "wrong" bin size can drasticaly decrease sensitivity to a class of light curves.

converted to the phase of the distant object. Thus, to convert a time in MET to phase, we must:

1. correct for light travel time, e.g., to the center of the earth, requiring knowledge of the S/C position at the given MET. Small general relativistic effects may also be included.

2. correct for the earth's position as well as other Solar System effects (e.g., if the direction of the source passes near the Sun at the given MET, it suffers a Shapiro delay). The unit of time at the Solar System barycenter is also slightly different from TAI. Light travel time and Doppler shifting from the earth's revolution about the sun are included.

3. correct to the distant object rotational phase. If the object is a member of a binary system, the correction to the distant system barycenter is nontrivial, and must account for Doppler shift due to orbital acceleration, Shapiro delays, and light travel time delays.

For *Fermi*-LAT data, there exists a Science Tool, *gtbary*, to convert MET to TDB (Barycentric Dynamical Time) or to geocentric time, i.e., time measured at the center of the earth. A plugin for *tempo2*[2] will also perform this transformation as well as fold the photons on standard-form ephemerides for pulsars.

For the remainder of this section, the exact procedure we use will be irrelevant, and we will simply assume some black box has produced from a set of event times, $\{t_i\}$, a set of phases, $\{\phi_i\}$. Our convention is to identify all values of phase that are congruent modulo 1, so a sweep of ϕ from 0 to 1 defines a complete cycle.

7.2   Statistics for Pulsation Searches

Suppose that in the null hypothesis—emission from the source in question is not periodic— the $\{\phi_i\}$ have the cumulative distribution function (cdf) $F_\Phi(\phi)$. That is, the probability to observe the random variable (rv) Φ with a value less than or equal to ϕ is $F_\Phi(\phi)$. If the

---

[2]http://www.atnf.csiro.au/research/pulsar/tempo2/

distribution admits a probability density function (pdf), $d/d\phi F_\Phi(\phi)$, we denote this function $f_\Phi(\phi)$. We follow this notation for distributions throughout. Often, the $\{\phi_i\}$ in the null case will be uniformly distributed, $\Phi \sim U[0,1) \rightarrow F_\Phi(\phi) = \phi$. However, the null distribution will *not* remain uniform if the instrument response varies on timescales comparable to the period of interest. Many of the results discussed below only hold under a uniform null distribution. Departure from uniformity is thus a serious complication, though we present two avenues of attack in Appendix B.

The alternative hypothesis specifies another distribution, say $F'_\Phi(\phi)$. Ideally, this distribution is known *a priori*; in practice, we must estimate it, that is, provide a morphology for the light curve. We concentrate here on the class of tests—a subclass of those considered by Beran[25]—that represent the light curve using a Fourier expansion estimated from the data, i.e., the true (unknown) light curve is

$$f(\phi) = 1 + \sum_{k=1}^{\infty} \alpha_k \cos(2\pi k\phi) + \beta_k \sin(2\pi k\phi), \tag{7.1}$$

and we will attempt to estimate the coefficients. In practice, we will truncate this sum at some finite harmonic. (We eschew the complex notation for Fourier series here to avoid the complication of complex rvs. However, for brevity in the sequel, we will refer to the α_k and the β_k collectively as the ψ_k.) We define

$$c_k(\phi) = \cos(2\pi k\,\phi); \tag{7.2}$$

$$s_k(\phi) = \sin(2\pi k\,\phi). \tag{7.3}$$

For brevity we will refer to these variables collectively as G_k. Since Φ is a rv, so are the G_k, although the G_k and are clearly mutually dependent. If $\Phi \sim U[0,1)$, the distribution function for these variables is simple and independent of k: $f_{G_k}(x) = [\pi\sqrt{-x^2}]^{-1}$. We consider the case of nontrivial distributions for Φ in Appendix B. We note in passing the characteristic function for the distribution $f(x) = [\pi\sqrt{1-x^2}]^{-1}$

$$\phi_X(t) = <\exp itx> = \int_{-1}^{1} \frac{dx\exp(itx)}{\pi\sqrt{1-x^2}} = \frac{2}{\pi}\int_{0}^{1}\frac{\cos(tx)}{\sqrt{1-x^2}} = J_0(t), \tag{7.4}$$

where we have appealed to symmetry in the integrand and the definition of the Bessel functions of the first kind. It is easier to calculate the moments directly than to differentiate

the above expression, and we have that the nth absolute moment is given by

$$< |x^a| > = \frac{2}{\pi} \int_0^1 dx \frac{x^a}{\sqrt{1-x^2}} = \frac{1}{\pi} \beta\left(\frac{a+1}{2}, \frac{1}{2}\right) = \frac{1}{\sqrt{\pi}} \frac{\Gamma(\frac{a+1}{2})}{\Gamma(\frac{a+2}{2})}, \tag{7.5}$$

where $\beta(x, y)$ is the Euler beta funciton and $\Gamma(x)$ is the usual Euler gamma function.

Using the new variables, we define the empirical trigonometric moments,

$$\alpha_k = \frac{1}{n} \sum_{i=1}^n c_{ki}; \quad \beta_k = \frac{1}{n} \sum_{i=1}^n s_{ki}, \tag{7.6}$$

where e.g. $c_{ki} = \cos(2\pi \phi_i)$. These quantities are essentially the Monte Carlo estimators for the Fourier coefficients for the kth harmonic. From the expression above for the absolute moments (or simply noting that $G_k \in [-1, 1]$ regardless of the distribution), the mean (μ) and standard deviation (σ) are finite and the central limit theorem (CLT) applies to the ψ_k. That is,

$$\lim_{n \to \infty} \sqrt{n}(\psi_k - \mu_{G_k})/\sigma_{G_k} \sim \mathcal{N}, \tag{7.7}$$

where $\mathcal{N}$ denotes the standard normal distribution. If $\Phi \sim U[0, 1)$, $\mu_{G_k} = 0$ and $\sigma_{G_k} = 1/\sqrt{2}$ (Eq. 7.5), then

$$\lim_{n \to \infty} \sqrt{2n} \, \psi_k \sim \mathcal{N}. \tag{7.8}$$

The third moment of G_k is $4/3\pi$, so the Berry-Esseen inequality gives the rate of convergence as $\propto 1/\sqrt{n}$. Specifically, let $\psi_k(n)$ denote one of the statistics above for a sample size of n. Then

$$\sup_x |F_{\psi_k(n)}(x) - \Phi(x)| \leq \frac{C \, 8 \sqrt{2}}{3 \, \pi \sqrt{n}} \approx \frac{0.847}{\sqrt{n}}, \tag{7.9}$$

i.e., the difference between the cumulative distribution of the statistic and the cdf of the normal distribution, $\Phi(x)$, is absolutely bounded. (Here, we have used $C = 0.7056$, the upper bound determined Shevtsova [76].) In practice, it is not this rate of convergence that will be the limiting step in later statistics we develop. However, before moving on, we mention that an exact expression for $F_{\psi_k(n)}(x)$ may be obtained by inverting the characteristic function noted earlier. Since ψ_k is simply a linear combination of G_k, the characteristic function for ψ_k is given by

$$\phi_{\Psi_k(n)}(t) = J_0 \left(\frac{t}{n}\right)^n. \tag{7.10}$$

Alternatively, if an approximate result, but more accurate than the CLT, is desired, one can use the known moments in an Edgeworth series.

The quantities ψ_k are used in several statistical tests in the literature; see [42] for a brief review. In particular, invariance under rotations of phase become manifest when using these variables. For instance, if we redefine our zero of phase $\phi \to \phi + \delta$, then

$$\begin{pmatrix} \alpha'_k \\ \beta'_k \end{pmatrix} = \begin{bmatrix} \cos(2\pi k\,\delta) & -\sin(2\pi k\,\delta) \\ \sin(2\pi k\,\delta) & \cos(2\pi k\,\delta) \end{bmatrix} \times \begin{pmatrix} \alpha_k \\ \beta_k \end{pmatrix}, \tag{7.11}$$

i.e., a phase shift amounts to a rotation in α/β space for each harmonic, so any statistic involving only functions of the magnitude of these vectors, $\alpha_k^2 + \beta_k^2$, will be independent of δ. The simplest such statistic, a sum of ψ_k^2 for the first m harmonics, has a long history in high-energy gamma-ray observations, and we introduce it below.

7.3   The Z_m^2 Test

The Z_m^2 statistic, defined as

$$Z_m^2 = 2n \sum_{k=1}^{m} \alpha_k^2 + \beta_k^2, \tag{7.12}$$

has been the workhorse of searches for γ-ray pulsars for years. A Z_2^2 test was used in a search for pulsations in COS-B data using timing solutions for 145 radio pulsars [32]. A similar search of EGRET data[69] used Z_m^2 tests with 1, 2, and 10 harmonics, the H test (see below), and the "Z_{2+4}^2" test which is defined as above but with summation restricted to the 2nd and 4th harmonics. It forms an integral part of the H test, and continues to see use in analysis of *Fermi*-LAT data, especially for sources where approximately sinusoidal modulation is expected.

From the discussion above, in the case of uniformly-distributed phases, $\sqrt{2n}\,\psi_k$ asymptotically follows the standard normal distribution. Z_m^2 is thus, asymptotically, distributed as the sum of the squares of $2m$ standard normal variates. As is well known, the square of a normal variate is distributed as χ_1^2, i.e., chi-square with one degree of freedom, and the sum of $2m$ squares of *independent*, normally distributed variates is distributed as χ_{2m}^2, chi-square with $2m$ degrees of freedom. Hence, if and only if these $2m$ ψ_k are statistically independent, then Z_m^2 is asymptotically distributed as χ^2 with $2m$ degrees of freedom. In

addition to this simple null distribution, the test is powerful[42], quick to compute, and has the already-discussed property of invariance under phase translation. It underpins much of the statistical framework we will develop, and so we take some time to assess its properties.

7.3.1 Validity of Asymptotic Calibration

First, there is the point of the asymptotic "independency" of the ψ_k. It is clear that, for only a few phases, these variables are highly dependent. Indeed, for a single observation, ψ_k can be inverted to find find the original rv, ϕ (up to uncertainty from congruence modulo 1), and hence any other ψ_k. However, for the uniform null distribution, this ability to infer the phase from the sum of sines and cosines breaks down with increasing sample size. E.g., if we have observed 10 phases and we have $\alpha_0 = 10$, then we know that $\phi = 0$ for each phase. However, $\alpha_0 = 10$ will almost surely not be observed, whereas $\alpha_0 \approx 0$ is quite likely, and there are many combinations of 10 phases that can yield it, i.e., it tells us very little about the constituent phases. As N increases, the tails are further suppressed, and consequently information about the $\{\phi\}$ and hence the distribution of ψ_i conditioned on the value of ψ_j. It is in this sense and limit that the ψ_k become "independent".

On the other hand, for some null distributions, it is impossible for the ψ_k to become independent, even as $N \to \infty$. A clear albeit unrealistic example is $f(\phi) = \delta(\phi - \phi_0)$, in which case ψ_k remain perfectly correlated. In a more realistic case, any peaking in phase of the null distribution will lead to long-surviving mutual depedence of the ψ_k. Thus, while knowledge of the null distribution will allow us to define ψ_k such that the marginal distributions are still asymptotically normal (i.e., the CLT applies), it will not be the case that the sum of the ψ_k becomes chi-squre distributed. More details can be found in Appendix B.

Related to the question of mutual independence of the ψ_k is the rate of convergence to the asymptotic χ^2 distribution. A Monte Carlo study of the converge as a function of both sample size and maximum harmonic (m) is shown in Figure 7.3.1. Evidently, a sample size of about 50 phases are required for robust significance estimation at the 3σ level. Convergence appears to improve for higher values of m. Interestingly, for low photon counts, low values

of m *underestimate* the significance, leading to an increased probability of making a Type II error. In contrast, high values of m *overestimate* the significance, increasing the chance of making a Type I error.

7.3.2   Modification for Event-weighting

A set of $\{\phi_i\}$ will typically contain background events. As discussed above, for most high-energy instruments, background events make up a significant fraction of the data selected, and may even dominate the signal. Since the significance of a detection scales with the signal-to-noise ratio (SNR), we may try to apply stringent cuts to the data to try to enrich the signal. We don't *a priori* know the optimal cuts, so we must often choose between sub-optimal cuts or tuning the cuts with the data. The former dilutes the power of the test, while the latter increases the risk of a Type I error if we fail to re-calibrate the test statistic, typically with Monte Carlo.

An alternate approach to enriching the SNR is event-weighting: we apply no stringent cuts to the data but instead assign each event a weight. Events associated with the source are weighted heavily, those with the background lightly. Below, we discuss a means of estimating the probability that the event in question was caused by detection of a source photon and we show that using these probabilities as weights is a powerful technique.

To include the weights, we define the weighted empirical trigonometric moments,

$$\alpha_{wk} = \frac{1}{n} \sum_{i=1}^{n} w_i \, c_{ki}; \quad \beta_{wk} = \frac{1}{n} \sum_{i=1}^{n} w_i \, s_{ki},$$

where w_i is the weight assigned to each event. (Here, the c_{ki} and s_{ki} can either be derived directly from the phases or from the cumulative distribution in the case of non-uniformity as outlined in Appendix B.) The weights may themselves be random variables (if determined from some model fit to the data), or they may be simple scalars. If the former, the distribution of the $\{w_i\}$ must possess finite mean and variance since $w \in [0, 1]$. In the null hypothesis, the weights and the phases are independent rvs, so the moments of the products are simply the product of the moments, and it can be shown that the Lyupanov Condition for the CLT to hold will apply for any choice of F_Φ and any such weights distribution. On the other hand, if the weights are taken as simple, non-random numbers, ψ_k is now simply
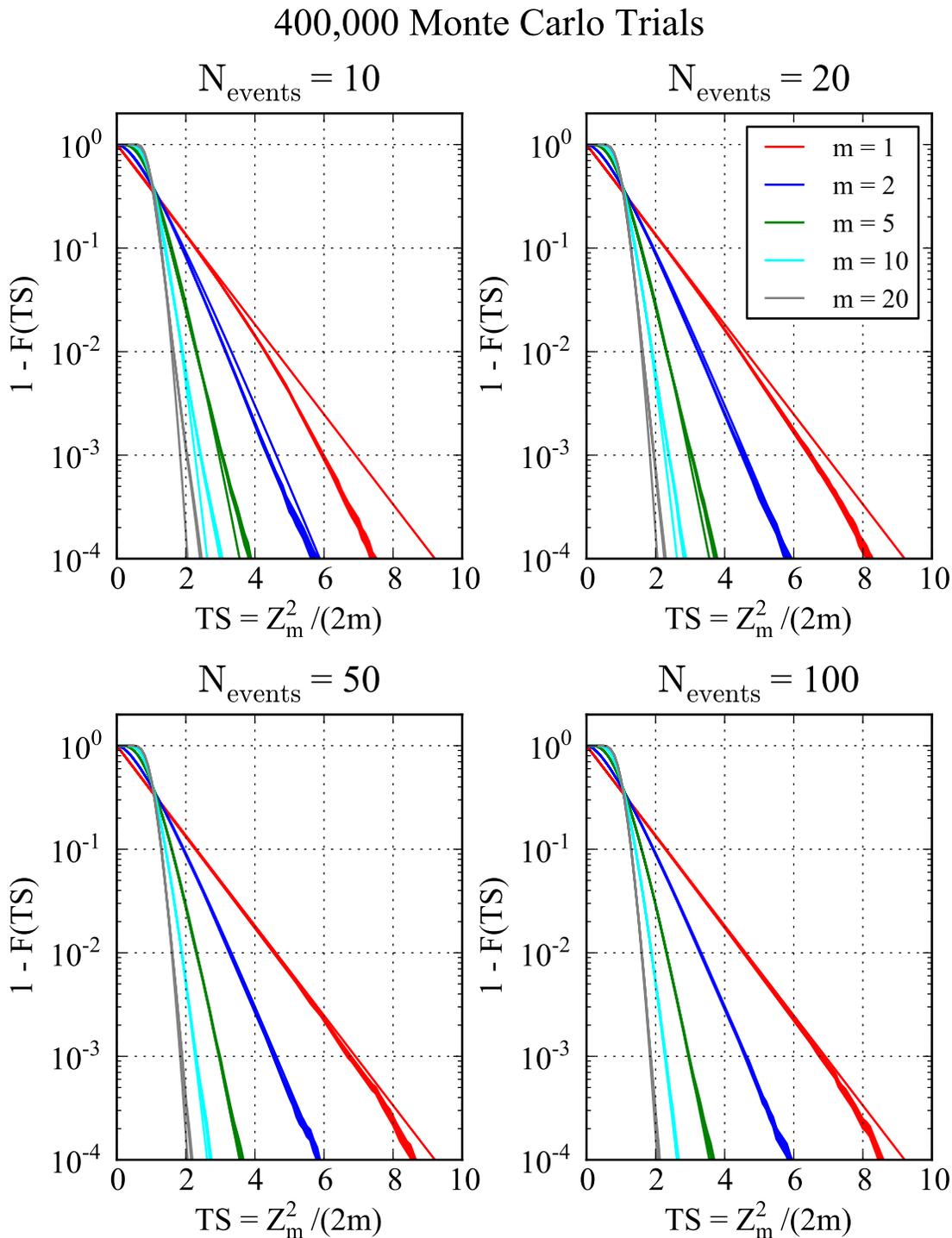
Figure 7.1: The observed distribution for the Z_m^2 statistic for various sample sizes and a collection of maximum harmonics. The asymptotic calibration is shown as a solid line, while the solid band gives the empirical distribution function of the Monte Carlo realizations. The width indicates the statistical uncertainty estimated as $\sqrt{N(\geq TS)}$, i.e., the square root of the number of counts with a TS greater than or equal to the current TS.

a different linear combination of g_k. The characteristic function for ψ_k becomes

$$\phi_{\Psi_k}(t) = \prod_{i=1}^{n} J_0(2w_i\sqrt{n}t) \approx \prod_{i=1}^{n} \left[1 - \frac{w_i^2\, t^2}{2\, n}\right] \approx \left[1 - \left(\sum_{i=1}^{n} w_i^2\right)\frac{t^2}{2}\right]. \qquad (7.13)$$

For this to match on to the characteristic function of a normal distribution, i.e., for the CLT to again apply, we must simply include $\sum_{i=1}^{n} w_i^2$ in a new definition of ψ_k, i.e.,

$$\lim_{n\to\infty} \sqrt{2\,n}\left(\sum_{i=1}^{n} w_i^2\right)^{\frac{-1}{2}} \psi_k \sim \mathcal{N}. \qquad (7.14)$$

Practically, the distribution function for the weights is unknown, so we would replace the distribution moments with the sample moments, i.e., we would follow the above prescription regardless of how we interpret the weights.

The weighted ψ_k are now asymptotically normally distributed. Their asymptotic mutual independence also remains, in the null case, as the distribution of weights and phases are independent. Thus, the sum of the weighted η_k will again be χ_{2k}^2. With this result, we define the weighted Z_m^2 test statistic, Z_{mw}^2,

$$Z_{2mw}^2 \equiv 2\left(\sum_{i=1}^{n} w_i^2\right)^{\frac{-1}{2}} \sum_{k=1}^{m} \alpha_{wk}^2 + \beta_{wk}^2. \qquad (7.15)$$

To verify the distribution, we generated a set of weights representative of those observed in actual data. Precisely, we drew a signal, s_i, from a χ_2^2 distribution and a background, b_i, from a χ_{50}^2, and calculated $w_i = s_i/(s_i + b_i)$. We incorporated the weights as in Eq. 7.15 and performed many Monte Carlo trials. The results, showing perfect equivalence between Z_m^2 and Z_{mw}^2, appear in Figure. 7.3.2.

7.4   The H Test

While the Z_m^2 test is unbinned in phase we still must choose the number of harmonics to include in the Fourier decomposition. Choosing m too small will result in a loss of power against sharply-peaked light curves, while m too large will lose power against broad, sinusoidal light curves. To ameliorate this problem, de Jager et al. ([42]) proposed the H test, which seeks to estimate the optimal m from the data. They defined

$$H_m = \max\left[Z_i^2 - c \times (i-1)\right], 1 \leq i \leq m \qquad (7.16)$$
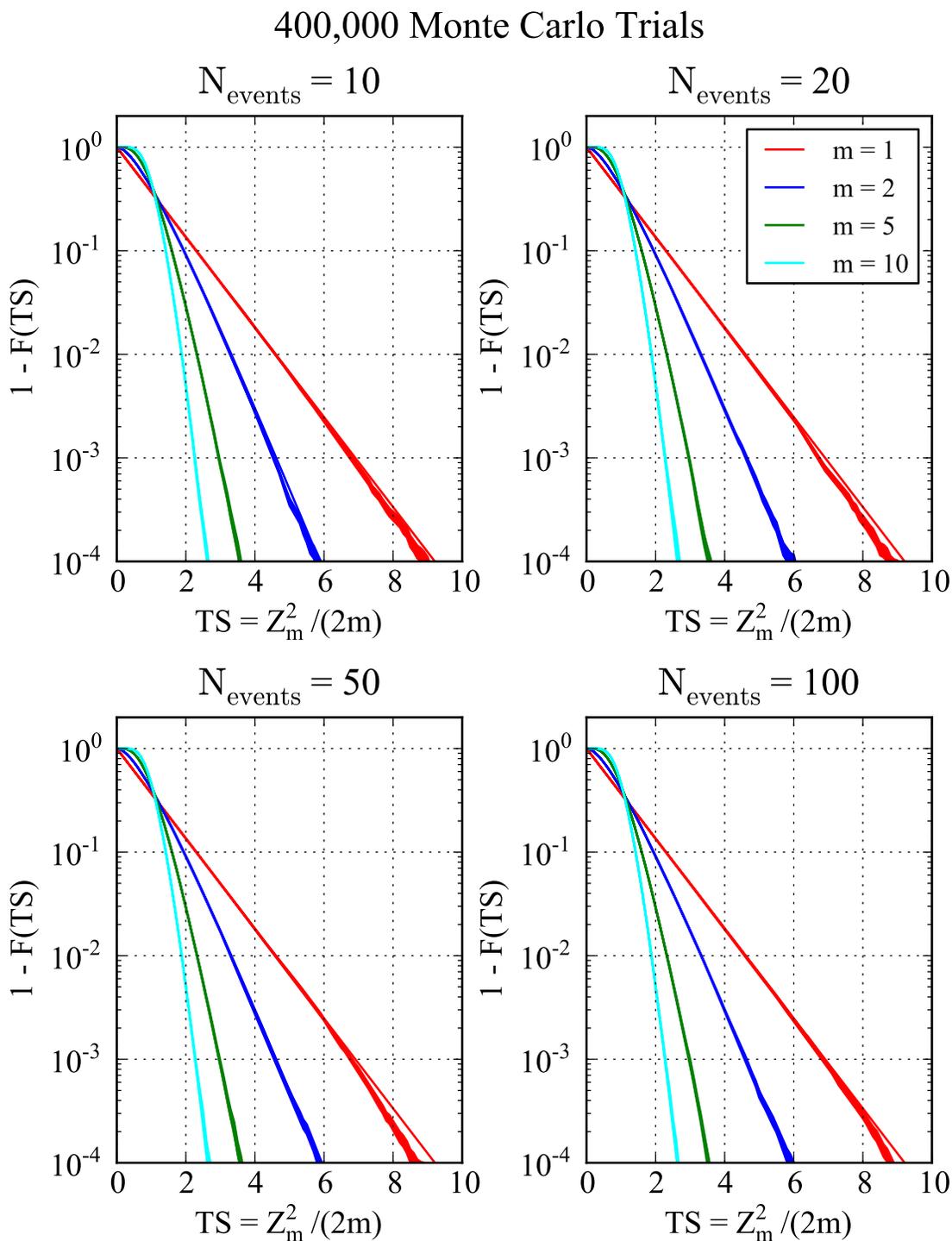
Figure 7.2: The observed (Monte Carlo) and predicted (asymptotic) distribution for the *weighted* test statistic Z^2_{mw}. The procedure is identical to that described for Figure 7.3.1 save for the addition of weights and the omission of the $m = 20$ case for decreased computational burden. The addition of weights is shown to make no difference in agreement with the predicted distribution; only insufficient sample size can leads to a discrepancy.

and specifically recommended $m = 20$ and $c = 4$ as an omnibus test. They provided a Monte Carlo calibration of the tail probability and, in a recent paper[41], increased the Monte Carlo statistics, giving an estimate $1 - F_{H_{20}}(h) \approx \exp(-0.4 \times h)$ good to $h \approx 70$. We derive the analytic, asymptotic calibration for all values of m, c, and h in Appendix A and use this calibration anywhere a conversion from h to chance of Type I error (i.e., "σ") is needed[3].

In connection with the preceeding material, we note that, since the calibration of the H-test relies only on the asymptotic χ^2_{2m} calibration and independence of the underlying η_k variables, any transformation leaving these properties unchanged (e.g. weighting) also leaves the Z^2_m and H test calibrations unchanged. Thus, we define a *weighted H-test statistic*

$$H_{mw} \equiv \max \left[Z^2_{iw} - c \times (i-1) \right], 1 \leq i \leq m. \tag{7.17}$$

In the sequel, we adopt the original values for c and m unless otherwise noted, i.e.,

$$H_w \equiv \max \left[Z^2_{iw} - 4 \times (i-1) \right], 1 \leq i \leq 20 \tag{7.18}$$

and likewise for H (unweighted).

7.5   Probability-weighted Statistics: Pulsars

With the definitions and derivations above, we are now ready to implement *probability-weighted* statistics for pulsation searches. By probability-weighted, we mean we seek to weight each phase (or rather, cosine or sine thereof) with some number that expresses a genuine probability that the photon associated with the phase originated from the (possibly pulsed) source. In constructing this quantity, we must first choose between an instantaneous or a time-averaged probability. Allowing for time dependence provides a more precise treatment of time-varying backgrounds, but it requires a complicated bookkeeping effort. However, for periodicity searches for periods on timescales comparable to or longer than the timescale for exposure variations, this bookkeeping is crucial. Time dependence also allows for *phase-dependence*, but we must assume a form for the light curve in order to calculate the weights, and this is more appropriate for a likelihood approach.

---

[3]Throughout this section, we use a two-tailed convention when converting to "σ" units, e.g., 2σ implies a chance probability of 5%.

Since we are concentrating on pulsars, we thus choose to compute the time-averaged probability. (See Appendix B for a treatment of searches for periods $>> 1$s.) To define a time-averaged probability, we refer to Eq. 4.10 and write down a slightly generalized version here, the expected rate from the jth source contributing to the ROI in an infinitesimal slice of observed energy and position

$$r_j(E, \vec{\Omega}) = \mathcal{F}(E, \vec{\lambda}) \, \epsilon(E, \vec{\Omega}_0) \, f_{\text{wpsf}}(\vec{\Omega}; \vec{\Omega}_0, E). \tag{7.19}$$

Here, we recall that $\epsilon(E, \vec{\Omega}_0)$ is the exposure integrated over the pointing history of the S/C, $f_{\text{wpsf}}(\vec{\Omega}; \vec{\Omega}_0, E)$ is the incidence-angle-averaged PSF, and $\mathcal{F}(E, \vec{\lambda})$ gives the intrinsic flux density of the source. In this expression, E and $\vec{\Omega}$ are the *measured* energy and position for a particular photon; this expression is unbinned in all quantities save time.

We now assume we have performed a likelihood analysis as outlined in Chapter 6 and have estimates for $\vec{\lambda}_j$, the parameters describing the spectral model for the jth source. Then, for a photon observed near E and $\vec{\Omega}$, the probability that it originated with the jth source is simply

$$w_j \equiv \frac{r_j(E, \vec{\Omega}, \vec{\lambda}_j)}{\displaystyle\sum_{i=1}^{N_s} r_i(E, \vec{\Omega}, \vec{\lambda}_i)}, \tag{7.20}$$

where N_s is the number of sources contributing to the ROI. w_j clearly satisfies the requirements of a probability: $w_j \in [0, 1]$ and $\sum_{i=1}^{N_s} w_i = 1$.

Having defined our approach, we can now make contact with previous efforts to improve periodicity searches. As we shall discuss in more detail in §7.5.1, applying any statistical test to phases only with an omnibus extraction scheme—e.g., all photons within a fixed radius, or within a PSF-dependent cone—is not at all optimal. At least two methods for improving on this basic scheme have been attemped. Brown et al. proposed constructing "photon probability" maps[31], essentially kernel density estimators for the sky constructed in a fashion analagous to that outlined in §6.3. By using these quantities in light curves constructed with COS-B data, they obtained an increase in the SNR ratio for the Crab pulsar. This method was profitably employed in searches for periodicity in EGRET data[72, 67].

This scheme—essentially probability weighting with a weight depending only on the

reconstructed photon position—is similar to our method but does not adequately account for varying SNRs. To see this, consider searching for pulsations from a very bright (very dim) source. At all energies of interest, the aperture over which the PSF is appreciably different from 0 will have a background component, particularly in the Galactic plane. If, in a given energy band, the source is much brighter (dimmer) than the background component, we want to include (exclude) events lying the tails of the PSF. This "choice" is naturally made by using probability weights derived from a full spectral model in which the relative strength of sources enters directly into the weight calculation.

A second approach was outlined by Özel and Mayer-Haßelwander[71] who proposed a method for constructing an optimal aperture of arbitrary shape in position-energy space based on the *expected* counts, i.e., a tentative spectral model for the candidate pulsar folded through the IRF. By adopting an approximate spectral model, the authors effectively estimated the SNR in each of the three energy bands they considered, overcoming partially the deficiency of PSF-weighting. However, their prescription was insensitive to light curve morphology, i.e., their algorithm returns the same aperture for a given position/spectrum, but the optimal aperture certainly depends on the light curve. As we shall show below, incorporating weights directly into the statistical test leads to a near-invariance in choice of aperture.

We now proceed with a comparison of weighted and un-weighted versions of the Z_m^2 and H test.

7.5.1 Performance: Type 1 Errors

The primary interest in constructing the weighted versions of pulsation search statistics (and the weights) is, of course, to craft tests that find more (real) pulsars. In statistical language, we want to minimize false positives (Type I Error) and false negatives (Type II error).

Type I error stems from two sources. First, there is the chance of a fluctuation in the test statistic sufficiently large to pass our pre-set threshold for rejection of the null hypothesis, i.e., claiming detection of a pulsar. As long as we understand the null distribution of the test

statistic, this particular source of error is easy to control: we simply determine in advance our overall tolerance to false positives and set the TS threshold accordingly. We must be cautious about applying the aysmptotic calibration of the null distribution to small sample sizes; in such cases, it is always a good idea to verify the chance probability with a Monte Carlo simulation.

A more insidious source is related to the nature of LAT data. As discussed in Chapter 4, the PSF is a strong function of energy, and at energies of a few GeV and below, there is strong source confusion, particularly in the Galactic plane. Simply going to higher energies for the better SNR is, however, not feasible for pulsar as the spectra are exponentially supressed above a few GeV. Thus, the sensitivity to pulsations can depend quite strongly on the particular selection criteria for the data, e.g., the energy-dependent extraction radius. Without knowing these cuts in advance, one is tempted to tune the cuts using the test statistic itself, a famous source of bias. To compensate, one should correct with a trials factor equal to the number of cuts tested. Alternatively, the use of omnibus cuts will result in a sub-optimal SNR, increasing the chance of a false negative. Finally, there is the issue of diminished sample size—and concomitant departure from the asymptotic calibration—with very stringent cuts.

We expect that *probability weighted statistics* will be all but immune to the issue of aperture selection. High energy photons falling in the core of the PSF will naturally receive a weight close to 1, i.e., we include all photons that would be accepted by a stringent cut on SNR[4]. However, we also include a large set of photons from lower energies that receive modest weights but undoubtedly still contain pulsed signal. Finally, photons falling in the tails of the PSF at any energy receive a very low—essentially zero—weight. Unweighted, these low-energy, distant photons would swamp the signal. Weighted, they are neglibible. We can then with impunity select a single, large region of interest and be guaranteed good performance. Practically, no signal is gained beyond 2 to 3° save for the very strongest pulsars, which are of course already known; see Figure 7.5.1. Since we never tune the data

---

[4]Since the weighted statistics are in some sense normalized, the actual proximity to the maximum value of 1 is irrelevant. A more precise statement: "photons falling in the core of the PSF at high energies will have significantly higher than average weights".

selection, we incur no undue risk of Type 1 error. In this regard, probability weighting significantly increases the performance of the Z_m^2 and H statistics.

To make these claims concrete, we compare the weighted H test (H_{20w}) and the standard H test (H_{20}) over a grid of selections in photon position and energy. The method and results are presented in Figure 7.5.1. Similar results in a single dimension—photons with energies above 200 MeV and a grid of position selection criteria—in terms of detection threshold flux (defined in the section below) appear in Figure 7.5.1. In both examples, the weighted statistics are largely insensitive to the extraction criterion. The result can be summarized as follows: unweighted statistics depend on SNR while weighted statistics depend primarily on *signal alone*, and the strategy is then to use a single, large aperture for all weighted pulsation tests.

7.5.2 Performance: Type II errors (Sensitivity)

We must also consider Type II error—false negatives. In astrophysical terms, this is directly related to the sensitivity of the method, i.e., the number of photons that must be collected before the test statistic passes (for some percent of an ensemble) the detection criterion.

The easiest way to quantify the sensitivity is by simulation of an ensemble of pulsed sources. As for the validation of spectral analysis, we use the Science Tool *gtobssim* to simulate a point source with a diffuse background. For the discussion below, we use a realistic pulsar spectrum, in particular

$$\frac{dN}{dE} \propto (E/GeV)^{-\Gamma} \exp{-E/E_c} \tag{7.21}$$

with $\Gamma = 1.57$ and $E_c = 3150 \text{GeV}$. (This spectrum is typical of many pulsars and is, in fact, close to the measured spectrum for the middle-aged Vela pulsar.) We place the source at the position of the Vela pulsar, (R.A., Decl.) = (128.8463, -45.1735). We choose this scenario since (a) many pulsars lie in the Galactic plane, in particular along the spiral arms as projected onto the sky and (b) detection in the Galactic plane is much more difficult than at high Galactic latitudes, so improved methods are particularly helpful here. As in Chapter 4, we simulate a very bright source ($\mathcal{F} = \mathcal{F}_5 = 10^{-5} \text{ph cm}^{-2} \text{ s}^{-1}$) and select subsets of the photons in order to achieve a target brightness. Unlike in the spectral analysis case, we use
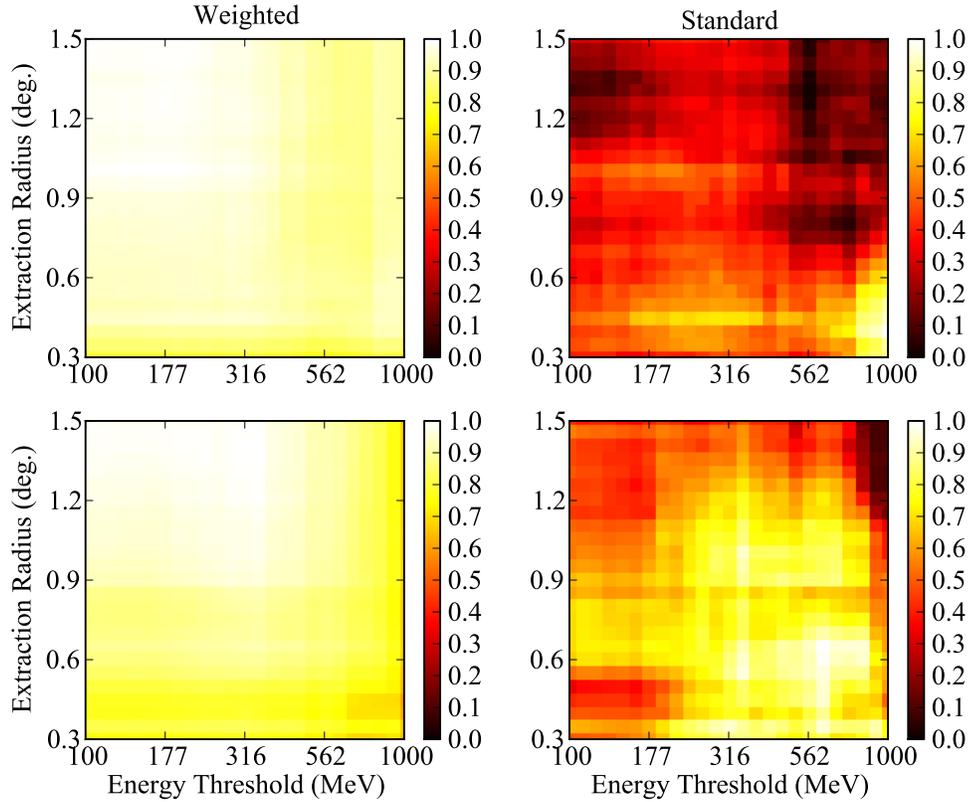
Figure 7.3: A comparison of the dependence on event selection of H_{20w} and H_{20}. A source with flux 10^{-8} ph cm^{-2} s^{-1} was simulated and a "Vela-like" light curve was added to the source photons. The spectrum was that of Eq. 7.21. The test statistics were calculated over a grid of selection criteria: the y-axis gives the extraction radius (maximum angular distance from the pulsar allowed) and the x-axis indicates the threshold energy, the minimum energy allowed in the selection. The lefthand (righthand) columns shows the results for the weighted (un-weighted) test statistic. The two rows are two indepedent Monte Carlo realizations of the source and background. The test statistics were converted to σ units and independently normalized to the maximum observed value. It is clear that the weighted statistics have a minimal dependence on the extraction criterion and, more importantly, do best when given the most data (maximum in upper lefthand corner of cut space). On the other hand, the significance can be strongly peaked for the standard statistic and the optimal cut varies from realization to realization.
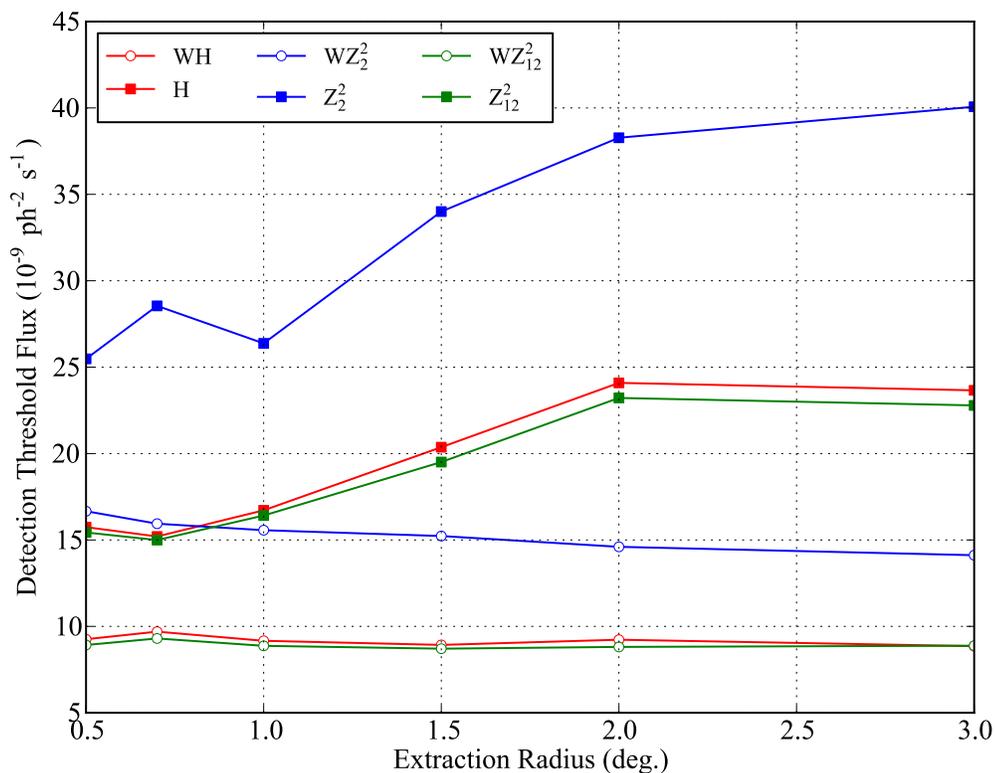
Figure 7.4: A comparison of the weighted and un-weighted versions of three statistics. Photons with energies above 200 MeV are selected according to the extraction radius on the x-axis. The light curve in this case is a Gaussian with $\sigma = 0.03$, so the H test and high-harmonic Z test are much more sensitive than the Z_2 tests. The salient feature is the relative independence of the detection flux threshold on extraction radius for the weighted methods. The sensitivity for the standard tests drops by about 50% as we increase the aperture size. The results indicate that little is gained for weighted methods with an extraction radius larger than $2°$.

a single realization of the diffuse background. This is not a serious impediment since we can randomize the phase of the diffuse photons for each MC realization. Using the same spectral models used to simulate the sources, we calculate the weights using *gtsrcprob*, a Science Tool designed to implement the probability-weighting scheme. The base weights must be modified to account for changing point source flux; we address this point below.

When the Monte Carlo events are generated by *gtobssim*, the simulated event times have no inherent pulsation. In order to convert "unpulsed" photons to a "pulsed" data set, we apply the following procedure:

1. Select a template for the light curve of the pulsar. This will typically be a sum of (wrapped) Gaussian functions, and will be normalized, i.e., a probability density function for photon phase.

2. Select photons originating with the pulsed source. This is possible because each event generated by *gtobssim* is tagged with a unique identification for the generating source.

3. For each selected photon, draw (by Monte Carlo) a phase from the light curve and assign it to the photon.

For background photons, we simply assign a phase drawn from the uniform distribution. In this prescription, there is no sense of a time-to-phase mapping. If we wish to use an explicit ephemeris, e.g., specifying a reference time, a period, and a period derivative, then this timing solution predicts the phase as a function of time. We then must perform the following additional steps to make the simulated times consistent with the simulated phases:

1. For the simulated time, use the time-to-phase mapping to calculate the phase *predicted* by the timing solution.

2. Use the timing solution to determine the pulsar period at the simulation time.

3. Adjust the simulated time by $\delta t = P(t) \times (\phi_{sim} - \phi_{mod}) \bmod 1$, with ϕ_{sim} the phase drawn from the light curve and ϕ_{mod} the phase predicted by the timing solution, i.e.,

the smallest possible adjustment that will align the simulated time with the desired light curve. Provided $P(t)$ is small enough, the change in exposure is negligible.

After applying this procedure, one can use the timing solution to convert the adjusted simulated times to phase, and these phases will follow the light curve distribution. These steps are not necessary for testing pulsed sensitivity and we include them here only for completeness.

With these procedures, we have in hand an ensemble of pulsars at a series of intrinsic fluxes with arbitrary light curves. To test the sensitivity, we determine the *flux threshold* for each method given a particular LC morphology, i.e., the flux for which a source would be detected according to some criteria (see below) a majority of the time. More specifically, we

1. At a sequence of intrinsic fluxes, evaluate the test statistic for each method for each member of the ensemble. At each flux, we must recalculate the probability weights. These were originally assigned with the *simulation* flux values, $\mathcal{F}_{sim}$, and we now wish to described a source with target flux $\mathcal{F}_{tar}$. The new weights are given by $w_{tar}^{-1} - 1 = (w_{sim}^{-1} - 1) \times \mathcal{F}_{sim}/\mathcal{F}_{tar}$.

2. Using the asymptotic calibration, calculate the chance probability of Type I error (the tail probability in the asymptotic distribution) and convert it to σ units, i.e., a value of $X\sigma$ is the two-sided tail probability of a normal distribution integrated from X to ∞.

3. Determine the flux for which 68% of the ensemble deliver a σ value above some predetermined threshold (see below). This is the *detection* or flux threshold. To determine the 68% level with some level of robustness, we fit the ensemble values with a normal distribution and report the appropriate quantile value.

To determine the flux threshold, we need to invert the calculated quantity (tail probability in σ units) to the desired quantity (flux threshold). Fortunately, the relationship between the two is linear. This dependence may be slightly surprising. Significance canonically scales
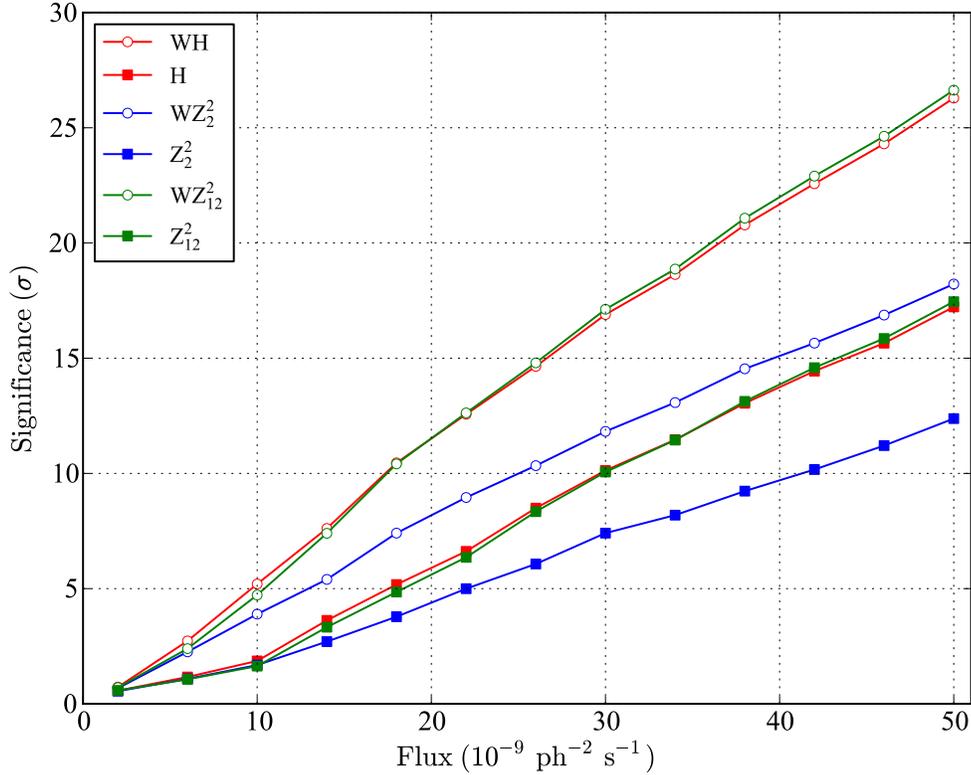
Figure 7.5: The dependence of significance in σ units as a function of flux. The reported value at a given flux is that attained by at least 68% of the ensemble of 50 MC realizations. The trend is approximately linear, as explained in the main text, making it simple to invert the relation. The light curve here is a single Gaussian peak with $\sigma = 0.03$.

as $\sqrt{N}$, the collected events, or in this case, since we integrate for exactly one year, $\sqrt{\mathcal{F}}$, the square root of the source flux. However, when we increase the source flux without increasing the background flux, we also increase the SNR, and in general significance is proportional to the square root of SNR. Taken together, the dependence is linear, i.e., $\sigma \propto \mathcal{F}$. This dependence for a particular source is shown in Figure 7.5.2.

Above, we mentioned a pre-determined threshold for the tail probability in σ units: test statistics above this threshold cause us to discard the null hypothesis, i.e., claim detection of a pulsed source. In the following, we choose 4σ for as the confidence level for detection.

It provides a very small probability of Type I error, about 1 in 15800, and it is also not too different from the LAT Collaboration's internal threshold of 5σ. On the other hand, as we have seen, sample sizes of order 100 (typical for a faint pulsar) begin to depart from their asymptotic calibrations at around this confidence level.

We are now ready to characterize the effect of adding probability-based weights to existing statistics on sensitivity. To give a modest survey of the statistics outlined above, we present weighted and un-weighted versions of H_{20}, Z_{12}^2, and Z_2^2. Recall that H_{20} is an "omnibus" test that depends little on light curve morphology, whereas we expect Z_{12}^2 to perform well for light curves with sharp peaks and Z_2^2 to perform well for light curves with broad features.

The primary result, in Figure 7.5.2, shows the flux threshold for each method as a function of "duty cycle", the fraction of the full phase for which there is appreciable pulsed emission. In this case, the light curves are single Gaussian peaks with a variety of values for their σ parameter; the templates are shown in Figure 7.5.2. For a fixed flux, increasing the duty cycle decreases the peak flux, or SNR, and the primary dependence is then an inverse relation between flux threshold and duty cycle. However, there is additional dependence from the nature of each test. It is clear, e.g., that the H_{20} and Z_{12}^2 perform significantly better than Z_2^2 for low duty cycle sources, while Z_2^2 maintains a slight edge for broad light curves. The H test does a good job for all duty cycles.

More importantly, it is clear that, independent of pulsar duty cycle (unsurprisingly), the weighted statistics enjoy about a factor of two smaller threshold for detection than their unweighted counterparts. This has nontrivial implications for the detected pulsar population. E.g., if the pulsars follow a linear logN-logS relation with slope α, using weighted statistics yields $\approx 2^\alpha$ additional detections.

Two-peaked light curves

Many pulsar light curves illustrate twin peaks, often separated by about 0.4-0.5 cycles. We repeat the analysis of the previous section using a template comprising two Gaussian peaks separated by 0.45 in phase. Real pulsar peaks often have unequal intensities (with an

Figure 7.6: The 68% flux detection threshold based on a 4σ detection criterion for a single-peaked Gaussian light curve as a function of duty cycle. Light curves with narrow (broad) peaks lie to the left (right). The un-weighted thresholds are about twice those of the weighted thresholds, indepedent of duty cycle or statistic.
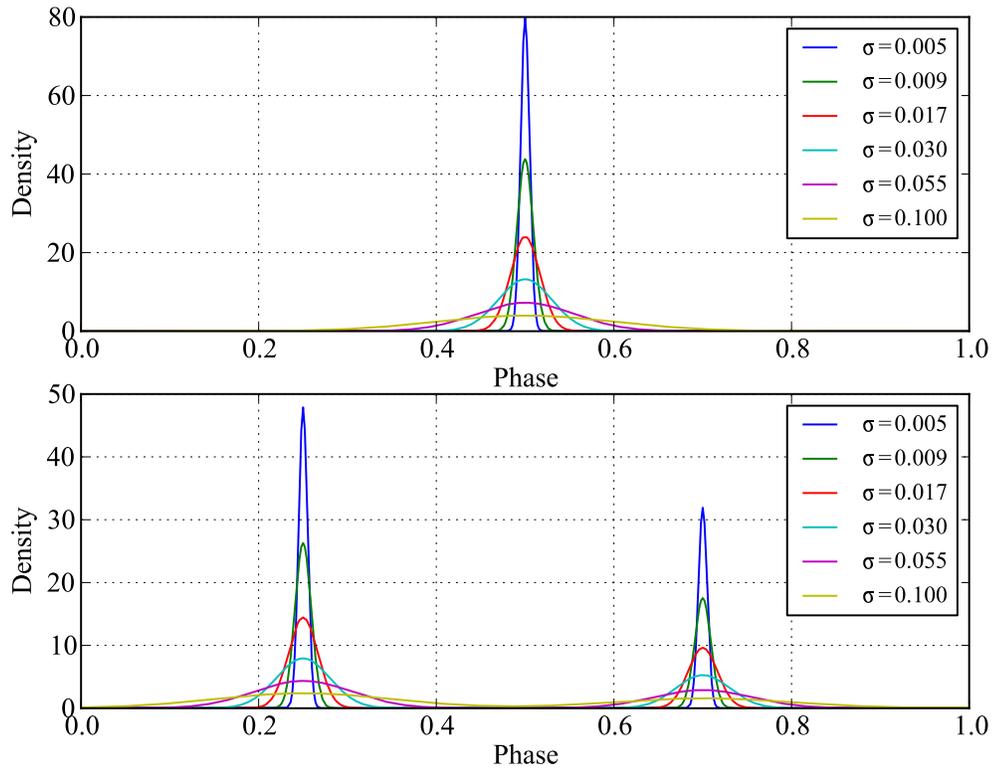
Figure 7.7: The templates used for the single- and double-peaked light curves in the determination of the detection flux thresholds. The functional form is a wrapped Gaussian, i.e., $f(\phi) = \sum\limits_{i=\infty}^{\infty} g(\phi + i)$ with $g(\phi, \mu, \sigma) = (2\pi)^{-0.5} \exp(-0.5\,(\phi - \mu)^2/\sigma^2)$. In the first panel, $\mu = 0.5$, while in the second panel, $\mu_1 = 0.25$, and $\mu_2 = 0.70$. In this panel, the ratio of the peak heights is $3/2$, and the peak widths are identical.

Figure 7.8: The 68% flux detection threshold based on a 4σ detection criterion for a double-peaked Gaussian light curve as a function of duty cycle. The flux threshold for the weighted statistic is $1.5 - 2.0$ times lower than the unweighted statistics.

energy-dependent ratio.) We reflect that here with a slightly-dominant leading peak. The templates are shown in Figure 7.5.2. As seen in Figure 7.5.2, the overall flux thresholds are unsurprisingly increased: at a fixed flux, spreading photons between multiple peaks decreases the overall SNR. The weighted statistics maintain a comfortably decreased flux threshold.

Pulsars with DC Emission Components

Some pulsars (not respecting their nomenclature[5]) emit an appreciable portion of their flux as unpulsed γ rays, e.g. PSR J1836+5925[13]. (Models of magnetospheric emission generally allow for a sustained GeV component under certain geometries and viewing angles.) Unpulsed emission could be a confounding factor for the weighted test, since unpulsed photons from the source will carry heavy weight but be uniformly distributed, decreasing the effective SNR. And indeed, in Figure 7.5.2, where we have used a single-peaked light curve with a pulsed fraction of 1/2, we see (a) an overall increased flux threshold on the order of two, as expected (b) a slight increase in the ratio of weighted-to-unweighted detection thresholds. However, the ratio is 1.5 to 2.0, indicating the weighted statistics still offer significantly improved sensitivity for pulsars with appreciable DC emission.

Effect of Uncertainties in Spectral Parameters

From these demonstrations, it is clear that the probability-weighted statistics have, on average, a factor of about 2 improved sensitivity relative to the unweighted versions. One potential objection is that, in determining the weights for this validation, we have used the known spectrum. With real data, of course, we must first estimate the spectrum in order to determine the weights. To assess the impact of using estimated parameters with concomitant uncertainty, we make use of the Monte Carlo ensembles developed for the validation of spectral analysis and described in Chapter 5.

For this application, we simulated 20 realizations of a point source at the position of the Vela pulsar with the same power law with exponential cutoff spectrum as employed above, Eq. 7.21. To this data we added phase from a single-Gaussian light curve with $\sigma = 0.03$. From Figure 7.5.2, we see that the detection threshold for such a configuration is estimated at $\approx 8 \times 10^{-9} \text{ph cm}^{-2} \text{ s}^{-1}$. Thus, following the procedure outlined in Chapter 5 for constructing ensembles of point sources with a particular flux, we generate 20 realizations of this "pulsar" with a flux of $8 \times 10^{-9} \text{ph cm}^{-2} \text{ s}^{-1}$ [6]. First, we calculate the probability

---

[5]Given the original notion of pulsar as a "pulsating source of *radio*", with the advent of "radio-quiet pulsars" this ship may already have sailed.

[6]At this flux, there are of order 100 source photons for the 1-year integration period, allowing for asymp-
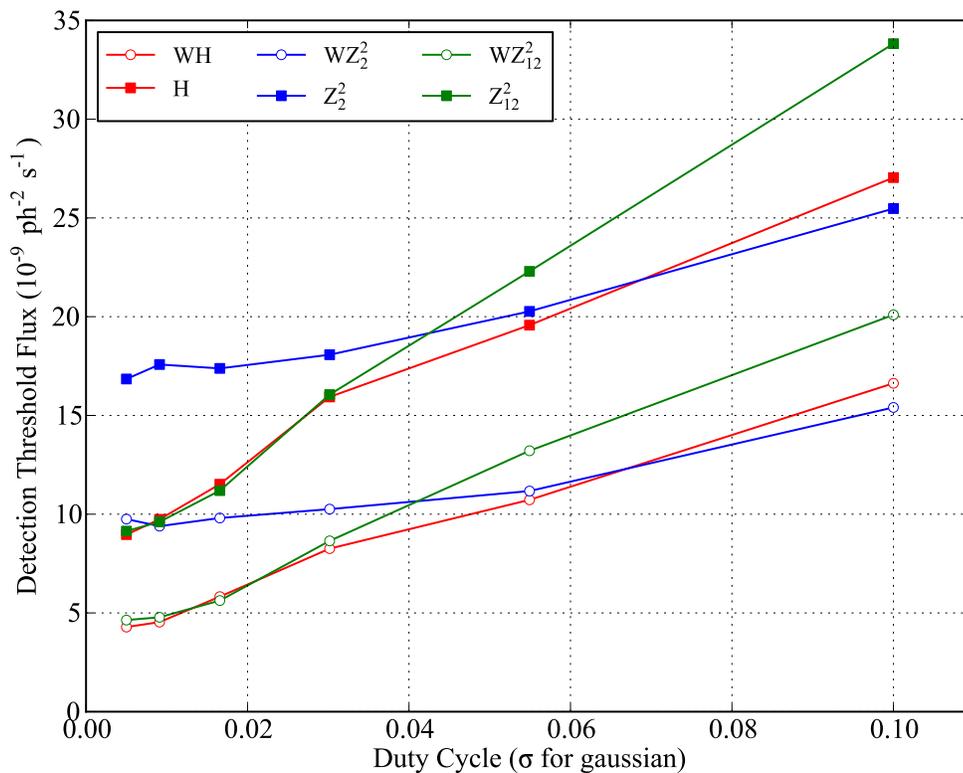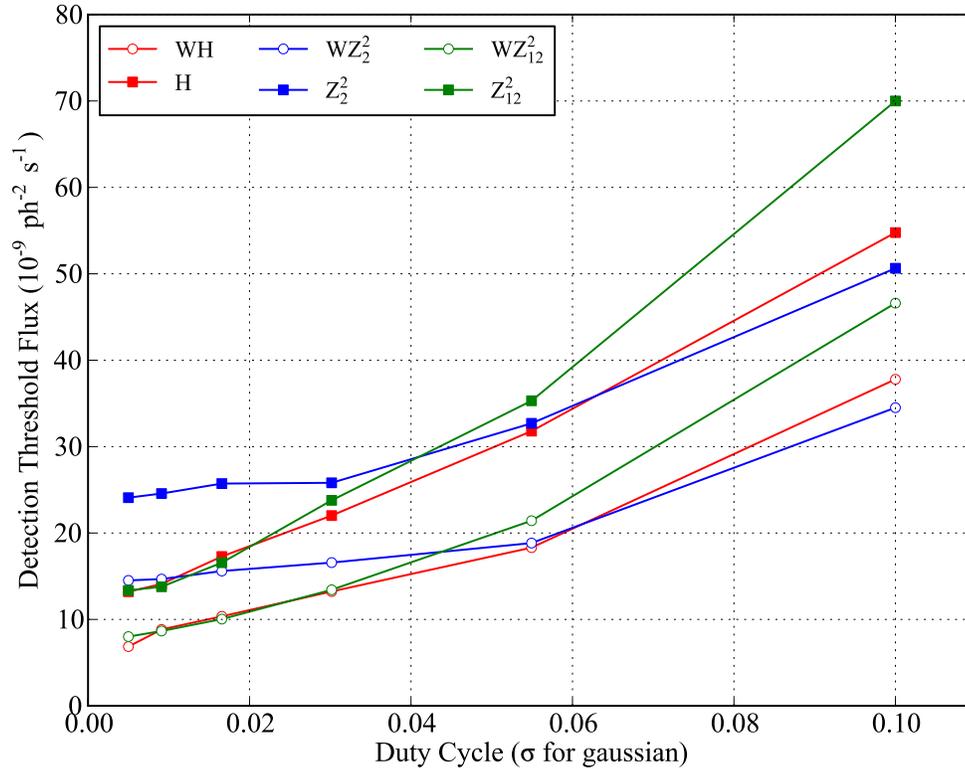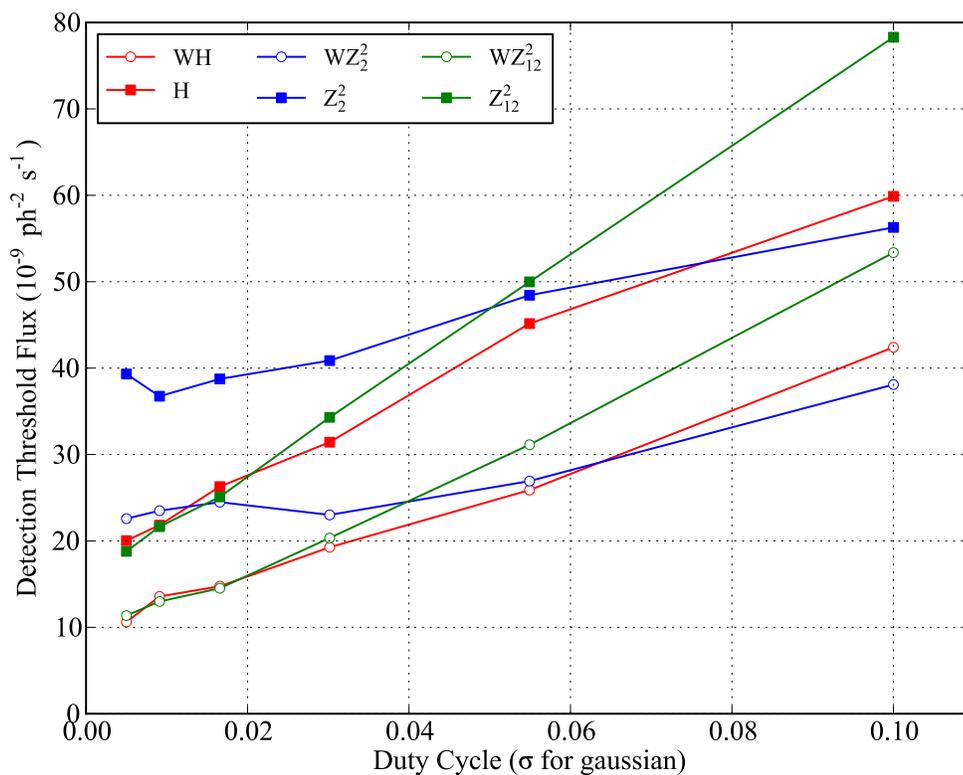
Figure 7.9: The 68% flux detection threshold based on a 4σ detection criterion for a single-peaked Gaussian light curve as a function of duty cycle. Here, the pulsed fraction has been decreased to 50%, i.e., half of the photons from the source are uniformly distributed and half are drawn from the Gaussian light curve. The flux threshold for the weighted statistic is $1.5 - 2.0$ times lower than the unweighted statistics, with some slight dependence on duty cycle and statistic.

weights using the known model parameters for the pulsar and the diffuse background, i.e., the "ideal" case. We then perform a ML spectral fit with *pointlike* to estimate the spectral parameters. Since the simulated point source is very dim relative to the background, it is inappropriate to attempt to fit a spectral model with three degrees of freedom. We therefore fix the cutoff energy to 100 GeV, essentially a power law spectrum. This approach is *conservative* since we are using an incorrect spectral model. Using the best-fit values for the flux density and the photon index, we calculate a new set of probability weights. Finally, we use a weighted H test to determine the significance (a) using the "ideal" weights and (b) using the "measured" weights.

We compare the results—in σ units—in Figure 7.5.2. In general, the detection significance obtained with the "measured" weights is slightly lower than that obtained with the "ideal" weights, although in a few cases statistical fluctuations lead to the opposite outcome. Comparing the populations means, the the overall significance using measured weights is decreased to 92% of the ideal case. Recalling that the flux threshold depends linearly on the significance, we estimate the flux threshold is increased in the case of measured weights by $\approx 10\%$. This effect is relatively small compared to the factor of $1.5 - 2.0$ increase in flux threshold seen between the weighted and unweighted versions of the H test. We therefore conclude that—even accounting for uncertainties in the spectral parameters used to calculate the probability weights—weighted statistics offer a significance improvement in sensitivity.

Comparison of Pulsed and Unpulsed Detection Thresholds

The machinery established above—performing spectral fits on an ensemble to compute the weighted statistics using probabilities estimated from an ML fit—also provides for directly comparing the DC (unpulsed) source significance with the pulsed significance. Recall from the discussion in Chapter 6 that there is an asymptotic calibration for the test statistic obtained from the likelihood ratio for a single parameter whose null value lies on a boundary. In that case and here the parameter was the flux of a point source which is zero in the null

totic calibration.
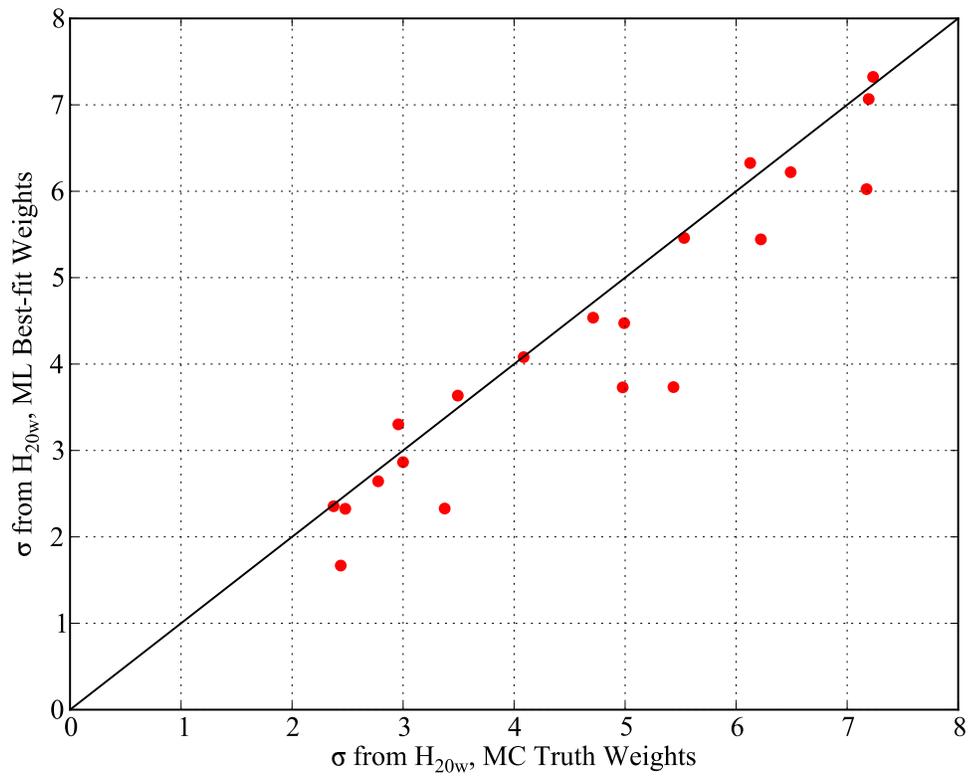
Figure 7.10: The significance for each member of the ensemble described in the text calculated using weights derived from the model used to simulated the data (the Monte Carlo "truth") and derived from the ML model. The ML-derived values are very similar to those obtained using the known spectrum.

case. To apply this calibration, we proceed as in Chapter 6 and fix all parameters but the flux. We set the cutoff energy to 100 GeV as above, and we set the photon index to $\Gamma = 2.0$.

Next, we perform a ML fit on the ensemble of sources described in the previous section to determine the best-fit value for the flux density. The DC significance is then determined by

$$\sigma_{DC} = \sqrt{2 \times \log \mathcal{L}_{opt}/\mathcal{L}_0}, \tag{7.22}$$

with $\mathcal{L}_{opt}$ the likelihood value obtained with the best-fit flux density and $\mathcal{L}_0$ the likelihood value obtained with the flux density set to zero.

As in the previous section, we calculate the probability weights using both the Monte Carlo truth values of the parameter and with the best-fit spectrum—in this case, with cutoff energy *and* photon index fixed—to estimate a pulsed significance with the H_{20w} test.

We compare the unpulsed and pulsed significances in Figure 7.5.2. The vast majority of ensemble members are detected more significantly through pulsations than through unpulsed emission. The measured significance for pulsed detections is a factor of 2.2 greater than for DC detection. If we assume that the detection flux threshold is linear in σ_{DC} as it is in the pulsed significance, this means we require sources to be about twice as bright on average to detection them through DC emission rather than pulsed emission[7]. It is amusing that this is about the same factor we observe when comparing weighted and unweighted statistics.

To determine how fitting the spectra with only one degree of freedom affects the performance of the weighted statistics, we compare in Figure 7.5.2 the significance computed from the "ideal" weights and from the weights determined from the best-fit spectrum (recalling that the photon index and cutoff energy are fixed). Interestingly, the "measured" weights deliver results comparable to those calculated in §7.5.2, indicating an insensitivity to the overall spectral shape.

7.6 Calculating Pulsed Sensitivity

One of the primary observables in constraining pulsar emission mechanisms is the overlap between radio-loud and γ-ray bright pulsars. An important extension of this is placing the

---

[7]This result, of course, depends strongly on the light curve morphology.
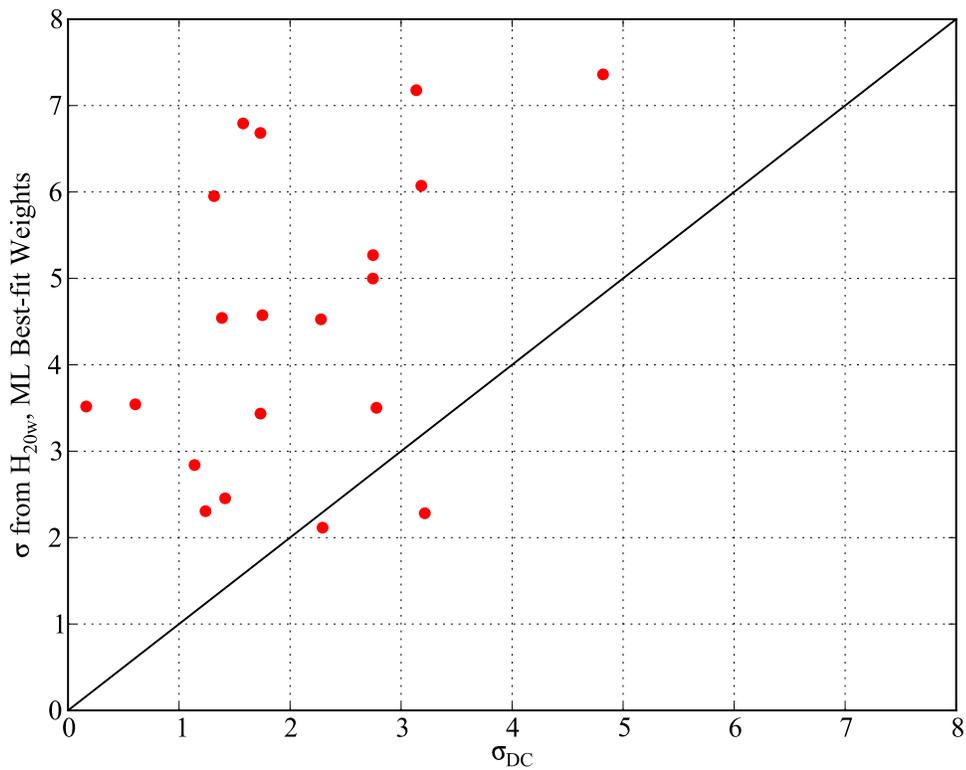
Figure 7.11: A comparison of the significance of pulsed detection versus unpulsed detection. The pulsed detection significance was calculated using the ML best-fit spectrum with the weighted H test, while the unpulsed significance was derived from a likelihood ratio test as described in the text. For nearly all sources, the pulsations are more strongly detected than the DC emission.
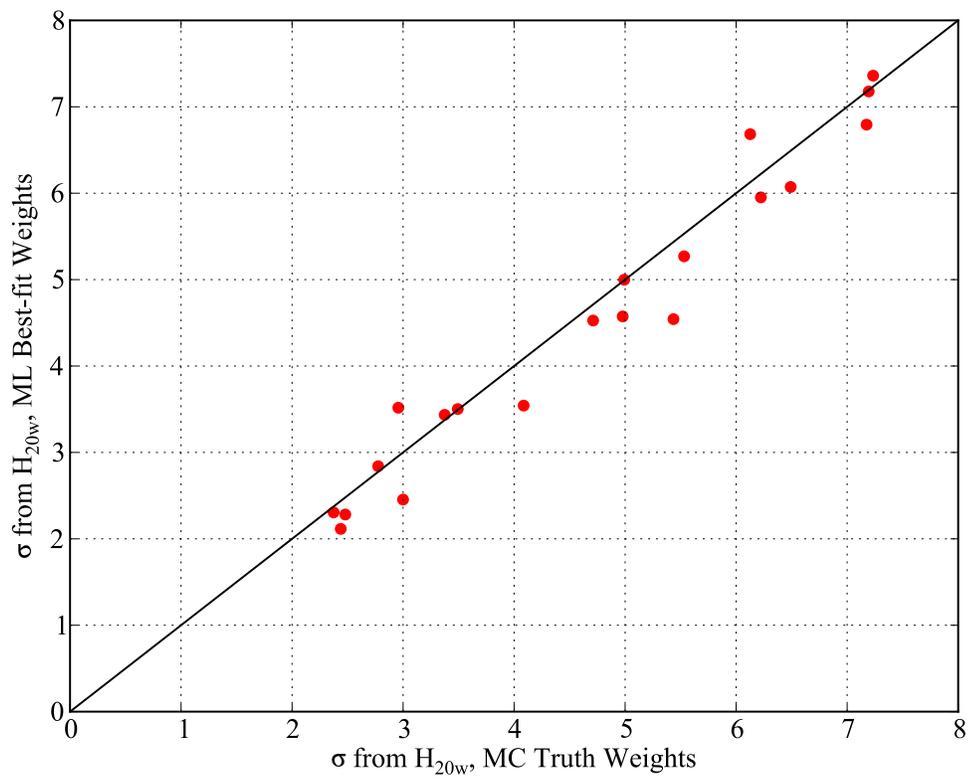
Figure 7.12: As Figure 7.5.2 but with weights derived from a ML fit with the pulsar spectrum fixed to a power law with $\Gamma = 2.0$ and only the flux allowed to vary.

strongest constraints possible on the γ-ray flux from radio-loud pulsars. In the material above, we saw that weighted statistics outperform both unweighted versions of the same statistics *and* unpulsed significance tests. Thus, pulsed upper limits are a natural candidate for constraining the γ-ray emission.

Although much work has been done on extracting analytic pulsed upper limits (see [40] for an enlightening discussion), the weighted method is not particularly amenable to this approach. Each source comes with its own set of probability weights that depends strongly on the pulsar spectrum and its position on the sky. Thus, while we can extract analytic upper limits once we know the weight distribution, by calculating the weights we have already done most of the work! The machinery we developed above for determining flux detection thresholds is precisely what is needed for an upper limit/sensitivity.

To determine the sensitivity of *Fermi*-LAT to pulsed detections, we essentially repeat the procedure outlined in the previous section, viz.

1. Simulate an ensemble of point sources with some spectral shape.

2. Add phase from, e.g., a single-peaked or a double-peaked light curve.

3. Select a significance level for detection, e.g., the 4σ criterion adopted in our analyses.

4. Determine the fraction of the ensemble that must exceed the detection significance to claim a detectable flux. (We used 68%.)

5. Calculate the flux detection threshold as outlined above.

This flux detection threshold is the desired sensitivity, i.e., the flux level at which we can typically (68% of the time) detect the source with significance exceeding our threshold (4σ.) Sensitivity maps giving the pulsed flux detection threshold as a function of position on the sky can be generated by repeating this exercise over a grid of positions.

7.7 Summary

We motivated the use of unbinned, Fourier-based statistics for the detection of pulsations and we outlined some of their properties. We showed that they could be modified to incorporate probability weights while retaining their calibration. Appealing to gains made in using likelihood—i.e., knowledge of the IRF—we proposed that likelihood-derived probability weights would offer improvement over statistics formulated with counts alone. We defined the time-averaged probability that a given photon comes from a particular source in terms of the ML spectral model for the source and its background. We found that the weighted versions of the Z_m^2 and H test had very little dependence on the particular data selection used in contradistinction to the unweighted versions (Figure 7.5.1). By evaluating test statistics for ensembles of simulated pulsars, we estimated the detection flux threshold for a given statistical test and showed that the flux threshold for weighted tests was a factor of $1.5 - 2$ lower than the unweighted test (Figures 7.5.2 and 7.5.2). Finally, we demonstrated that these improvements depended very little on the fraction of pulsed emission (Figure 7.5.2) or on exact knowledge of the spectrum (Figures 7.5.2 and 7.5.2). Additionally, we found that pulsed sensitivity is about twice that of the unpulsed sensitivity (Figure 7.5.2) for a fully-pulsed source.

Taken together, these results represent a significant gain in the search for periodic signals from pulsars and suggest the adoption of weighted statistics—in particular, H_{20w}—as an omnibus test for pulsation.

BIBLIOGRAPHY

[1] A. A. Abdo et al. The Fermi Gamma-Ray Space Telescope Discovers the Pulsar in the Young Galactic Supernova Remnant CTA 1. *Science*, 322:1218–, November 2008.

[2] A. A. Abdo et al. A Population of Gamma-Ray Millisecond Pulsars Seen with the Fermi Large Area Telescope. *Science*, 325:848 (MSP Population), August 2009.

[3] A. A. Abdo et al. Detection of 16 Gamma-Ray Pulsars Through Blind Frequency Searches Using the Fermi LAT. *Science*, 325:840–, August 2009.

[4] A. A. Abdo et al. Fermi large area telescope observations of the cosmic-ray induced γ-ray emission of the Earth's atmosphere. *Phys. Rev. D*, 80(12):122004–+, December 2009.

[5] A. A. Abdo et al. Fermi Large Area Telescope Observations of the Vela Pulsar. *ApJ*, 696:1084 (Vela), May 2009.

[6] A. A. Abdo et al. Fermi LAT Observation of Diffuse Gamma Rays Produced Through Interactions Between Local Interstellar Matter and High-energy Cosmic Rays. *ApJ*, 703:1249–1256, October 2009.

[7] A. A. Abdo et al. Fermi LAT Observations of LS I +61deg303: First Detection of an Orbital Modulation in GeV Gamma Rays. *ApJL*, 701:L123 (LS I +61° 303), August 2009.

[8] A. A. Abdo et al. Pulsed Gamma-rays from PSR J2021+3651 with the Fermi Large Area Telescope. *ApJ*, 700:1059 (PSR J2021+3651), August 2009.

[9] A. A. Abdo et al. Pulsed Gamma Rays from the Millisecond Pulsar J0030+0451 with the Fermi Large Area Telescope. *ApJ*, 699:1171 (PSR J0030+0451), July 2009.

[10] A. A. Abdo et al. The on-orbit calibration of the Fermi Large Area Telescope. *Astroparticle Physics*, 32:193–219, October 2009.

[11] A. A. Abdo et al. Detection of Gamma-Ray Emission from the Starburst Galaxies M82 and NGC 253 with the Large Area Telescope on Fermi. *ApJL*, 709:L152–L157, February 2010.

[12] A. A. Abdo et al. Fermi Large Area Telescope First Source Catalog. *ApJS*, 188:405–436, June 2010.

[13] A. A. Abdo et al. Fermi Large Area Telescope Observations of PSR J1836+5925. *ApJ*, 712:1209–1218, April 2010.

[14] A. A. Abdo et al. Fermi Observations of Cassiopeia and Cepheus: Diffuse Gamma-ray Emission in the Outer Galaxy. *ApJ*, 710:133–149, February 2010.

[15] A. A. Abdo et al. Spectrum of the Isotropic Diffuse Gamma-Ray Emission Derived from First-Year Fermi Large Area Telescope Data. *Physical Review Letters*, 104(10):101101–+, March 2010.

[16] A. A. Abdo et al. The First Catalog of Active Galactic Nuclei Detected by the Fermi Large Area Telescope. *ApJ*, 715:429–457, May 2010.

[17] A. A. Abdo et al. The First Fermi Large Area Telescope Catalog of Gamma-ray Pulsars. *ApJS*, 187:460–494, April 2010.

[18] F. A. Aharonian and A. M. Atoyan. Broad-band diffuse gamma ray emission of the galactic disk. *A&A*, 362:937–952, October 2000.

[19] Z. Arzoumanian, D. J. Nice, J. H. Taylor, and S. E. Thorsett. Timing behavior of 96 radio pulsars. *ApJ*, 422:671–680, February 1994.

[20] W. B. Atwood, M. Ziegler, R. P. Johnson, and B. M. Baughman. A Time-differencing Technique for Detecting Radio-quiet Gamma-Ray Pulsars. *ApJL*, 652:L49–L52, November 2006.

[21] W. B. Atwood et al. Design and initial tests of the Tracker-converter of the Gamma-ray Large Area Space Telescope. *Astroparticle Physics*, 28:422–434, December 2007.

[22] W. B. Atwood et al. The Large Area Telescope on the Fermi Gamma-Ray Space Telescope Mission. *ApJ*, 697:1071–1102, June 2009.

[23] X.-N. Bai and A. Spitkovsky. Modeling of Gamma-Ray Pulsar Light Curves with Force-Free Magnetic Field. *ArXiv e-prints*, October 2009.

[24] A. S. Bennett. The revised 3C catalogue of radio sources. *MmRAS*, 68:163–+, 1962.

[25] R. J. Beran. Asymptotic theory of a class of tests for uniformity of a circular distribution. *The Annals of Mathematical Statistics*, 40(4):1196–1206, 1969.

[26] D. L. Bertsch et al. Pulsed high-energy gamma-radiation from Geminga (1E0630 + 178). *Nature*, 357:306–+, May 1992.

[27] G. F. Bignami et al. The COS-B experiment for gamma-ray astronomy. *Space Science Instrumentation*, 1:245–268, August 1975.

[28] J. Binney and S. Tremaine. *Galactic dynamics.* 1987.

[29] J. A. Biretta, W. B. Sparks, and F. Macchetto. Hubble Space Telescope Observations of Superluminal Motion in the M87 Jet. *ApJ*, 520:621–626, August 1999.

[30] M. Brigida. Moon and quiet sun detection with fermi-lat observatory. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, In Press, Uncorrected Proof:–, 2010.

[31] L. E. Brown, D. D. Clayton, and D. H. Hartmann. Gamma ray pulsar analysis from photon probability maps. In C. R. Shrader, N. Gehrels, & B. Dennis, editor, *NASA Conference Publication*, volume 3137 of *NASA Conference Publication*, pages 267–272, February 1992.

[32] R. Buccheri et al. Search for pulsed gamma-ray emission from radio pulsars in the COS-B data. *A&A*, 128:245–251, November 1983.

[33] F. Camilo et al. Radio Detection of LAT PSRs J1741-2054 and J2032+4127: No Longer Just Gamma-ray Pulsars. *ApJ*, 705:1–13, November 2009.

[34] K. S. Cheng, C. Ho, and M. Ruderman. Energetic radiation from rapidly spinning pulsars. I - Outer magnetosphere gaps. II - VELA and Crab. *ApJ*, 300:500–539, January 1986.

[35] Herman Chernoff. On the distribution of the likelihood ratio. *The Annals of Mathematical Statistics*, 25(3):573–578, 1954.

[36] J. Chiang and R. W. Romani. An outer gap model of high-energy emission from rotation-powered pulsars. *ApJ*, 436:754–761, December 1994.

[37] James W. Cooley and John W. Tukey. An algorithm for the machine calculation of complex fourier series. *Mathematics of Computation*, 19(90):297–301, 1965.

[38] J. K. Daugherty and A. K. Harding. Electromagnetic cascades in pulsars. *ApJ*, 252:337–347, January 1982.

[39] J. K. Daugherty and A. K. Harding. Gamma-Ray Pulsars: Emission from Extended Polar CAP Cascades. *ApJ*, 458:278–+, February 1996.

[40] O. C. de Jager. On periodicity tests and flux limit calculations for gamma-ray pulsars. *ApJ*, 436:239–248, November 1994.

[41] O. C. de Jager and I. Büsching. The H-test probability distribution revisited: Improved sensitivity. *ArXiv e-prints*, May 2010.

[42] O. C. de Jager, B. C. Raubenheimer, and J. W. H. Swanepoel. A poweful test for weak periodic signals with unknown light curve shape in sparse data. *A&A*, 221:180–190, August 1989.

[43] A. J. Deutsch. The electromagnetic field of an idealized star in rigid rotation in vacuo. *Annales d'Astrophysique*, 18:1–+, January 1955.

[44] R. DuPlain, S. Ransom, P. Demorest, P. Brandt, J. Ford, and A. L. Shelton. Launching GUPPI: the Green Bank Ultimate Pulsar Processing Instrument. In *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 7019 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, August 2008.

[45] J. Dyks and B. Rudak. Two-Pole Caustic Model for High-Energy Light Curves of Pulsars. *ApJ*, 598:1201–1206, December 2003.

[46] C. Baixeras et al. Commissioning and first tests of the magic telescope. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 518(1-2):188 – 192, 2004. Frontier Detectors for Frontier Physics: Proceedin.

[47] T. C. Weekes et al. Veritas: the very energetic radiation imaging telescope array system. *Astroparticle Physics*, 17(2):221 – 243, 2002.

[48] Fermi LAT Collaboration. Modulated High-Energy Gamma-Ray Emission from the Microquasar Cygnus X-3. *Science*, 326:1512–, December 2009.

[49] C. E. Fichtel et al. High-energy gamma-ray results from the second small astronomy satellite. *ApJ*, 198:163–182, May 1975.

[50] A. S. Fruchter, D. R. Stinebring, and J. H. Taylor. A millisecond pulsar in an eclipsing binary. In H. Ögelman & E. P. J. van den Heuvel, editor, *Timing Neutron Stars*, pages 163–+, 1989.

[51] T. K. Gaisser. *Cosmic Rays and Particle Physics*. January 1991.

[52] P. Goldreich and W. H. Julian. Pulsar Electrodynamics. *ApJ*, 157:869–+, August 1969.

[53] K. M. Górski et al. HEALPix: A Framework for High-Resolution Discretization and Fast Analysis of Data Distributed on the Sphere. *ApJ*, 622:759–771, April 2005.

[54] J. P. Halpern, M. Eracleous, R. Mukherjee, and E. V. Gotthelf. 3EG J2016+3657: Confirming an EGRET Blazar behind the Galactic Plane. *ApJ*, 551:1016–1023, April 2001.

[55] J. P. Halpern and S. S. Holt. Discovery of soft X-ray pulsations from the gamma-ray source Geminga. *Nature*, 357:222–224, May 1992.

[56] A. K. Harding, J. V. Stern, J. Dyks, and M. Frackowiak. High-Altitude Emission from Pulsar Slot Gaps: The Crab Pulsar. *ApJ*, 680:1378–1393, June 2008.

[57] A. Hewish, S. J. Bell, J. D. H. Pilkington, P. F. Scott, and R. A. Collins. Observation of a Rapidly Pulsating Radio Source. *Nature*, 217:709–713, February 1968.

[58] J. A. Hinton. The status of the hess project. *New Astronomy Reviews*, 48(5-6):331 – 337, 2004. 2nd VERITAS Symposium on the Astrophysics of Extragalactic Sources.

[59] J. D. Jackson. *Classical Electrodynamics, 3rd Edition*. July 1998.

[60] W.N. Johnson, J.E. Grove, B.F. Phlips, J. Ampe, S. Singh, and E. Ponslet. The construction and performance of the csi hodoscopic calorimeter for the glast beam test engineering module. *Nuclear Science, IEEE Transactions on*, 48(4):1182 –1189, aug 2001.

[61] T. Kamae, N. Karlsson, T. Mizuno, T. Abe, and T. Koi. Parameterization of γ, $e^{+/-}$, and Neutrino Spectra Produced by p-p Interaction in Astronomical Environments. *ApJ*, 647:692–708, August 2006.

[62] G. Kanbach et al. The project EGRET (Energetic Gamma-Ray Experiment Telescope) on NASA's Gamma-Ray Observatory (GRO). *Space Sci. Rev.*, 49:69–84, 1988.

[63] W. L. Kraushaar, G. W. Clark, G. P. Garmire, R. Borken, P. Higbie, V. Leong, and T. Thorsos. High-Energy Cosmic Gamma-Ray Observations from the OSO-3 Satellite. *ApJ*, 177:341–+, November 1972.

[64] M. D. Lehnert and T. M. Heckman. The Nature of Starburst Galaxies. *ApJ*, 472:546–+, December 1996.

[65] M. S. Longair. *High energy astrophysics. Vol.1: Particles, photons and their detection*. March 1992.

[66] J. R. Mattox et al. The Likelihood Analysis of EGRET Data. *ApJ*, 461:396–+, April 1996.

[67] M. A. McLaughlin and J. M. Cordes. Gamma-Ray Pulsars: Modeling and Searches. *ArXiv Astrophysics e-prints*, October 2003.

[68] A. A. Moiseev, R. C. Hartman, J. F. Ormes, D. J. Thompson, M. J. Amato, T. E. Johnson, K. N. Segal, and D. A. Sheppard. The anti-coincidence detector for the GLAST large area telescope. *Astroparticle Physics*, 27:339–358, June 2007.

[69] H. I. Nel et al. EGRET High-Energy Gamma-Ray Pulsar Studies. III. A Survey. *ApJ*, 465:898–+, July 1996.

[70] Jorge Nocedal and Stephen J. Wright. *Numerical Optimization*. Springer, August 2000.

[71] M. E. Özel and H. A. Mayer-Haßelwander. A method to improve the visibility of time-variable gamma-ray sourcesin structured background. *A&A*, 125:130–135, August 1983.

[72] P. V. Ramanamurthy, C. E. Fichtel, D. A. Kniffen, P. Sreekumar, and D. J. Thompson. Possible Evidence for Pulsed Emission of High-Energy Gamma Rays by PSR B0656+14. *ApJ*, 458:755–+, February 1996.

[73] R. W. Romani. Gamma-Ray Pulsars: Radiation Processes in the Outer Magnetosphere. *ApJ*, 470:469–+, October 1996.

[74] R. W. Romani and I.-A. Yadigaroglu. Gamma-ray pulsars: Emission zones and viewing geometries. *ApJ*, 438:314–321, January 1995.

[75] P. M. Saz Parkinson et al. Eight gamma-ray pulsars discovered in blind frequency searches of Fermi LAT data. *ArXiv e-prints*, June 2010.

[76] I. G. Shevtsova. Sharpening of the upper bound of the absolute constant in the berry–esseen inequality. *Theory of Probability and its Applications*, 51(3):549–553, 2007.

[77] M. Sikora, M. C. Begelman, and M. J. Rees. Comptonization of diffuse ambient radiation by a relativistic jet: The source of gamma rays from blazars? *ApJ*, 421:153–162, January 1994.

[78] D. A. Smith et al. Pulsar timing for the Fermi gamma-ray space telescope. *A&A*, 492:923–931, December 2008.

[79] A. Spitkovsky. Time-dependent Force-free Pulsar Magnetospheres: Axisymmetric and Oblique Rotators. *ApJL*, 648:L51–L54, September 2006.

[80] M. Tavani et al. The AGILE Mission. *A&A*, 502:995–1013, August 2009.

[81] D. J. Thompson, C. E. Fichtel, R. C. Hartman, D. A. Kniffen, and R. C. Lamb. Final SAS-2 gamma-ray results on sources in the galactic anticenter region. *ApJ*, 213:252–262, April 1977.

[82] D. J. Thompson, C. E. Fichtel, D. A. Kniffen, and H. B. Ogelman. SAS-2 high-energy gamma-ray observations of the VELA pulsar. *ApJL*, 200:L79–L82, September 1975.

[83] D. J. Thompson et al. Calibration of the Energetic Gamma-Ray Experiment Telescope (EGRET) for the Compton Gamma-Ray Observatory. *ApJS*, 86:629–656, June 1993.

[84] S. S. Wilks. The large-sample distribution of the likelihood ratio for testing composite hypotheses. *The Annals of Mathematical Statistics*, 9(1):60–62, 1938.

[85] C. Winkler et al. The INTEGRAL mission. *A&A*, 411:L1–L6, November 2003.

[86] S. Woosley and T. Janka. The physics of core-collapse supernovae. *Nature Physics*, 1:147–154, December 2005.

[87] W.-M. et al. Yao. Review of Particle Physics. *Journal of Physics G*, 33:1+, 2006.

[88] L. Zhang and K. S. Cheng. X-ray and gamma-ray emission from millisecond pulsars. *A&A*, 398:639–646, February 2003.

Appendix A

THE ASYMPTOTIC NULL DISTRIBUTION OF H_M

Recall the H_m statistic is defined as

$$H_m = \max[Z_i^2 - c(i-1)] \equiv \max[X_i],\ 1 \leq i \leq m. \tag{A.1}$$

In its original formulation, $m = 20$ and $c = 4$. The addition of the constant term $c(i-1)$ suppresses contributions from the higher harmonics in the null case. (If $c = 0$, the mth value is always the maximum.)

From the above definition, H_m is an (extreme) order statistic of the X_i, a collection of m dependent, non-identically distributed RVs. While the distribution of such a statistic is often difficult or impossible to obtain in a useful form, in this case, the relatively simple mutual dependence of the variables (they satisfy the Markov property), together with the form of the conditional distributions (exponential), admits a concise, closed form solution for the asymptotic null distribution of H_m.

A.1 The Joint Probability Density Function of $\vec{X}$

Let $X_i \equiv Z_i^2 - c(i-1)$. (For convenience, we interpret Z_0^2 as 0, giving $X_0 = c$.) If we assume the asymptotic distribution for Z_m^2, then X_{i+1} can be obtained from X_i by adding a χ_2^2 distributed variable and subtracting c. That is,

$$f_{X_{i+1}|X_i}(x_{i+1}|x_i) = \chi_2^2(x_{i+1} - x_i + c). \tag{A.2}$$

Now, let $\vec{X}_m$ be a random vector in $\mathbb{R}^m$, such that H_m is the maximum element of $\vec{X}_m$. We construct the joint pdf for $\vec{X}$ as a product of conditional distributions:

$$\begin{aligned}
f_{\vec{X}_m}(\vec{x}_m) =& f_{X_m|\vec{X}_{m-1}}(x_m|\vec{x}_{m-1}) \times f_{X_{m-1}|\vec{X}_{m-2}}(x_{m-1}|\vec{x}_{m-2}) \times \cdots \\
& \times f_{X_2|X_1}(x_2|x_1) \times f_{X_1}(x_1).
\end{aligned}$$

From Eq. A.2, this reduces to

$$f_{\vec{X}_m}(\vec{x}_m) = \prod_{i=1}^{m} \chi_2^2(x_i - x_{i-1} + c), \tag{A.3}$$

demonstrating the Markov property. From this form, it is easy to show that the marginal distribution for X_i is χ_{2i}^2 by reducing the integral over the remaining $m-1$ terms to a series of convolutions. The normalization may also be verified in this way.

Inserting the explicit form for $\chi_2^2(x) = \frac{1}{2}\exp(-\frac{x}{2})\,\theta(x)$, where $\theta(x)$ is the Heaviside step function restricing support to positive arguments, yields a significant simplification:

$$f_{\vec{X}_m}(\vec{x}_m) = \frac{\alpha^{m-1}}{2} \times \left[\prod_{i=1}^{m} \theta(x_i - x_{i-1} + c) \right] \exp\left(-\frac{x_m}{2}\right), \tag{A.4}$$

where we have defined for convenience $\alpha \equiv \frac{1}{2}\exp\left(-\frac{c}{2}\right)$.

A.2 The Cumulative Distribution Function of H_m

H_m is just the maximum element of the vector $\vec{X}_m$. Thus, the probability to observe a value less than or equal to h_m is simply the integral of the $f_{\vec{X}}(\vec{x})$ over all values of $\vec{x}$ with all elements of $\vec{x}$ less than or equal to h_m. That is,

$$F_{H_m}(h_m) = \prod_{i=1}^{m} \left(\int_{-\infty}^{h_m} dx_i \right) f_{\vec{X}}(\vec{X}).$$

Inserting the form obtained in Eq. A.4 yields

$$F_{H_m}(h_m) = \frac{\alpha^{m-1}}{2} \prod_{i=1}^{m} \left[\int_{-\infty}^{h_m} dx_i \theta\left(x_i - x_{i-1} + c\right) \right] \exp\left(-\frac{x_m}{2}\right). \tag{A.5}$$

The two important features of the joint pdf—the simple relation between the dependent variables and the simple form of the conditional distributions—yields an integral expression for F_{H_m} in which the integrand consists of an exponential in a single variable. The main difficulty lies in determining the support of the integrand specified by the product of step functions.

A.2.1 Reduction of $F_{H_m}(h_m)$

We can develop the integral in Eq. A.5 recursively. First, we make a change of variables in the rightmost integral: $u_m \equiv x_m - x_{m-1} + c$. This integral is then

$$\int_0^{h-x_{m-1}+c} du_m \, \exp\left(\frac{-u_m - x_{m-1} + c}{2}\right) = \alpha^{-1} \left[\exp\left(\frac{-x_{m-1}}{2}\right) - \exp\left(-\frac{h+c}{2}\right)\right].$$

The lefthand term, togther with the remaining integrals, is the cumulative distribution function for an $m-1$ harmonic H-test, $F_{H_{m-1}}$. The righthand is the integral of unity with nontrivial limits of integration. That is,

$$F_{H_m}(h_m) = F_{H_{m-1}}(h_m) - \alpha^{m-1} \times \exp\left(-\frac{h_m}{2}\right) I_{m-1}, \tag{A.6}$$

where

$$I_n(h) = \prod_{i=1}^{n} \left[\int_{-\infty}^{h} dx_i \, \theta\left(x_i - x_{i-1} + c\right)\right].$$

Fully reducing F_{H_m} yields a power series in α:

$$F_{H_m}(h_m) = 1 - \exp\left(-\frac{h_m}{2}\right) \times \sum_{n=0}^{m-1} \alpha^n I_n(h_m). \tag{A.7}$$

Thus, the only remaining task is to evaluate I_n.

A.2.2 Evaluation of I_n

We begin with a change of variables to eliminate the step functions. Let $u_i \equiv x_i - \sum_{j=1}^{i-1}(x_j - c)$. Then

$$I_n(h) = \prod_{i=1}^{n} \left[\int_0^{B_i} du_i\right].$$

Here, $B_i = h + (i-1)c - \sum_{j=1}^{i-1} x_j$, and we note that $B_n = B_{n-1} + c - u_{n-1}$. We can evaluate the n integrals recursively. With each integration, we make the change of integration variable to $q_i \equiv B_i + (n-i+1)c - u_i$. For instance, evaluating the rightmost integral, we have

$$\begin{aligned}
I_n(h) &= \prod_{i=1}^{n-1} \left(\int_0^{B_i} du_i\right) B_n \\
&= \prod_{i=1}^{n-2} \left(\int_0^{B_i} du_i\right) \int_0^{B_{n-1}} du_{n-1} \, B_{n-1} + c - u_{n-1} \\
&= \prod_{i=1}^{n-2} \left(\int_0^{B_i} du_i\right) \int_{2c}^{B_{n-1}+2c} dq_{n-1} \, (q_{n-1} - c) \\
&= \left[\prod_{i=1}^{n-2} \left(\int_0^{B_i} du_i\right) \int_{2c}^{B_{n-1}+2c} dq_{n-1} \, q_{n-1}\right] - cI_{n-1}.
\end{aligned}$$

This form is typical as one continues to integrate. The change of variable always produces a monomial in the integration variable. The upper boundary then produces a term in the integration variable of the next integral while the lower boundary produces a monomial of c. This separation allows a recursive development for I_n, and combining the recursive terms yields

$$I_n(h) = \frac{(h + nc)^n}{n!} - \sum_{j=1}^{n} I_{n-j-1} \frac{(jc)^j}{j!}. \tag{A.8}$$

We note that $I_0(h) = 1$ and $I_1(h) = h$. The use of logarithms facilitates the numerical evaluation of this expressions for large h and n.

A.2.3   Monte Carlo Validation

To check our results, we performed Monte Carlo simulations of the H_{20} statistic in the asymptotic null case. Specifically, for each realization of H_{20}, we drew 20 realizations from a χ^2 distribution with one degree of freedom and determined H accordingly (with $c = 4$.) In Figure A.2.3, we show the results of 10^9 Monte Carlo trials for a variety of maximum harmonics, viz. H_1, H_2, H_4, H_8, and H_{20}. The results are in good agreement with the asymptotic distribution derived here. Next, we focus on H_{20} with 10^{10} trials, shown in Figure A.2.3. These results are also in good agreement.

A.2.4   Behavior at Large H

Finally, we consider the behavior of the null distribution at large values of H_{20}. We have seen that $\exp(-0.4H)$ is an excellent approximation for the true survival function for any "practical" application[1]. Nonetheless, it is interesting to consider very large values of H to investigate the trend of the true distribution relative to the simple exponential approximation. We show the ratio of the two expression in Figure A.2.4. The exponential first-order in H becomes a poor approximation after $H \approx 100$ (again, sufficiently large for any practical need!). By performing a least squares fit to minimize the difference between the asymptotic distribution and an approximation of the form $\exp(-aH - bH^2)$, we obtain an expression

---

[1]Finite sample size effects are almost certainly larger than the discprepancy between the approximation and the asymptotic distribution once H_{20} is of order 50.
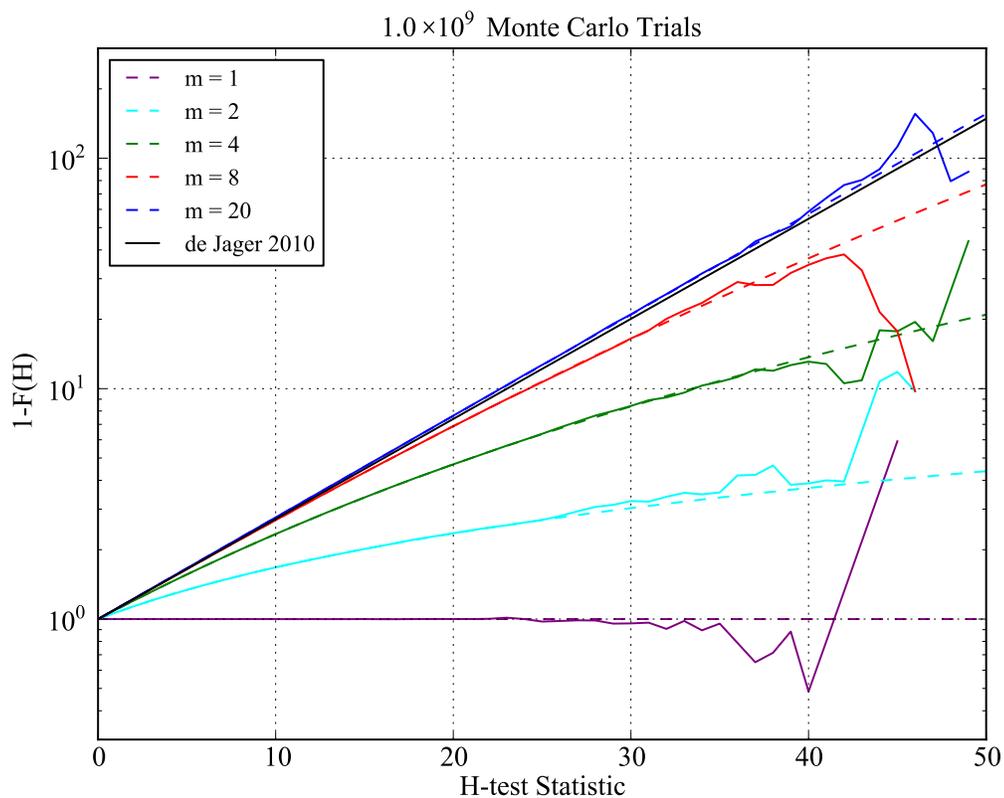
Figure A.1: We show the survival function $(1 - F(H))$ of the asymptotic distribution for H and for a sample distribution $(N = 10^9)$ drawn from the null distribution by simulation. To reduce the scale, we divide the survival function for a χ^2 variable with two degrees of freedom, viz. $1 - F(x) = \exp(-0.5x)$. For each maximum harmonic, the sample distributions agree with the asymptotic calibration.
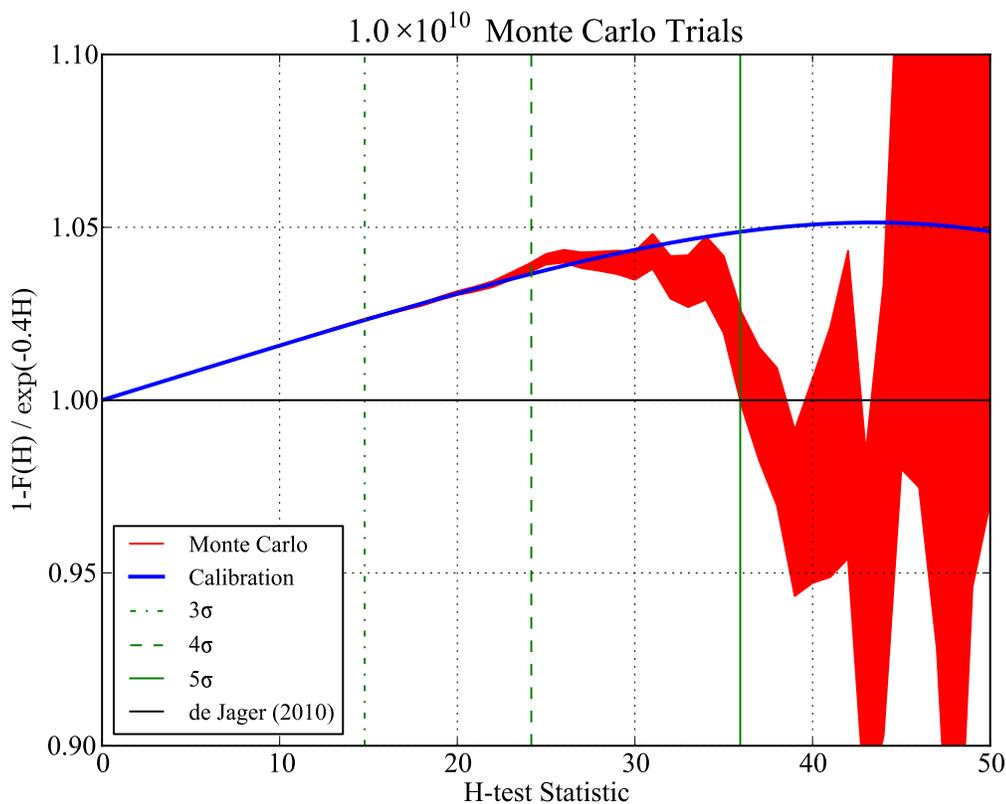
Figure A.2: We show the survival function $(1 - F(H))$ of the asymptotic distribution for H and for a sample distribution $(N = 10^{10})$ drawn from the null distribution by simulation. To reduce the scale, we divide the survival function by the "first-order" approximation $1 - F(H) \approx \exp(-0.4H)$[41]. This approximation gives excellent results ($< 5\%$ deviation from the true distribution) to $H = 50$ and above. We have marked the values H corresponding to (two-sided) significance levels of 3σ, 4σ, and 5σ for reference.
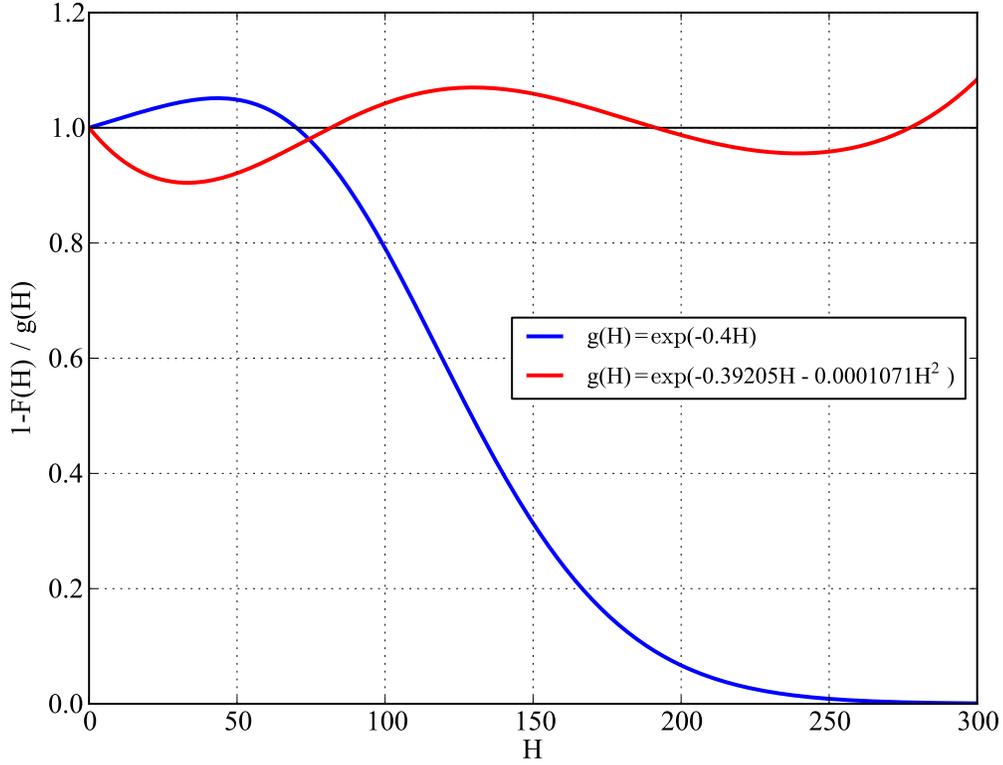
Figure A.3: The ratio of the true asymptotic distribution for H_{20} to the approximate expression $\exp(-0.4H_{20})$. At $H_{20} \approx 40$, the survival function of the asymptotic distribution begins to decrease more rapidly than exponential approximation, indicating the importance of a quadratic term. We show such a term, determined by least squares, with the red curve, which provides an accurate estimation of the survival function over a broad range.

accurate to 10% over the range $0 \leq H \leq 300$,

$$1 - F(H) \approx \exp(-0.39205H - 0.0001071H^2).$$
(A.9)

Appendix B

PULSATION SEARCHES WITH NON-UNIFORM NULL DISTRIBUTIONS

As discussed above in Chapter 7^1, when the null distribution of the phases departs from uniformity, the ψ_k can retain independence even for large sample sizes. Below, we outline two approaches to partially overcome this deficiency.

First, if the departure from non-uniformity is in some sense small, we can proceed as before. We must modify the definition of the ψ_k to account for the non-uniformity so that they are once again asymptotically normally distributed. We begin by explicitly calculating the null distribution $F_\Phi(\phi)$, based on the IRF and S/C pointing history. Essentially, $f_\Phi(\phi) \approx \epsilon(\phi)$, i.e., the null distribution is roughly the exposure calculated as a function of phase. The approximation comes in choosing a shape for the (typically unknown) source spectrum used as a weight when summing the exposure over energy, and in the form of time-dependent background leakage. Provided we are able to adequately characterize the IRF and model the time-dependent background, we can calculate the null distribution $F_\Phi(\phi)$ and thus the moments of the G_k. Then,

$$\sqrt{n}(\psi_k - \mu_{G_k})/\sigma_{G_k} \sim \mathcal{N}, \tag{B.1}$$

as before. The moments are somewhat cumbersome to calculate as the distributions of G_k are now not monotonic. For an arbitrary phase distribution $F_\Phi(\phi)$, the distribution of c_k is given by

$$F_{C_k}(x) = \sum_{i=1}^{k} \int_{(2i-1+\delta(x))/2k}^{(2i-1-\delta(x))/2k} d\phi \, f_\Phi(\phi) \tag{B.2}$$

with $\delta = 1 - \arccos{(x)}/\pi$. The distribution for S_k may be obtained from Eq. B.2 by letting $i \to i + 1/4$. These distribution functions are in turn used to calculate the required moments μ and σ.

---

[1] We adopt the notation of Chapter 7 here.

Now, provided that the departure from uniformity is mild, we expect the same arguments about the asymptotic dependence of ψ_k to apply, at least approximately, and a Z_m^2 statistic constructed using these new ψ_k will follow, again approximately, a χ_m^2 distribution. However, the risk of overestimating the significance of a detection due departure from the χ_m^2 calibration leads us to prefer a second method.

In a second approach, we again calculate the null distribution $F_\Phi(\phi)$ explicitly from the exposure. However, we now switch our fundamental set of rvs from $\{\phi_i\}$ to $\{F_\Phi(\phi_i)\}$, i.e., the cdf evaluated at the observed phases. If we have calculated $F_\Phi(\phi)$ correctly, then the $\{F_\Phi(\phi_i)\}$ are guaranteed to follow a uniform distribution in the null case. Further, $F_\Phi(\phi_i) \in [0, 1)$, so we can use the $\{F_\Phi(\phi_i)\}$ directly in the (weighted and unweighted) tests we have already developed. The drawback to this approach is a distortion of the input signal by "processing" by $F_\Phi(\phi_i)$. Even worse, this processing depends on absolute phase, i.e., we lose the invariance under translations on the circle that has been a property of the statistics we have outlined[2]. However, we should not be surprised by this. The departure from uniform exposure inevitably increases our sensitivity to some signals and decreases it for others. A few examples appear in Figure B where we have taken some simple analytic distributions—a uniform distribution, a decaying exponential[3], and a sinusoid— for the exposure and processed a sharp signal—represented by a von Mises distribution. We see that, e.g., when the signal is aligned with the peak of the sinusoidal exposure, the processed signal is broadened, whereas if it is near the minimum of the exposure, the signal is sharpened. (This gain is offset by the appreciable decrease in the number of collected photons.)

---

[2]To be clear—the test statistic itself is still independent of absolute phase, but the processed signal $F_\Phi(\phi)$ *does* depend on the absolute phase

[3]An exponential distibution is a somewhat pathological example since it is discontinuous.
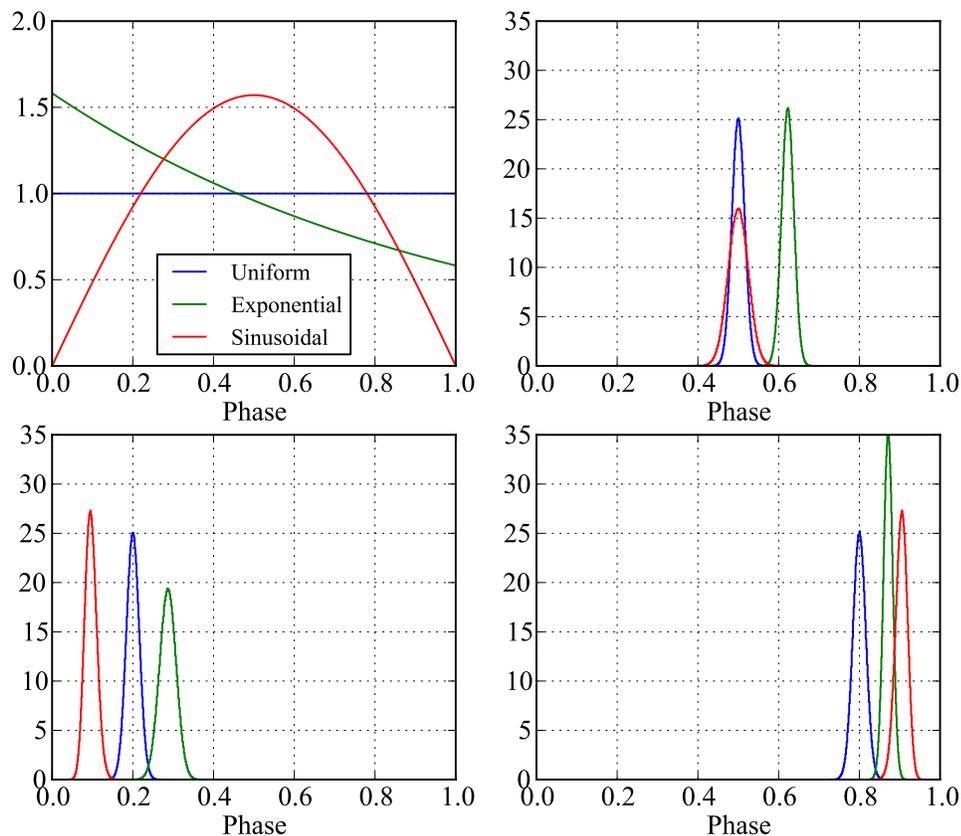
Figure B.1: The resulting "signal" after processing an input with $F_\Phi(\phi_i)$ as outlined in the text. We simulate 10^6 realizations from a von Mises distribution with a sharp peak and then generate histograms of $F_\Phi(\phi_i)$ for each of three distributions. The uniform distribution is trivial and represents the unprocessed signal. A sinusoidal distribution might arise, e.g., if the period of interest is very close to twice the orbital period of the S/C. An exponential distribution is difficult to produce physically but provides insight into asymmetric distributions. In the upper left panel, we plot $f_\Phi(\phi)$ for each of the exposure scenarios. In the remaining three panels, we show the processed signal for three different absolute phases of the signal. Clockwise from upper right, the signal is centered $\phi = 0.5$, 0.8, and 0.2.