A Transdimensional Perspective on Dark Matter

Citation

Permanent link

Terms of Use

Share Your Story

A Transdimensional Perspective on Dark Matter

Thesis advisor: Professor Douglas P. Finkbeiner   Tansu Daylan

A Transdimensional Perspective on   Dark Matter

Abstract

Robust uncertainty propagation and marginalization over nuisance parameters is the key to perform robust inference. This thesis introduces a transdimensional, hierarchical, threshold-free, and Bayesian inference framework. Colloquially referred to as *probabilistic cataloging*, the paradigm propagates within and across model covariances, reduces mismodeling and information loss due to thresholding based on statistical significance, and accelerates computation to ensure scalability. The main product is a `Python` implementation of probabilistic cataloging called the Probabilistic Cataloger (`PCAT`). The code and its documentation are available on `GitHub` and `readthedocs`, respectively, for the use of the scientific community.

A central problem in contemporary cosmology is dark matter. Astrophysical data sets such as telescope images that aim to probe the characteristics of dark matter, require covariant models and aim to recover information contained in data features potentially caused by model elements with low statistical significance. Using probabilistic cataloging, the thesis presents two inferences of the properties of dark matter, e.g., self-annihilation of Weakly Interacting Massive Particles (WIMPs) in the inner Milky, and the small-scale structure of dark matter in galactic halos. The role of various priors in probabilistic cataloging are also studied.

Contents

To the Universe, which has placed intricate puzzles into its workings so we may keep on asking more questions than we can answer, making this precious life meaningful and worth living.

Acknowledgments

First of all, I thank my thesis advisor Douglas P. Finkbeiner for his contributions in the development of the research program presented in this thesis. His interest and deep insight in data science has largely motivated me to explore transdimensional inference and allowed me to work on problems that I genuinely enjoyed, making this PhD a truly pleasant and productive intellectual exercise. I also thank the other members of my thesis advising committee, Daniel Eisenstein and Matt Reece, for all the useful discussions during my PhD.

I thank my lovely wife Ece for the endless joy she brings to my life. Her compassion and attachment have alleviated the hard times of this PhD. I also thank my mother Tülay for bringing me up to a life full of opportunities and for her endless support in my academic work. Although my father did not live long enough to see my scientific career, he continues to live in my vision, and I thank him for everything he has done to broaden my horizon. My brother Tonra has been a great support and a true loadstar in life. I thank him and his beloved wife Kezban for their invaluable amity. My cute nieces Beren and Duru appeared halfway through my PhD, and have cheered me up with their ultimate sweetness.

During my PhD, I had the benefit of interacting with a vibrant academic community. I deeply thank Cora Dvorkin, Tracy Slatyer, Dan Hooper, Lisa Randall, Chistopher W. Stubbs, David W. Hogg, Charlie Conroy, Howard M. Georgi, Andrew Strominger, Cumrun Vafa, Manoj Kaplinghat,

founders of the Republic of Turkey, especially Mustafa Kemal Atatürk, for establishing the pillars

of the secular education system in which I was able to perceive the world from all angles and freely

wonder about the workings of the Universe.

0

Introduction

As humankind collects more data on the observable Universe, relating these observations to the underlying models becomes increasingly harder. The trivial approach in this process would be to construct an independent, adhoc model for *each and every* observation, successfully explaining *all* possible observations. This would be an example of extreme overfitting and not useful for making predictions based on the learned models. Instead, the scientific method makes progress by explain-

ing as much data as possible with the least model complexity. The ultimate success towards this goal would be the construction of a theory of everything that is consistent with *any* observation up to the measurement uncertainty. However, as a model grows in context during this process and needs to be consistent with a larger and more diverse data set, updating one's beliefs on this model based on the arrival of new data, becomes increasingly difficult.

A fundamental problem in contemporary cosmology is the missing matter in the Universe. A transparent, gas-like substance that dominates the dynamics at galactic and larger scales, dark matter hypothesis is at the forefront of research in cosmology. In order to test the existence of dark matter or to infer its mass and potential interactions, one has to collect large data sets, counting cosmic rays and photons as a function of space, time, and energy.

Whether recorded by a particle detector or a Charged Coupled Device (CCD), cosmological data sets are essentially digitized outputs of voltage amplifiers. Because there is only a *rather* indirect relation between these voltage measurements and the underlying cosmological model parameters, a large number of nuisance parameters need to be marginalized over when performing inference. Moreover, both the observed data and the model parameters can have covariances, i.e., multiple configurations can yield the same level of consistency between the model and the data. The former arises when there is correlated noise in the measurements and the latter can be due to the parametrization choice of the underlying model. These covariances are crucial in robustly executing the scientific method, since uncertainty propagation has to be performed based on the knowledge of such covariances.

This thesis implements a novel statistical inference framework, *probabilistic cataloging*, that can

learn about hypotheses based on observed data. It generalizes the treatment of such covariances to the *transdimensional* context, where multiple models of different complexity can be simultaneously consistent with the observed data. This is a very frequently encountered, but rarely properly handled inference mode that has ubiquitous applications across all frontiers of science, including cosmology.

The remaining of this thesis is structured as follows. Chapter 1 presents the problem of missing matter in the Universe. Then, Chapter 2 introduces the notion of probabilistic cataloging. The following two chapters showcase two different astrophysical applications of probabilistic cataloging to the problem of dark matter. In particular, Chapter 3 presents results regarding the gamma-ray sky including the anomalous emission in the inner Milky Way, Chapter 4 illustrates a transdimensional approach to modeling gravitationally lensed images. Chapter 5 then presents the software implementation of probabilistic cataloging, PCAT. Chapter 6 concludes the thesis. The contents of the Sections 3.2, 4, 5.2, and 5.3 have been previously published in[74] and[73] with minor changes.

1

Missing matter in the Universe

Photons in the Universe carry information about the underlying physical processes. A unique, isotropic component of this electromagnetic field is the Cosmic Microwave Background (CMB)[175]. The nearly-thermal spectrum of the emission as measured by the Far-Infrared Absolute Spectrophotometer (FIRAS)[152] is consistent with the Universe having a hot and dense past[81]. In this Big Bang model, the Universe adiabatically cools to eventually allow the electrons and protons to recom-

bine, causing CMB photons to stream freely. The observed temperature of the CMB monopole, 2.725 ± 0.002 K[153], allows one to infer that the Universe must have been about 380 thousand years old when CMB photons last scattered with the cosmic plasma. Likewise, as shown in Figure 1.1[10], the small ($\sim 10\mu$K) spatial anisotropies of the CMB temperature[205,117,76,206,179] indicate that the density perturbations at the time of last scattering, must have also been small. However, if these initial density fluctuations are evolved forward in time under the influence of gravity, no galaxies are predicted to form within the inferred lifetime of the Universe. As shown in the top left panel of Figure 1.2, consistency with the observation that the Universe today is structured up to supercluster scales, then, requires that there be an additional, at most weakly interacting (i.e., in addition to gravitation), cold (i.e., non-relativistic long before the last scattering of the CMB photons), dominant matter component in the Universe. This missing form of matter is referred to as dark matter.

The case for the existence of dark matter would be less convincing had it not been predicted before the discovery of the CMB. While studying the line-of-sight motions of galaxies in the Coma cluster, Fritz Zwicky observed in 1930s that the galaxies were moving at a speed (i.e., with a line-of-sight velocity dispersion of $\sigma \sim 1000$ km s^{-1}) that would not allow the luminous (i.e., stellar) mass of the cluster to hold the galaxies together, naming the missing matter as "Dunkle Materie" (Dark Matter)[244]. Since then, similar Doppler observations of the optical light from stars in galaxies[187] and the 21 cm line emission of neutral Hydrogen at the outer regions of galaxies[185] have indicated a non-Keplerian rotation curve. Furthermore, X-ray emission of early type galaxies also revealed presence of hot gas in their halos[97], which again lead to the prediction of additional mass to support the observed fast rotation of the X-ray emitting gas. Another series of data sets that require dark matter are
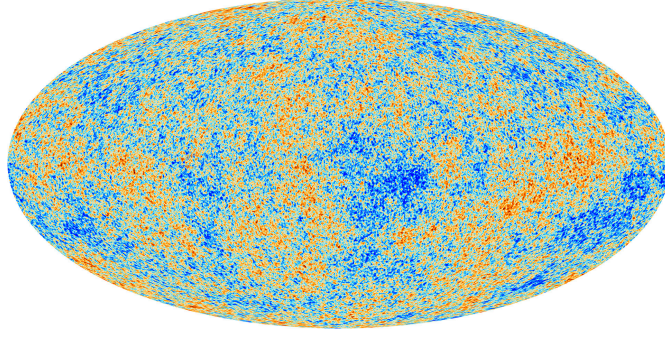
Figure 1.1: The full sky map of CMB temperature anisotropies (i.e., leaving out the monopole) inferred by finding an independent component of the observed microwave sky common in all bands as measured by the Planck satellite [10].

observations of gravitational lensing. Gravitational lensing is the coupling of photons to gravitation, leading to the bending of light due to the presence of mass[85]. Consistency between the observed light deflection with the mass models, require that the mass budget of the deflectors be dominated by non-luminous matter.

Furthermore, in the Big Bang model, the hot beginning of the Universe allows efficient thermonuclear reactions. As a result, light nuclei must have been synthesized in the first few minutes of the Universe[16,46,64]. Requiring consistency between the predicted abundances of such stable primordial elements, such as ^{2}H, ^{3}He, ^{4}He, and ^{7}Li, with their measurements in low metallicity clouds at high redshift, allows the determination of the primordial baryonic abundance relative to that of photons, indicating that baryons cannot make up more than $\sim 5\%$ of the mass-energy budget of the Universe.

Dark matter is not the only hypothesis that can address these data sets. The leading alternative theories are Modified Newtonian Dynamics (MOND)[157], a phenomenological modification to the Newtonian gravity and Modified Gravity (MOG)[160], a Yukawa theory of gravity in the weak field approximation, as well as their relativistic generalizations TeVeS[25] and Scalar-Tensor-Vector Gravity
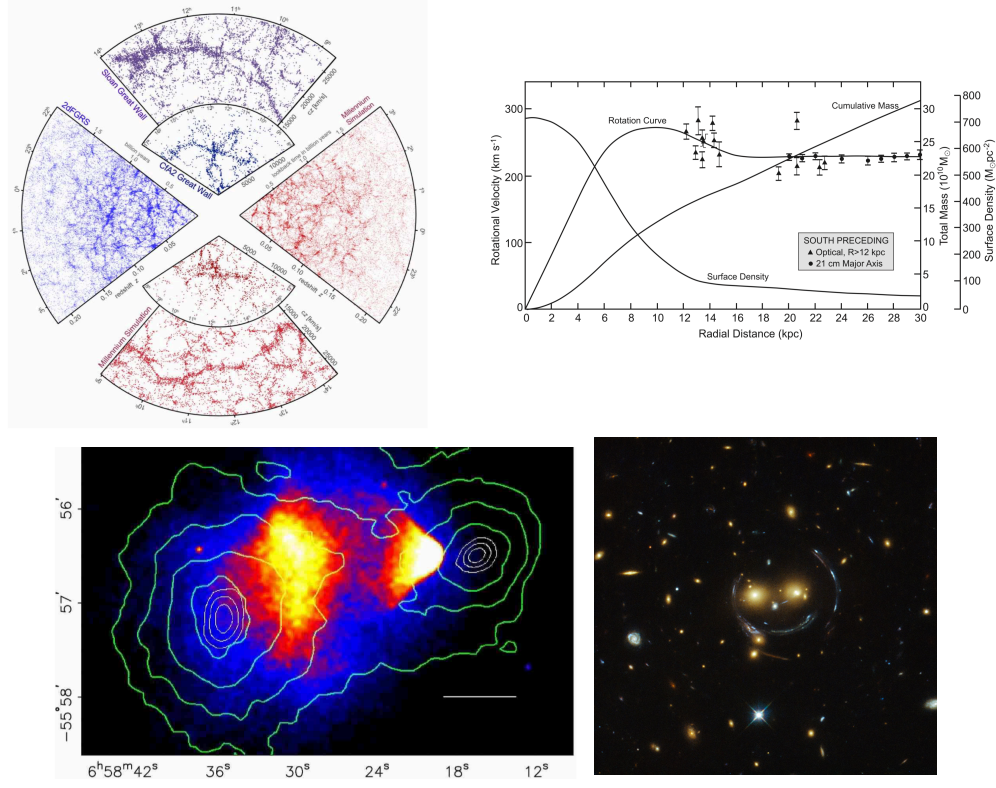
Figure 1.2: Some of the other data sets that require dark matter. Top left: Redshift distribution of galaxies spectroscopically measured by the CfA Redshift Survey [103] (darker blue), the Two Degree Field Galaxy Redshift Survey (2dFGRS) [60] (brighter blue), and the Sloan Digital Sky Survey (SDSS) [238] (violet), compared with the simulated catalogs drawn from the Millenium simulation [210], as reproduced from [208]. The filaments and walls in the large-scale structure of the Universe are the pancake-like structures due to the gravitational instability predicted by Zeldovich [242]. In addition to explaining the observed large-scale structure, cosmological N-body simulations with dark matter also predict the observed excess correlation at the sound horizon at the time of decoupling, i.e., the Baryonic Acoustic Oscillations [86]. Top right: The observed rotation curve of M31, reproduced from [184]. The triangles indicate the Doppler measurements of optical light from stars [187], whereas circles show those of the 21 cm line emission of neutral Hydrogen [185]. Bottom left: The X-ray image [150] of a pair of merging galaxy clusters, 1E 0657-558, taken by the Chandra telescope, superposed with the gravitational potential inferred [57] from the lensing of optical emission of background galaxies. The X-ray emission from the collisional hot gas does not correlate well with the mass distribution of the system, suggesting that collisionless dark matter accounts for most of the mass in the system and does not interact with the baryons. Bottom right: Hubble Space Telescope (HST) image of a strong lens system, SDSS J1038+4849, where the foreground galaxy cluster bends the light from the background galaxies, producing multiple images. The shape, radius, and brightness of the resulting arcs require the foreground system to be more massive than the stellar and gas mass inferred from the observed optical light.

(STVG), respectively. The former modifies the acceleration at the small-acceleration limit, leading to higher rotation at larger galactocentric radii, whereas the latter proposes a stronger gravitational interaction that is partly canceled by a repulsive, finite-range force only at small scales to reproduce the Newtonian gravity. Even though such models make predictions consistent with some observations, they fail to be consistent with all anomalies that the dark matter hypothesis can explain.

It is also true that some of the non-luminous matter is in the form of black holes, neutron stars, planets, and possibly dim stars such as brown and white dwarfs. However, observations of microlensing indicate that the abundance of such compact objects with masses above $\sim 10^{-7} M_\odot$, except around a window of $\sim 10 - 100 M_\odot$ cannot account for the required amount of dark matter[15,219,98].

1.1 Standard Model of Cosmology

Modeling of the observed CMB anisotropies requires that the geometry of the Universe be close to flat, i.e., that the inner angles of a cosmologically-sized triangle sum to $180°$. However, the observed abundance of galaxy clusters only yields 30% of the mass that could close the Universe, leading to a cosmological model where an additional, spatially uniform component, i.e., dark energy, accounts for the rest of the mass-energy budget of the Universe.

There are several observations that probe the equation of state of dark energy. For instance, the luminosity distance to type IA Supernovae (i.e., standard candles with fixed energy output) are observed to be higher than that predicted for a universe decelerating under the gravitational influence

of matter alone[182,176]. Measurements of the baryonic acoustic oscillations in the large-scale structure also require a dark energy component, in order to explain the redshift evolution of the angular diameter distance. In this model, because dark energy becomes dominant only at late times, acoustic peaks in the CMB power spectrum alone cannot break the geometric degeneracy and constrain dark energy. Nevertheless, the correlation of galaxy density and the CMB temperature at large scales indicates that gravitational potentials of voids and clusters must have been decaying recently[113], which can be explained by dark energy via the late-time integrated Sachz-Wolfe effect[189]. Further evidence for dark energy comes from the gravitational lensing of the CMB by the large-scale structure, given observations of the CMB power spectrum at small multipoles[201].

Bringing dark energy and dark matter together along with the baryonic matter and radiation, produces the concordance Λ Cold Dark Matter (ΛCDM) model. As shown in Figure 1.3, A joint fit to the cosmological data sets yields the abundance of dark matter, dark energy, and baryonic matter to be $\sim$ 27%, 68%, and 5% of the mass-energy budget of the Universe[10].

CMB observations also introduce a challenge to the Big Bang model. The observation of a large monopole of the CMB temperature requires distant points on the sky with non-intersecting particle horizons, to be causally connected. This is achieved in the inflation model[116,14,147], where the Universe goes through an exponential expansion before the Big Bang, which allows CMB photons to be in thermal contact prior to inflation. The field responsible for the expansion, the inflaton, eventually decays, reheating the Universe and reproducing the hot and dense conditions of the Big Bang model. Along with ΛCDM, Big Bang and inflation form the standard model of cosmology. Testing and updating this concordance model is the central problem of cosmology.
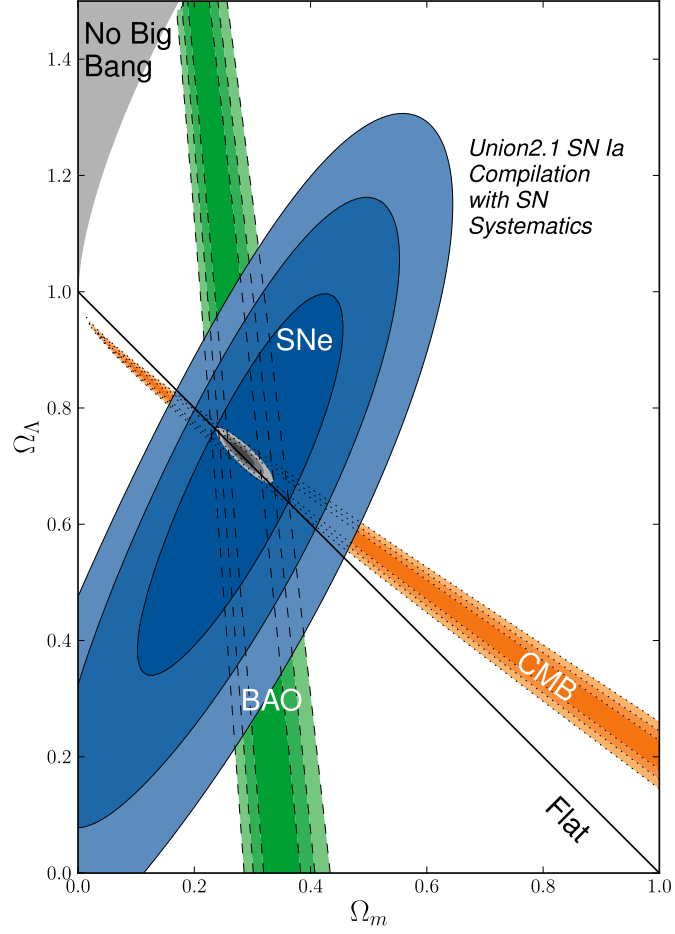
Figure 1.3: Energy-mass budget of the Universe, modeled by ΛCDM and constrained by the cosmological data sets [215].

1.2 SMALL-SCALE STRUCTURE OF DARK MATTER

In ΛCDM, the structure is thought to have formed by the initial collapse of the rarest (most un-likely) matter density fluctuations that grew to be nonlinear into self-gravitating halos, and then, by the successive accretion of smaller halos by the more massive halos. The radial profile of the mass density of these halos can be modeled with a broken power-law, i.e., the Navarro Frenk White (NFW) profile[84,164],

$$\rho(r) = \frac{\rho_0}{(r/r_s)^\gamma (1 + r/r_s)^{3-\gamma}} \tag{1.1}$$

where r is the halocentric radius, r_s is the scale radius, γ is the inner log-slope, and ρ_0 is a normaliz-ing constant. Since structure formation still continues today, only few and dynamically unrelaxed structures exist at the largest scales.

ΛCDM is consistent with observations of the large-scale structure of the Universe[71,31,30,72]. It can fit the power-spectrum of the temperature and polarization anisotropies of the CMB, with some exceptions at the largest scales, e.g., the cold spot[63], whose downward temperature fluctuation could plausibly be due to cosmic variance. At smaller scales, around $k \sim 10 \, \text{Mpc}^{-1} h$, observations of the Lyman-α forest, i.e., absorption spectra of distant quasars, are also consistent with the nonlin-ear growth of perturbations in ΛCDM. In particular, the one-dimensional power spectrum of the Lyman-α forest[173,172,237] measured using the Baryon Oscillation Spectroscopic Survey (BOSS)[87] and

XQ-100[148], can be fitted well with the predictions of the ΛCDM.

At yet smaller scales, nonlinear growth of structure causes the power spectrum of matter fluctuations to be ineffective in characterizing the matter distribution. Furthermore, there is little evidence to believe that ΛCDM is the correct description of the Universe at smaller scales. Discrepancies between observations and ΛCDM predictions have been claimed to remain at subgalactic scales[45,78]. For instance, the number of observed satellites in the Local Group (∼ 50) falls short of the much larger numbers of low-mass subhalos (> 1000) predicted by N-body simulations of ΛCDM[139,162]. Furthermore, the projected density profiles of the observed subhalos are rather shallow in the central regions compared to the much cuspier profiles predicted by ΛCDM[96,161,47,106,231,171].

These apparent discrepancies can be resolved in several ways. First, the inclusion of baryonic feedback processes within numerical simulations[43,192,235,80,193,62,102,212,167,43,17] can significantly reduce the tension between predictions and observations. Another venue to resolve the tension is to modify the microphysics of dark matter by, for instance, postulating that dark matter particles interact with each other[207,239,70,59], or with a relativistic species[51,33,6,94,133,221,66,54]. Both possibilities result in a modification to the internal density profile of dark matter halos, while the latter could also suppress the number of small-mass subhalos[44,195,230]. Alternatively, the discrepancy could be due to a suppression of the small-scale matter power spectrum prior to structure formation. This may arise from the free streaming of dark matter particles after they kinetically decouple from the rest of the cosmic plasma, which partially erases inhomogeneities at the ∼ kpc comoving scale[36,69,32,37,38]. Yet another way to modify ΛCDM is to relax the coldness assumption. Given that the main motivation to postulate dark matter is to seed the Universe with density fluctuations, dark matter cannot be hot

(i.e., relativistic), which would allow it to free-stream after decoupling from the rest of the cosmic plasma in the early Universe. Only some, less than 1% of dark matter is hot, in the form of neutrinos. However, current measurements of the Lyman-α forest are consistent with some amount of free-streaming. Furthermore, dark matter can also be an ultralight (with mass $\sim 10^{22}$ eV) scalar field, in which case the Heisenberg uncertainty principle would prevent the resulting bosonic dark matter waves with a de Broglie wavelength of ~ 1 kpc from collapsing, such as in the Fuzzy Dark Matter model[130]. Last, the discrepancy can be resolved by introducing a small-scale suppression, e.g., a kink or running spectral index, to the *primordial* power spectrum assuming an inflation with broken scale-invariance[132,158].

In any case, the possibility of resolving the potential small-scale crisis of ΛCDM makes a strong case for probing the subgalactic structure of dark matter observationally. Chapter 4 will present an inference framework that learns about the statistical properties of mass substructure in strong lenses given high-resolution optical images of such systems.

1.3 Particle nature of dark matter

Cosmological dark matter problem described above requires an electromagnetically neutral, non-relativistic, dissipationless, and long-lived particle. However, there is no such state in the Standard Model of particle physics[123,89,233]. Furthermore, although there are some anomalies, the current evidence for dark matter comes solely from its gravitational interactions.

A large spectrum of hypotheses have been put forward to realize a particle theory for dark matter.

Some of these are asymmetric dark matter that relates the observed asymmetry between matter and antimatter to a dark sector[190,169]; axions, i.e., a Bose-Einstein condensate due to the misalignment mechanism proposed as a solution to the strong Charge Parity (CP) problem in quantum chromodynamics[174]; localized, stable, and bosonic field configurations called Q-balls[186,99,58]; the gauge fermion mediating supergravity interactions, the gravitino[88]; sterile neutrinos that do not interact via the weak force, but that still mix with the active neutrinos[82,202,1]; and primordial black holes that may have collapsed in the early Universe[119]. However, this thesis will discuss a yet different candidate, Weakly Interacting Massive Particles (WIMPs)[131,142] and their indirect detection.

The first WIMP candidate was proposed as a new generation of heavy neutrinos[142]. However, this is now ruled out[115] since the Large Electron Positron Collider (LEP) measurement of the decay width of the Z boson with mass m_Z[163], constraints the number of active (i.e., those with nonnegligible left-handed component) neutrinos with mass below $m_Z/2 \sim 45$ GeV, to three.

WIMPs can also be produced by several ultraviolet-complete theories. One way is to extend the Standard Model with a symmetry between fermions and bosons called supersymmetry. This leads to the prediction of heavy (i.e., $\mathcal{O}(100 \text{ GeV})$) fermionic states called neutralinos that are linear combinations of the fermionic superpartners of the neutral gauge bosons and the Higgs scalar in the Standard Model. However, heavy particles eventually decay, unless this is forbidden by a symmetry, i.e., conservation law. Further imposing the conservation of a discrete parity, $\mathbb{Z}_2$, makes neutralino stable, yielding a WIMP candidate. A similar symmetry allows the first Kaluza-Klein excitation[52,200,18] and the lightest odd-parity particle of the little Higgs[234,20,107], to produce WIMPs.

In the Big Bang model supported by the discovery of the CMB, the Universe must have been hot

in the beginning following the reheating of the cosmic plasma with the decay of the inflaton field. The simplest picture in this environment is to propose that WIMPs denoted by χ with mass m_χ and Standard Model particles, denoted by SM, are efficiently exchanging energy, i.e., are in thermal equilibrium. Furthermore, because particles in the early Universe have high kinetic energies and are ultra-relativistic, both backward and forward reactions,

$$\chi\chi \rightleftharpoons \text{SMS}\bar{\text{M}}, \tag{1.2}$$

proceed readily. However, as the temperature of the cosmic plasma, T, goes below m_χ, backward reactions into χ and $\bar{\chi}$ become exponentially unlikely. As a result, the continuing forward reaction suppresses the number density of χ,

$$n_{\chi,eq} \propto (m_\chi T)^{3/2} \exp\left(-m_\chi/T\right). \tag{1.3}$$

The suppression continues until the rate of forward interactions drops below the rate with which the Universe expands, resulting in the freeze-out of the WIMP density at $n_{\chi,fo}$, as shown in Figure 1.4. The relic WIMP abundance in units of the critical density of the Universe then becomes

$$\Omega_\chi h^2 \propto \frac{3 \times 10^{-27} \, \text{cm}^3 \text{s}^{-1}}{\langle \sigma v \rangle} \tag{1.4}$$

for the mass at which ultraviolet-complete theories predict the existence of new, heavy states, i.e., $m_\chi \sim 100$ GeV. Here, h is the reduced Hubble constant, $h \equiv 10^{-2} H$. Given the cosmological measurement of $\Omega_\chi \sim 0.1$, this yields the annihilation cross section of $\langle \sigma v \rangle \sim 3 \times 10^{-26}$ cm^3 s^{-1}.

The annihilation cross section of WIMPs depend on the WIMP mass as $\propto m_\chi^2 / m_Z^4$ for masses below m_Z and as $\propto 1/m_\chi^2$ at larger masses. Therefore, since the relic WIMP abundance is inversely proportional to the annihilation cross section, m_χ cannot be smaller than ~ 2 GeV[142] or larger than ~ 400 GeV[115] since the coupling constant cannot grow indefinitely because of the conservation of probability (i.e., wavefunction unitarity). Mediator particles can also exist that increase the annihilation cross section, relaxing the bounds on m_χ.

Currently, there are only upper limits on the potential non-gravitational interaction, i.e., scattering[19] and annihilation[11], cross sections of WIMPs. An observation that can only be explained by a nongravitational interaction of dark matter would be an independent and stronger evidence for the existence of dark matter.

There are three ways with which one can expect to probe WIMPs as illustrated in Figure 1.5. First, WIMPs can be produced in high energy particle collisions such as in the Large Hadron Collider (LHC). The observational signature would be missing energy and momentum, since the produced WIMPs would escape the detector without any further interaction. Second, nuclei in low-
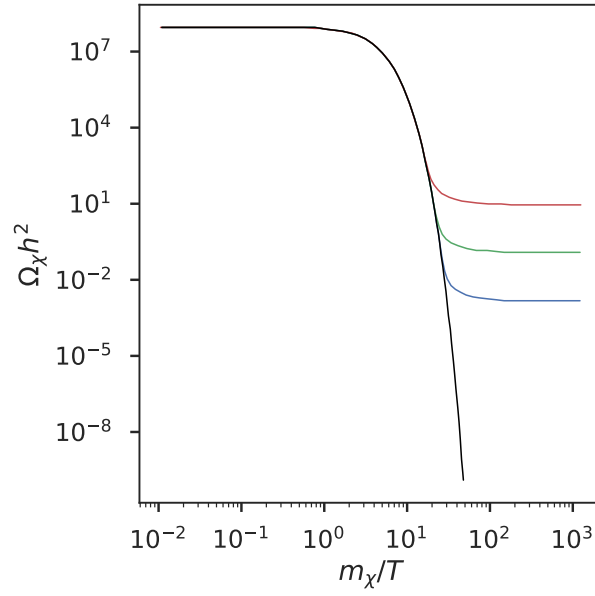
Figure 1.4: Dark matter abundance during thermal freeze-out[27]. If WIMPs with $m_\chi \sim 100$ GeV, thermally freeze out, their relic abundance can be consistent with the amount of non-baryonic matter required to close the mass-energy budget of the Universe. Colored lines show the abundance of WIMPs that freeze out with different annihilation cross sections that increase from red to blue. The black line shows the equilibrium abundance.

background detectors can be expected to collide with WIMPs moving in the galactic halo, resulting in observable the recoil of nuclei. Energy distribution of the recoiling particles as well as their temporal (annual) variation due to the varying velocity of the Solar System in the galactic frame, would allow discrimination of WIMPs from sources of background recoils. In the summer of the northern hemisphere, the velocity of the Earth around the Sun aligns with the (current) velocity of the Sun around the galactic center, potentially amplifying and spectrally hardening the recoil energy distribution[83]. However, direct detection experiments are sensitive to WIMP masses much larger than that of protons, i.e., above ~ 10 GeV/c^2. The current upper limit for the spin-independent scattering cross section is 7.7×10^{-47} cm^3 at 90% confidence level[19].

Third, WIMPs can interact with each other and produce Standard Model states that then produce showers of lower energy particles. An important class of such final states is photons, which propagate to the observer without interacting with the intervening magnetic field, unlike those of charged products such as electrons, positrons, protons, and anti-protons. The flux, i.e., number of photons per detector area, solid angle, energy, and time, expected from these WIMP annihilations is proportional to the line-of-sight integral of the WIMP number density squared,

$$\phi_{chi} = \frac{1}{2} \frac{\langle \sigma v \rangle}{4\pi} \frac{dN_\gamma}{dE} \int_{los} n_\chi^2 dl = \frac{\langle \sigma v \rangle}{8\pi m_\chi^2} \frac{dN_\gamma}{dE} \int_{los} \rho_\chi^2 dl, \quad (1.5)$$

since annihilation is a two-body process. Here, ρ_χ is the energy density of WIMPs, which is independently measured to be ~ 0.3 GeV cm^{-3}, $\langle \sigma v \rangle$ is the thermally averaged annihilation cross section
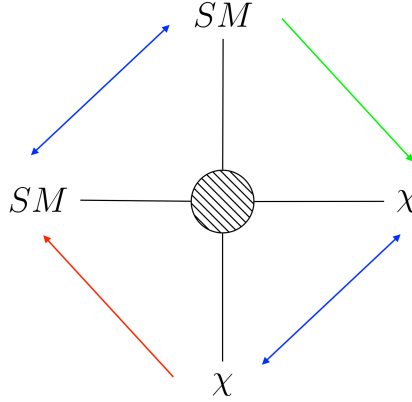
Figure 1.5: Three ways to detect WIMPs. The reaction proceeds to the lower right in collider searches and in all directions in the early Universe. Direct detection is any scattering along the line connecting the lower left and upper right. Indirect detection is the reaction that proceeds to the upper left. SM stands for a Standard Model particle and can be any of $W^\pm, Z, \gamma, g, H, q^\pm, l^\pm$.

and dN_γ/dE is the photon spectrum (i.e., number of photons per energy) in a single annihilation.

Depending on the annihilation channel, the Standard Model generically predicts the spectrum of these photons to have a peak at some fraction of m_χ. In particular, annihilations into quark anti-quark pairs result in more efficient partitioning of initial energy to photons, generating broader photon spectra with smaller average energies. Annihilations into charged leptons, however, produce more energetic photons in units of m_χ. Other particles such as electrons and positrons co-produced in WIMP annihilations can increase the spectrum at low energies via inverse Compton scattering and Bremsstrahlung radiation due to their interactions with the galactic gas, dust and cosmic rays, the latter being a significant factor for indirect detection in the direction of the galactic plane and the galactic center.

1.4 The GeV excess

The gamma-ray sky is a rich source of information on the non-thermal processes in the Universe. Gamma-rays allow characterization of energetic sources such as pulsars, active galactic nuclei (AGN), supernovae remnants, as well as galactic and extragalactic diffuse high energy interactions.

Fermi Large Area Telescope (LAT)[21] is a spaceborn gamma-ray detector that has performed a full-sky survey. Thanks to its low instrumental backgrounds and high angular resolution, it has detected over three thousand gamma-ray sources, the galactic bubbles[213] and characterized the diffuse emission of the Milky Way and the extragalactic emission. The 3FGL[4] is the full-sky catalog of gamma-ray sources detected by the Fermi-LAT.

The observed gamma-ray sky can be modeled as the sum of diffuse emission components, point and extended sources. The dominant contribution to the diffuse model[5] is the gamma-rays produced by the scattering of high energy protons with interstellar gas and dust, which produce neutral pions, π_0, that quickly decay into a pair of high energy photons. Another diffuse component is the inverse Compton scattering of interstellar optical and infrared radiation field by the high energy cosmic ray electrons. Lastly, when the high energy electrons scatter on the gas and dust, they produce stopping radiation (Bremsstrahlung) in gamma-rays. In addition, the gamma-ray sky has a significant isotropic emission component, which is mainly the total redshifted gamma-ray emission from distant galaxies.

Using the known sources of diffuse emission and point sources in the 3FGL, the Poisson regression of the Fermi-LAT data towards the inner galaxy reveals an anomalous emission, i.e., the GeV
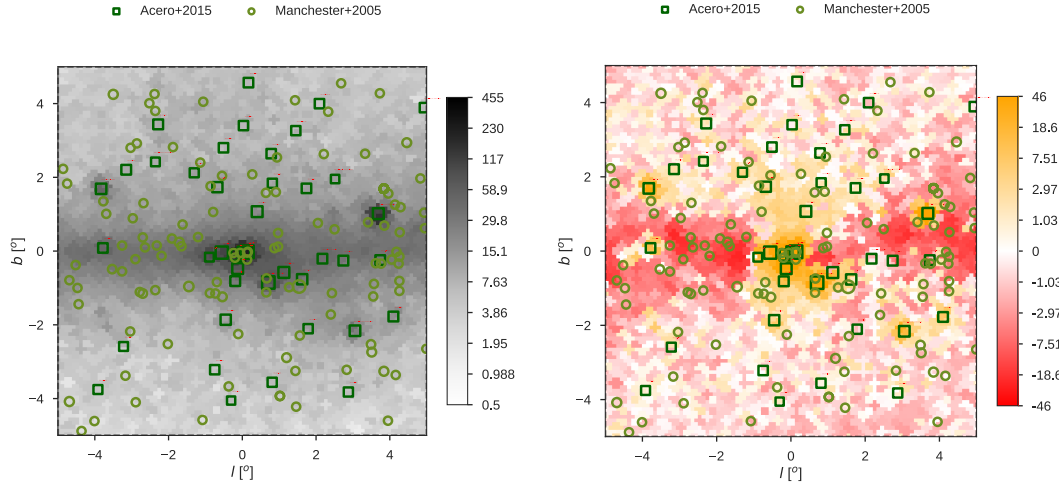
Figure 1.6: Left: gamma-ray emission inside a square of side $10°$ around the galactic center in the 1 - 3 GeV energy bin. Right: residual gamma-ray emission after the diffuse and isotropic emission components have been subtracted. The markers show the Australia Telescope National Facility (ATNF) pulsar catalog (light green) and the 3FGL gamma-ray source catalog (dark green).

excess. The GeV excess was initially discovered by Lisa Goodenough and Dan Hooper[111] and since then, has been scrutinized in detail[125,126,2,112,127,75,49,12]. The Fermi-LAT data and the anomalous residual emission are shown in Figure 1.6.

The inner galaxy is also one of the regions on the sky from which gamma-rays potentially produced by WIMP annihilations are expected to come from. Therefore, the potential consistency of the GeV excess and the photons from WIMP annihilations can generate additional evidence for the dark matter hypothesis as well as yielding an observational probe of its potential particle interactions. In particular, the best-fit spectrum of an emission template as in Equation 1.5 is given in the left panel of Figure 1.7, which peaks at $\sim$ 2 GeV and is consistent with that expected from WIMP annihilations. However, it is important to note that this spectrum is fitted for a single realization of

the diffuse background. When uncertainties in the background emission are accounted, the uncertainties in the spectrum of the GeV excess become larger[48], allowing a range of hypotheses regarding the origin of the GeV excess to be consistent with the marginal spectrum.

In this section, in order to characterize the spatial extent of the GeV excess and test its consistency with the morphology expected from WIMP annihilations, another Poisson regression of the Fermi-LAT data is performed.

The Fermi-LAT data (i.e., Reprocessed Pass 7, ultraclean class, front converting events) in the 40° by 40° square centered at the galactic center, taken between August 4, 2008 and December 5, 2013 is binned spatially in `HealPix` with side resolution 256 and spectrally in logarithmically spaced energy bins. Standard data quality cuts are applied to the data such as taking events with zenith angle less than 100°, instrumental rocking angle less than 52°, `DATA QUAL`, and `LAT CONFIG` tags equal to 1.

The regressors are taken to be the Fermi-LAT diffuse model (i.e., `p6v11`) evaluated at each energy, a free isotropic emission component in each energy bin, and 8 rotationally symmetric ring templates, each 1° wide, centered at the galactic center. The radii of the rings vary from 2° to 10° and are smoothed with the Fermi Point Spread Function (PSF) similar to the diffuse model, using the Fermi Science Tools. Data around the known point sources are masked.

A design matrix, A, that contains these templates as columns, is constructed such that

$$Y = A\beta + \epsilon \tag{1.6}$$

where the number of rows and columns of A are the number of pixels times the number of en-

ergy bins used in the fit, and the number of regressors, β is the vector of template coefficients, Y is the observed gamma-ray counts, and ϵ is the residual, which is assumed to be independent and identically distributed. Because collection of gamma-rays is a Poisson process, ϵ should be Poisson distributed. However, in order to obtain an approximate solution, it is initially assumed that ϵ is Gaussian distributed with standard deviation equal to $\sqrt{Y}$, and the solution for β that minimizes $\epsilon^T \epsilon$,

$$\beta_0 = (A^T A)^{-1} A^T Y, \tag{1.7}$$

is obtained. Then, starting from β_0, a gradient-descent optimizer is used to find the maximum likelihood solution for β.

The best-fit ring coefficients and their uncertainties at 68% confidence level, are shown in the right panel of Figure 1.7. This implies that the GeV excess is detected up to $10°$ from the galactic center, which corresponds to a projected distance of ~ 1.5 kpc. It is also inferred that, up to $8°$ from the galactic center, the morphology of the GeV excess is consistent with the morphology of a generalized NFW profile with index 1.2, squared and projected onto the sky.

The resulting best-fit annihilation cross section is 2.25×10^{-26} cm^{3}s^{-1} for a 43 GeV $b\bar{b}$. It is also found that, although the GeV excess can be fitted with annihilations into $\tau\bar{\tau}$ with lower (i.e., ~ 20 GeV) masses, annihilations into $b\bar{b}$ yield better consistency with the GeV excess. In addition, the inferred annihilation cross sections would be two times as much, if the WIMP is Dirac (i.e., not self-conjugate) instead of Majarona. It is also important to note that, when the regression is repeated
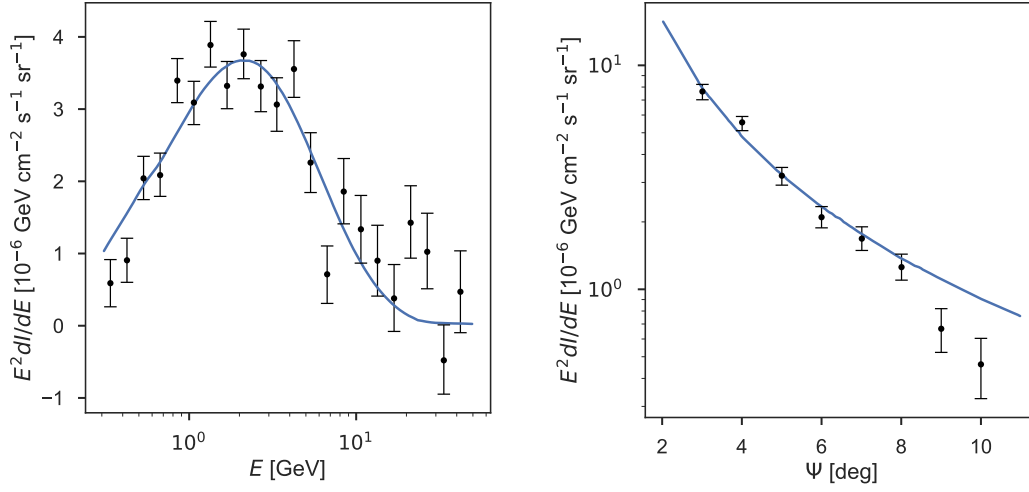
Figure 1.7: The data (black) and the best-fit (blue) spectrum (left) and morphology (right) of the GeV excess [75].

with the full-sky data, the best-fit annihilation cross sections are reduced by a factor of ~ 2.

The inferred annihilation cross section is consistent with the annihilation cross section needed for the thermal freeze-out of WIMPs. Furthermore, if WIMPs are annihilating with this cross section, the annihilation has to be s-wave (i.e., independent of the WIMP velocity), and in order to be consistent with the constraints from direct detection experiments, the elastic scattering cross section should be suppressed by the WIMP momentum. Nevertheless, the best-fit WIMP annihilation cross section is likely to be biased, since the systematic uncertainties of the excess is large below 1 GeV [49].

The finding that the GeV excess extends up to $10°$ from the galactic center, is inconsistent with the hypothesis that young pulsars may be producing the GeV excess, because young pulsars are expected to become faint by the time they reach this distance given a typical kick velocity of ~ 1000

24

km s^{-1}. However, it is plausible that the GeV excess is generated by millisecond pulsars, which can remain bright over billions of years.

It is important to note, however, that despite the consistency between the GeV excess and the WIMP annihilation hypothesis, other predictions of the WIMP annihilation hypothesis are in tension with other data sets. For example, if WIMP annihilations are producing the GeV excess, WIMPs should also be annihilating in dwarf galaxies. However, no such emission has been observed, which has ruled out thermally annihilating WIMPs below 50 GeV at 95% confidence level[9,13,11]. Furthermore, the absence of observations of new particles at the LHC has weakened the theoretical priors of $\mathcal{O}(100)$ GeV WIMPs.

2

Probabilistic cataloging

Imagine taking a long exposure photograph. Initially, because no photon has arrived at the CCD yet, *any* model about what is behind the lens, will be *equally likely* given the observed data. However, that there is no photon arriving at the CCD, would be the most *credible* model from a reductionist point of view, given that it is the simplest model and the closest to being tested. Once some photons have been collected by the CCD, imagine that the surviving model is that the camera is pointed at

a very distant ferry on the ocean. Although the observation that the hull of the ferry is partially occluded by the ocean, can be propagated as far as to infer that the Earth is round, there will still be infinitely many models not discriminated by the available data. For instance, the noise in the photograph may not allow the individual detection of most people on board. Yet, the model that there is one, ten or hundred people on board, may not be equally credible, given subtle discriminating features in the observed image. Moreover, given the noisy image of the distant ferry, any inference, e.g., whether there are humans or bears on board the ferry, require taking into account all possible multiplicities, heights, and colors of humans and bears. Furthermore, even if the CCD had collected photons for a long time, models regarding even finer scale may remain degenerate, i.e., individual eyelashes of a given person may not be detected, although the size distribution of eyelashes may again be constrained.

The essence of the scientific method is to construct a model (i.e., hypothesis), M, with certain a priori degree of belief, i.e., probability [*], distribution on its parameter space, $P(\Omega_M)$. The probability distribution over Ω_M can be updated a posteriori to yield another probability distribution $P(\Omega_M|D)$ [26] based on the consistency of the model with the observed data, D,

$$P(\Omega_M|D) = \frac{P(D|\Omega_M, M)P(\Omega_M, M)}{P(D|M)}, \tag{2.1}$$

---

[*]This thesis is written from a Bayesian standpoint. It treats probability as degree of belief in a proposition, as opposed to the frequentist definition of frequency of observation of experimental outcome. Hence, it allows assignment of probability to hypotheses unlike the latter, which can only assign probability to observed data conditional on hypotheses.

where $P(D|\Omega_M)$ is the probability of observing D given the model M, i.e., the likelihood. Since the data arrives incrementally over time, this update can happen indefinitely, adopting the posterior of previous inferences as the prior for new inferences.

In the rest of this thesis, the probability distributions over the parameter space of the model, M, i.e., $P(\Omega_M)$ and $P(\Omega_M|D)$, will be referred to as the *prior* and *posterior*, respectively. Furthermore, the union set of a group of models, $M_0, M_1, ..., M_{N_{max}}$, will be referred to as the *metamodel*, $\hat{M}$, whose parameter space becomes $\Omega_{\hat{M}} = \Omega_{M_0} \times \Omega_{M_2} \times ... \times \Omega_{M_{N_{max}}}$.

Models can contain a number of identical parameter subspaces. In this thesis, these will be referred to *elements*. An element is a group of parameters that represent a meaningful entity in the model only when considered together. Hence, the number of elements in a given model, N, is the *model indicator* and also a discrete parameter of the metamodel, $\hat{M}$, such that the Nth member of the metamodel contains N elements. Furthermore, the qth parameter of the nth element in the Nth member of the metamodel will be denoted with $\xi_{N,n,q}$. Note that elements and their parameters are model constructs and do not belong to the data space.

The parameter space spanned by element parameters will also be referred to as the *catalog space*,

$$\Omega_C = \bigcup_{N=1}^{N_{max}} \Omega_{C_N} = \bigcup_{N=1}^{N_{max}} \prod_{n=0}^{N-1} \prod_{q=0}^{N_{ep}-1} \xi_{N,n,q}, \tag{2.2}$$

where N_{ep} is the number of parameters per element. Because the metamodel can also contain parameters other than those of elements, the parameter space of the metamodel, $\Omega_{\hat{M}}$, can be written

as the union of the product spaces of the catalog subspaces with the space of a *common parameter* vector $\vec{\rho}$,

$$\Omega_M = \bigcup_{N=0}^{N_{max}} \vec{\rho} \times \Omega_{C_N}.$$

<div align="right">(2.3)</div>

Equivalently, the catalog space is the metamodel marginalized over the common parameters, $\vec{\rho}$.

Cataloging is the inference of elements and their parameters. A traditional approach [†] to cataloging is to determine whether adopting an alternative model with an additional element increases the maximum likelihood as compared to the null model without the additional element. Iteratively performing this hypothesis test over the catalog *subspace* while potentially optimizing other parameters, e.g., $\vec{\rho}$, one can determine the likelihood improvement for the additional element. This can then be used to define elements as detected when the data is at least a 5σ deviation from the prediction of the null model.

If the model elements only have Gaussian covariances, traditional cataloging can capture uncertainties of element parameters with the covariance matrix. However, when elements have non-Gaussian covariances due to the non-Gaussianity of the likelihood (e.g., Section 3) or nonlinearities in forward-modeling (e.g., Section 4), covariances between element parameters cannot be fully captured by the covariance matrix. The maximum likelihood catalog also either blends elements together or infers spurious elements, failing to propagate uncertainties across models with different

---

[†]The term *traditional* cataloging is used to refer to any approach that does not use at least one of the four features of probabilistic cataloging listed later in this section.
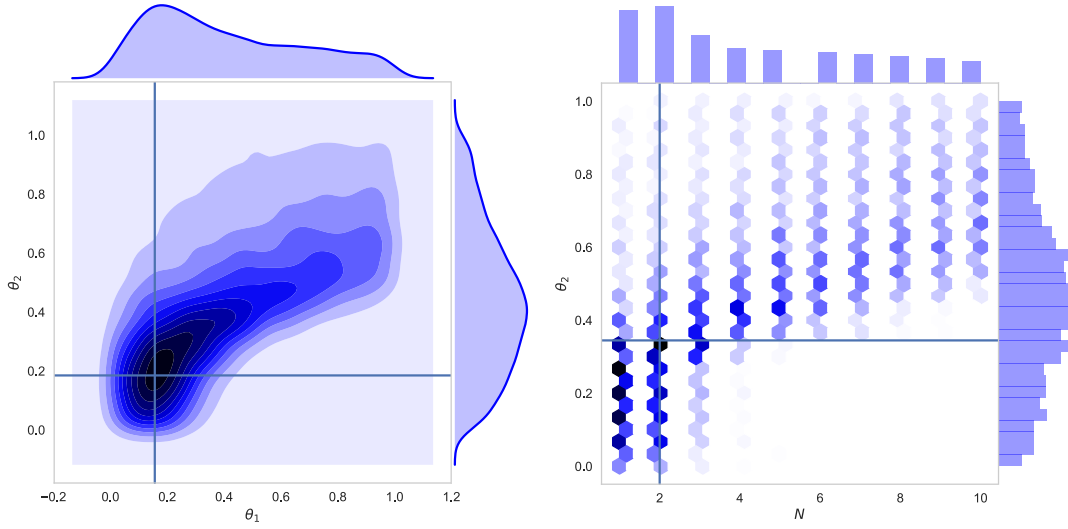
Figure 2.1: Left: Covariances between two continuous parameters. Right: Covariance between the model indicator and a continuous parameter.

complexity. Furthermore, encapsulation of the parameter covariances with the covariances matrix, also fails when the likelihood is multimodal. These issues with the propagation of within and across model covariances can be overcome in a Bayesian framework, by representing one's state of knowledge with an *ensemble* of configurations with a variable number of elements consistent with the data.

Figure 2.1 shows two examples of non-Gaussian covariances. The left panel contains the joint probability distribution of two parameters, θ_1 and θ_2. Even though the maximum likelihood is at $(\theta_1, \theta_2) = (0.2, 0.2)$ and the uncertainty estimate at this point is $(\sigma_{\theta_1}, \sigma_{\theta_2}) = (0.1, 0.1)$, the global topology of the joint probability distribution is significantly different from this estimate. Therefore, the marginal probability distributions do not agree with the maximum likelihood estimate. The right panel shows a similar joint probability distribution, but over a continuous parame-

ter θ_2 and a model indicator, N.

Traditional cataloging can dimensionally reduce large amounts of observations to relatively compact lists of elements, precluding false positives with the use of a hard significance threshold that discards subthreshold information. Nevertheless, traditional catalogs are still models that describe one's state of knowledge consistent with the given data up to statistical and systematic errors. As exemplified above, despite collection of large amounts of data, members of the metamodel, $\hat{M}$, can be approximately equally consistent with the data. Therefore, inferences should take such within and across-model degeneracies into account by first learning the joint probability distribution of $\Omega_{\hat{M}}$, and then marginalizing this probability distribution. This can be achieved by collecting an ensemble of metamodel realizations.

A fundamental problem in inference is that of mismodeling, i.e, the data not being a draw from the generative model used to fit it. This is because models of observations in the Universe are never perfect descriptions of the data due to unmeasurable degrees of freedom. Therefore, the derived statistical uncertainties underestimate the total uncertainties due to the presence of systematic errors. Inferences must marginalize over the dominant sources of systematic uncertainties to ensure that the remaining systematic uncertainties is below the inferred statistical uncertainty. Bayesian inference also allows robust marginalization over such nuisance parameters.

In what follows, an element that improves the goodness-of-fit below and above the 5σ significance threshold, will be referred to as *subthreshold* and *superthreshold*, respectively. Traditional cataloging relies on retaining superthreshold elements as "detected" and discarding subthreshold elements from the model. By requiring a high threshold, a low false discovery rate is ensured at the

expense of also lowering the completeness of the inferred catalog. As a result, the fact that some features in the data are not modeled due to the missing subthreshold elements, causes information loss as well as biasing the parameters of superthreshold elements and other parameters in the model. A high significance threshold especially degrades inference of the population characteristics of subthreshold and marginally superthreshold elements.

In this thesis, a *transdimensional*, *threshold-free*, *hierarchical*, and *Bayesian* inference framework called *probabilistic cataloging*, is implemented that performs covariant modeling and robust uncertainty propagation. The framework is based on taking fair samples from the posterior of a metamodel, consistent with the given data. Each of these four ingredients independently adds functionality to the framework. The thesis builds on previous ideas in [114,191,124,41,42].

Probabilistic cataloging does not distinguish subthreshold and superthreshold elements and does not place any threshold on the significance of elements. Subthreshold elements allow one to recover more information from the data. When they enter posterior samples more frequently than what one would expect from the null model, this propagates information from subtle features in the data to the posterior of their parameters, even though the degree of belief in their individual existence can be below detection threshold. Their existence in the metamodel also reduces the *degradation* (i.e., bias and variance) when inferring the parameters of the superthreshold elements.

The transdimensionality of the inference allows across-model covariances to be propagated. The posterior of the *count function* $S(\xi_q > \xi_{q,*})$, i.e., the number of elements with its qth parameter above some $\xi_{q,*}$ in a given *sample* from the posterior, encapsulates the population characteristics of the qth parameter. It accounts for how the uncertainties in the parameters as well as the existence of

32

elements reflect to the population characteristics of the q element. The posterior of $S(\xi_q > \xi_{q,*})$, $P(S(\xi_q > \xi_{q,*})|D)$, is marginalized over all within and across-model covariances.

In astrophysics, applications of probabilistic cataloging is ubiquitous. Fundamentally, the data is either the number of photons collected in spatial, spectral, and temporal bins with some measurement noise, or secondary quantities derived from such observations. Some examples of data sets that require covariant modeling are crowded images (spectrograms or light curves) across the electromagnetic spectrum (e.g., optical, X-rays, and gamma-rays), where there are more than one source (line or feature) per detector resolution; lensed images, where the lens model contains covariant mass clumps; kinematic data sets, where stellar streams form covariant clusters on the space of integrals of motion; CMB power spectra, where covariant features may exist in the primordial power spectrum. The common feature of all these data sets are that multiple models of different complexity can be approximately equally consistent with data, resulting in an uncertainty on the number of their elements.

2.1   PRIOR ON THE METAMODEL

Given a metamodel $\hat{M}$, the prior on its Nth member can be written as

$$P(\Omega_{M_N}) = P(N, \mu, \vec{\rho}, \{\{\xi_{N,n,q}\}_{n=0,1,\ldots,N-1}, \beta_q\}_{q=0,1,\ldots,N_{ep}-1}) \tag{2.4}$$

where β_q is the *hyperparameter* that parametrizes the prior on the qth parameter of all elements,

$$P(\xi_{N,n,q}, \beta_q) = P(\xi_{N,n,q}|\beta_q)P(\beta_q), \qquad (2.5)$$

where $q = 0, 1, ..., N_{ep} - 1$.

An additional hyperparameter is the one that parametrizes the prior on N, assuming that N is Poisson realization of underlying mean number of elements μ,

$$P(N|\mu) = \frac{\mu^N}{N!}e^{-\mu}. \qquad (2.6)$$

The hyperparameter μ is taken to be log-uniform distributed, i.e., scale-free, such that

$$P(\mu) = \frac{1}{\ln\mu_{max} - \ln\mu_{min}}\frac{1}{\mu} \qquad (2.7)$$

for $\mu_{min} = 0.1 < \mu < \mu_{max} = 10N_{max}$ and vanishes otherwise.

Changing a hyperparameter does not change the likelihood, but only reparametrizes the prior. The prior on the hyperparameter, $P(\beta_q)$, will be referred to as the *hyperprior*. Probabilistic cataloging allows hyperparameters to be assigned a hyperprior other than a delta-function. When the distribution of a given element parameter is apriori believed to belong to a certain a family, e.g., power-law distributed, but the exact form of the prior, e.g., log-slope, is not known, hierarchical modeling allows reduced prior informativeness and the forward propagation of prior uncertainties.

Throughout this thesis, in order to keep the notation simple, both the probability *densities* of

continuous, e.g., μ, and the probability *distribution* of the discrete parameter N are denoted simi-larly with $P(\mu)$ and $P(N)$, with the implication that the former is a probability density whereas the latter is a probability distribution and they are both referred to as *probability distributions*.

Element parameters are assumed to be independent and identically distributed draws from an underlying population and no correlations are believed to exist a priori. Therefore, uniform priors are placed on the n-point correlations of element parameters, allowing the posterior to reveal correlations, if any.

Taking into account the hierarchy and independence of parameters, the joint prior on the meta-model becomes

$$P(\Omega_{M_N}) = P(N|\mu)P(\vec{\rho}) \prod_q P(\beta_q) \prod_n P(\xi_{N,n,q}|\beta_q). \tag{2.8}$$

2.1.1 PARSIMONY

A model can explain an observed data set by remaining consistent with it. However, given an unconstrained number of degrees of freedom, any data can be explained with arbitrary complexity. Nevertheless, the resulting model would fail in making predictions for future data because of overfitting the randomness in the current data. Therefore, the figure of merit that distinguishes models is their explaining power weighted against their predictive power. A fundamental principle that encodes this principle is the Occam's razor, which maintains that models should make as few assumptions as

possible while explaining phenomena.

The degree of belief in a member of the metamodel, M_N, is given by the integral of the likelihood, $P(D|N, \theta_N)$, over the parameter space of the model, weighted by the prior degree of belief, $P(N, \theta_N)$,

$$P(D|M_N) = \int P(D|N, \theta_N)P(N, \theta_N)d\theta_N. \tag{2.9}$$

where θ_N is an arbitrary parameter of the Nth member of the metamodel. This fully marginalized posterior will be referred to as the Bayesian evidence. Assuming a flat prior of width Σ_{θ_N} and a flat posterior of width σ_{θ_N} for θ_N, the Bayesian evidence becomes $P(D|N, \theta_{N,\text{max}})(\sigma_{\theta_N}/\Sigma_{\theta_N})^{d_N}$, where d_N is the dimensionality of θ_N and $\theta_{N,\text{max}}$ is the maximum likelihood estimate of θ_N. Therefore, the Bayesian evidence contains the ratio of the posterior volume to the prior volume and disfavors models that waste prior parameter space, i.e., those that marginally increase the goodness-of-fit at the expense of making other predictions inconsistent with the observed data. In probabilistic cataloging, Bayesian evidence is the primary source of parsimony.

Figure 2.2 illustrates the Bayesian evidence for three different models M_1, M_2, and M_3. The horizontal axis enumerates the set of all possible data that could be observed. Probability of observing some data, is nonzero over different intervals for the three models. M_1 fails in assigning nonzero probability to the observed data, D. Even though M_3 predicts the observed data with some probability, it spreads the probability over a larger region in the data space. The model that assigns the highest probability of observing the data, D, is the model M_2. Therefore, M_2 should be chosen
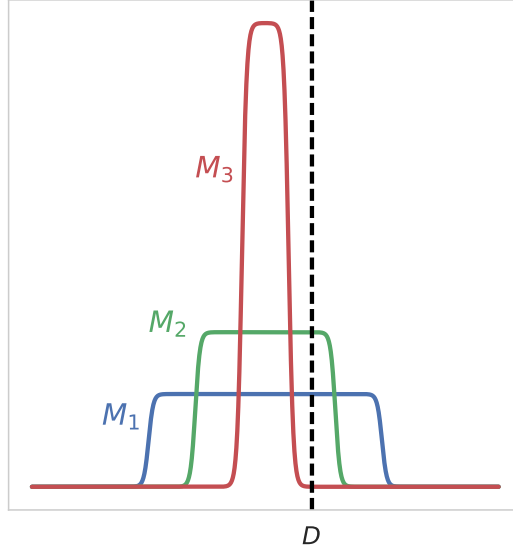
Figure 2.2: The Bayesian evidences of three models M_1, M_2, and M_3, as a function of data. M_2 has the highest evidence for the observed data, D.

model based on the parsimony principle.

In what follows, the element parameter whose correlation with its significance is highest, will be referred to as the *amplitude* parameter denoted by ξ_a. In probabilistic cataloging, in order to make the prior distribution of element amplitudes scale-free, the prior on the amplitude of the nth element is taken as a power-law between some $\xi_{a,min}$ and $\xi_{a,max}$ with the index $-\beta_a$.

$$P(\xi_{a,n}|\beta_a) = \begin{cases} \dfrac{1 - \beta_a}{\xi_{a,max}^{1-\beta_a} - \xi_{a,min}^{1-\beta_a}} \xi_{a,n}^{-\beta_a} \text{ for } \xi_{a,min} < \xi_{a,n} < \xi_{a,max} \\[2ex] 0 \text{ otherwise} \end{cases} . \qquad (2.10)$$

Furthermore, a uniform prior is placed on the angle described by this log-slope,

$$P(\beta_a) = \begin{cases} \dfrac{1}{\tan^{-1}(\beta_{a,max}) - \tan^{-1}(\beta_{a,min})} \dfrac{1}{1 + \beta_a^2} & \text{for } \beta_{a,min} < \beta_a < \beta_{a,max} \\ 0 \text{ otherwise} \end{cases} \quad . \quad (2.11)$$

As a result, most of the prior volume is assigned to the low end of the amplitude distribution. When the minimum of the prior on the amplitude is high, most of the prior volume will be inconsistent with data, yielding a low Bayesian evidence for bad models. However, as the minimum of the amplitude prior is lowered, more of the prior volume becomes consistent with the data. In the limit of taking this minimum to zero, addition of a random element drawn from the prior into the model, does not change the likelihood. This results the integral in Equation 2.9 to evaluate to unity, since Bayesian evidence does not penalize model complexity that does not affect the likelihood.

The likelihood as a function of a typical amplitude parameter is given in Figure 2.3. It shows that the amplitude parameter has a Gaussian likelihood around the true value, when the parameter value is above the detection threshold. However, at smaller amplitudes, the likelihood loses sensitivity to the parameter and levels off at the likelihood of the null (i.e., zero-element) model.

Probabilistic cataloging places a scale-free prior on the Poisson *mean* of the number of elements, which also causes smaller number of elements to be favored. This is the second (weaker) source of parsimony, but is independent of the response of the likelihood to the new dimensions.

Adding new dimensions to a model along which the likelihood is flat, is not useful for learning information from data because the posterior reduces to the prior. However, some information can
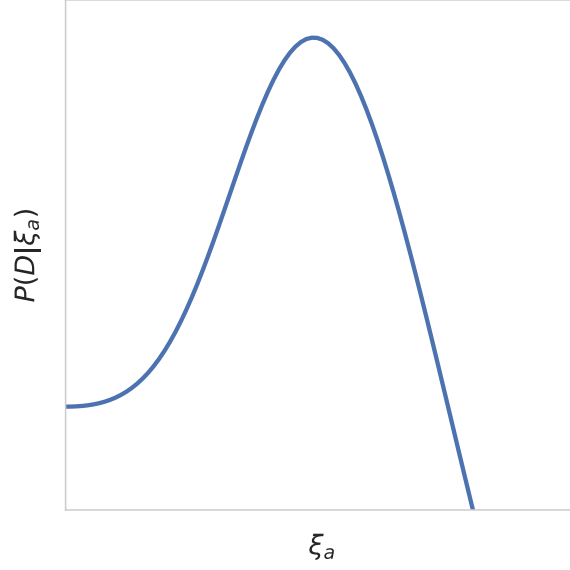
Figure 2.3: The likelihood as a function of typical amplitude parameter.

be inferred from the data when the likelihood is only *approximately* flat along the new dimension, although the posterior can be highly informed by the prior. Therefore, depending on the intended meaning of the posterior, an additional regularization prior may be needed. In particular, if the posterior is to represent the *complete* catalog space allowed by the data, then no additional regularization is needed. However, if the posterior is to represent a *compact* description of the catalog space consistent with the data, then model complexity should be penalized further.

Because of the many uncontrollable degrees of freedom, observations made in the physical Universe are always stochastic. When testing an alternative model, the randomness (i.e., uncertainty) of all observations have to taken into account in the null model. As a result, there should always be a nonzero *expected* discrepancy between the data and the fitting model. Furthermore, a model with

a larger number of elements can fit the data at least as well as another with fewer elements, if the priors on the element parameters are chosen appropriately. Just as a too-low goodness-of-fit shows that the model is not consistent with the data, a too-high goodness-of-fit indicates that the degrees of freedom of the model causes the noise in the observations to be mismodeled. Not respecting the noise properties of the data can bias a hypothesis test at low signal-to-noise since many subthreshold elements can overfit the data by increasing the goodness-of-fit above that expected from the noise in the null model. In addition, it can also cause the amplitudes of superthreshold elements to be underestimated.

When the posterior represents the *complete* catalog space allowed by the data, the computational complexity of inference grows exponentially with the addition of each element. However, this increase in the computational complexity is not justified when judged on the basis of information gain per computation, since little information is gained while spending a disproportionate amount of computational resources. Therefore, unless the minimum amplitude is high enough to ensure manageable computational complexity, an additional prior becomes needed to regulate model complexity when elements are subthreshold. This prior will be referred to as the *regularization prior.*

An effective way to regulate inferences is cross validation [211], which involves iteratively evaluating consistency of a predicted subset of the data when the model explains the rest of the data. However, in probabilistic cataloging this type of model regularization would introduce excessive computational complexity, since the forward-modeling has to be performed millions of times. Instead, models can also be regulated based on the some norm of the parameter vector, $L_p \equiv (\theta_0^p + \theta_1^p + ... + \theta_{N-1}^p)^{1/p}$ in the form of a negative log-prior term. Although L_0 norm offers

a penalty independent of the parameter scale, L_1 as in LASSO [217] (Least Absolute Shrinkage and Selection Operator), and L_2 as in ridge regression, are more frequently employed in optimization problems since they yield differentiable cost functions. In probabilistic cataloging, however, differentiability is not a requirement and the L_0 norm can be used to select models.

There are several information criteria that perform model selection based on the L_0 norm that are valid at different levels of signal-to-noise and likelihood topologies. For example, assuming that the true model is among the members of the metamodel and that the prior is approximately flat, the Bayesian evidence can be Taylor-expanded around the posterior mode up to second order to obtain an approximation valid at high signal-to-noise (i.e., when number of data points, N_D, is much larger than that of parameters, d_N),

$$P(D|M_N) \approx P(D|N, \theta_{N,*}) P(N, \theta_{N,*}) (2\pi)^{d_N/2} \left| \nabla^2 P(D|N, \theta_N) \right|_{\theta_N = \theta_{N,*}}^{-1/2} N_D^{-d_N/2} \quad (2.12)$$

yielding the Bayesian (Schwarz) Information Criterion (BIC) [198],

$$\ln P(d_N) = -\frac{1}{2} d_N \ln N_D. \quad (2.13)$$

When inference is at low signal-to-noise, this approximation does not hold true and only yields an upper (conservative) limit. If the assumption that the true model is in the metamodel is relaxed, one can then attempt to find a model that best approximates the true model. One way to characterize

the proximity of a probability distribution P to another probability distribution Q is to use the Kullback Leibler (KL) divergence[141], i.e., information gain,

$$D_{KL} = \int d\theta P(\theta) \ln \frac{P(\theta)}{Q(\theta)}. \tag{2.14}$$

When the KL divergence between the approximate fitting model and the true model, is minimized, the bias (i.e., discrepancy between the maximum and expected likelihood) becomes equal to the dimensionality of the model, yielding the Akaike Information Criterion (AIC)[24].

$$\ln P(d_N) = -d_N \tag{2.15}$$

However, both BIC and AIC are only valid for point estimates. In probabilistic cataloging, the regularization prior must instead be imposed *during* the sampling. Therefore, assuming that the residuals of the null and alternative models are χ^2 distributed with $N_D - NN_{ep}$ and $N_D - (N + 1)N_{ep}$ degrees of freedom, respectively, the regularization prior is defined as

$$\ln P(N+1) - \ln P(N) = \langle \frac{1}{2}\chi^2_{N_D-(N+1)N_{ep}} - \frac{1}{2}\chi^2_{N_D-NN_{ep}} \rangle$$
$$= \frac{1}{2}\left(\langle \chi^2_{N_D-(N+1)N_{ep}} \rangle - \langle \chi^2_{N_D-NN_{ep}} \rangle \right) = -\frac{1}{2}N_{ep}. \tag{2.16}$$

Therefore, by construction, this regularization prior removes complexity that results in $\lesssim 1\sigma$ improvements to the goodness-of-fit. As a result, when regularization prior is applied, probabilistic cataloging is only useful for learning about elements that are *individually* more significant than $\sim 1\sigma$.

2.2 SAMPLING

When inferring the posterior of the metamodel, given infinite time and computational resources, the posterior could be evaluated on a regular grid of the metamodel parameter space. However, the required number of likelihood evaluations grows exponentially with the number of dimensions and a brute-force calculation fails even for moderate number of parameters. Therefore, probabilistic cataloging requires taking fair samples from the posterior.

Sampling is performed by constructing a reversible Markov chain, whose stationary distribution is the target (e.g., posterior) distribution. Known as Markov Chain Monte Carlo (MCMC), this random process allows fair samples to be taken from a probability distribution that cannot be directly sampled from. A sufficient but not necessary condition to obtain a reversible Markov chain is to establish *detailed balance*

$$
\int_{\Omega_{M_N}} P(N, \theta'_N | D) Q(\theta_N, \theta'_N) \alpha(\theta_N, \theta'_N) d\theta_N =
$$

$$
\int_{\Omega_{M_N}} P(N, \theta_N | D) Q(\theta'_N, \theta_N) \alpha(\theta'_N, \theta_N) d\theta'_N. \quad (2.17)
$$

where θ_N and θ'_N are the current and proposed states in the parameter space of the Nth member of the metamodel, Ω_{M_N}, $Q(\theta_N, \theta'_N)$ is the one-to-one transition from θ_N to θ'_N and $\alpha(\theta_N, \theta'_N)$ is the probability of accepting the transition. Because both the current and proposed states are in the same model, M_N, the Bayesian evidences on both sides cancel, yielding

$$\int_{\Omega_{M_N}} P(D|N, \theta'_N) P(N, \theta'_N) Q(\theta_N, \theta'_N) \alpha(\theta_N, \theta'_N) d\theta_N =$$

$$\int_{\Omega_{M_N}} P(D|N, \theta_N) P(N, \theta_N) Q(\theta'_N, \theta_N) \alpha(\theta'_N, \theta_N) d\theta'_N. \quad (2.18)$$

Once $Q(\theta_N, \theta'_N)$ is specified, $\alpha(\theta_N, \theta'_N)$ can be uniquely determined,

$$\alpha(\theta_N|\theta'_N) = \min\left(1, \frac{P(D|N, \theta'_N)}{P(D|N, \theta_N)} \times \frac{P(N, \theta'_N)}{P(N, \theta_N)} \times \frac{Q(\theta'|\theta)}{Q(\theta|\theta')}\right). \quad (2.19)$$

When sampling from the posterior of the metamodel parameter space, $\Omega_{\hat{M}}$, the current and proposed states can have different dimensions, e.g., θ_N and $\theta'_{N'}$ with $N \neq N'$. Therefore, the chain can no longer be reversible in the transdimensional case, since the transition loses its one-to-one property. This can be alleviated by drawing random auxiliary parameters, u and u', using the probability distributions $P(u)$ and $P(u')$, to match the dimensions of the current and proposed states such that the transition $(\theta_N, u) \leftrightarrows (\theta'_{N'}, u')$ is a diffeomorphism H,

$$H(\theta_N, u) = (\theta'_{N'}, u') \quad (2.20)$$

$$H^{-1}(\theta'_{N'}, u') = (\theta_N, u). \quad (2.21)$$

44

The dimension matching,

$$\dim(\theta_N) + \dim(u) = \dim(\theta'_{N'}) + \dim(u'), \tag{2.22}$$

where $\dim()$ denotes the dimension operator, conceals the transdimensional nature of the proposal and ensures reversibility. Therefore, $H(\theta_N, u) = (\theta'_{N'}, u')$ replaces the probabilistic transition, $Q(\theta'|\theta)$.

When making a proposal across models, there are infinitely many ways to construct distinct diffeomorphisms. However, only some are useful transitions to explore degeneracies in the posterior of the catalog space. Denoting the probability of proposing a transition from model M_N to M'_N, $P(N \rightarrow N')$, the ratio,

$$\frac{P(N' \rightarrow N)}{P(N \rightarrow N')}, \tag{2.23}$$

is included in the acceptance probability to ensure detailed balance. Lastly, the coordinate transformation H is required to conserve probability such that,

$$d\theta'_{N'} du' = \left| \frac{\partial(\theta'_{N'}, u')}{\partial(\theta_N, u)} \right| d\theta_N du, \tag{2.24}$$

for all θ_N and u.

This yields

$$\alpha(\theta'_{N'}|\theta_N) = \min(1, \alpha_0), \tag{2.25}$$

$$\alpha_0 = \frac{P(N', \theta'_{N'}|D)}{P(N, \theta_N|D)} \underbrace{\frac{P(u')}{P(u)}}_{\alpha_u} \underbrace{\frac{P(N' \to N)}{P(N \to N')}}_{\alpha_c} \underbrace{\left|\frac{\partial(\theta'_{N'}, u')}{\partial(\theta_N, u)}\right|}_{\alpha_j} \tag{2.26}$$

where the common factor of $P(D|\hat{M})$ on both sides, cancel again since the transformation is performed within the same metamodel, giving

$$\alpha_0 = \frac{P(D|N', \theta'_{N'})}{P(D|N, \theta_N)} \frac{P(N', \theta'_{N'})}{P(N, \theta_N)} \times \alpha_u \times \alpha_c \times \alpha_j \tag{2.27}$$

This sampling framework is known as the Reversible Jump MCMC (RJMCMC)[114,232,118], which is a variant of MCMC that allows across-model proposals in a pool of models indexed by their dimensionality. Probabilistic cataloging inherits the RJMCMC framework in implementing transdimensional proposals.

The purpose of sampling is to traverse the parameter space with a frequency proportional to the posterior. Therefore, given a finite number of samples, the sampler visits regions with the highest posterior, which are disjoint across models and may also be disconnected in a given model. In order to sample efficiently, the proposals (i.e., choices of H and $P(u)$) should approximate the posterior and result in forward-model changes that yield comparable likelihoods.

WITHIN-MODEL PROPOSALS   Within-model proposals explore the posterior along a given member of the metamodel by making changes to the parameter vector. Even though within-model proposals do not require the RJMCMC formalism, they can be incorporated into this framework when both u and u' have the same dimension.

Within-model proposals are made in a transformed space, where the prior is uniform. The transformation is performed using the cumulative distribution function (CDF) of the modified parameter, θ_N, which is either an element or common parameter. This yields a transformed parameter, $\tilde{\theta_N}$, which, by construction, has a uniform prior between 0 and 1. The within-model proposals are made such that

$$(\tilde{\theta_N}, u) \rightarrow (\tilde{\theta}'_N, u') \tag{2.28}$$

$$\tilde{\theta}'_N = \tilde{\theta}_N + u \tag{2.29}$$

$$u' = -u. \tag{2.30}$$

As a result, the prior ratio in Equation 2.27 is set to unity. During forward-modeling with $\tilde{\theta}'_N$, the inverse CDF is used to calculate θ'_N, which is then used to determine the forward-model.

When the posterior is multimodal, barriers between high-posterior regions can significantly increase the number of proposals required for convergence. In order to increase the sampling efficiency, the auxiliary parameter u is drawn from a heavy-tailed Gaussian distribution with mean zero. Furthermore within-model proposals are made symmetric as in Metropolis proposals, which sets α_u

47

to unity. Within-model proposals are also self-antagonist such that both forward and reverse transitions can be achieved using the same type of proposal. Therefore, since there are as many parameter proposals possible as the number of parameters, which is invariant during a within-model proposal, α_c in Equation 2.27, becomes unity. Similarly, the Jacobian, i.e., α_j in Equation 2.27, evaluates to unity for all within-model proposals.

BIRTH AND DEATH PROPOSALS   Although sampling from a given Ω_{M_N} can be performed using within-model proposals, sampling from $\Omega_{\hat{M}}$ requires across-model proposals that incorporate a change in N. In these across-model proposals, u and u' have different dimensions. These across-model proposals implement the volume factor in the Bayesian evidence into probabilistic cataloging, because the auxiliary parameters are randomly sampled.

Birth and death proposals, where an element is either added to or deleted from the element list, are the elementary across-model proposals that allow the exploration of models of different dimensionality. Because they are the reverse proposals of each other, both have to be present in the set of possible proposal types in order for detailed balance to be respected.

During a birth, the transformation

$$(\theta_N, u) \rightarrow (\theta_N, \xi) = \theta'_{N+1}. \tag{2.31}$$

is proposed so that the auxiliary parameter vector, u, is the same as the parameters of the added ele-

ment, ξ. Conversely, during a death proposal it is required that

$$\theta_{N+1} = (\theta_N, \xi) \to (\theta_N, u') = \theta'_N \qquad (2.32)$$

where the auxiliary parameter vector u' carries the parameters of the element to be killed, ξ.

Because the new element parameters are drawn from the prior, this causes the probability of the auxiliary vector to be same as the prior of ξ, resulting in the cancellation of the prior and the probability distribution of the auxiliary parameters in Equation 2.27, i.e.,

$$\frac{P(\xi)}{P(u)} = \frac{P(u')}{P(\xi)} = 1. \qquad (2.33)$$

During birth and death proposals, the list of elements is treated as ordered. Given N elements in the current state, if the list of elements were treated as an unordered list, the number of possible distinct birth proposals would be 1, whereas the number of possible distinct deaths in the reverse proposal would be $N + 1$. However, there would also be an implicit $N + 1$ fold degeneracy in u, reducing $P(u)$ by the same factor. At the end, the resulting α_c would be unity. An equivalent scheme that does not introduce degeneracies to u can be obtained, however, by treating the element list as ordered. Then, both the forward and reverse proposals during a birth have $N + 1$ possibilities, resulting in α_c being equal to unity. Adoption of an ordered element list makes the bijective nature of the transformation manifest.

Since the auxiliary parameter vector is identical to the added element parameters, the Jacobian,

49

α_j, also becomes unity. Hence, α_0 becomes

$$\alpha_0 = \frac{P(D|N',\theta'_{N'})}{P(D|N,\theta_N)} \exp\left(-\frac{1}{2}N_{ep}\right) \tag{2.34}$$

When the regularization prior is applied, after a birth or death proposal, the proposal is accepted with a 100% probability only if the log-likelihood increases by $0.5N_{ep}$. Hence, the regularization prior effectively shifts the log-likelihood cutoff in Metropolis-Hastings proposals, down by N_{ep}. Figure 2.4 shows the trace (left) and histogram (right) of the log-likelihood changes during accepted birth proposals. The top and bottom rows correspond to inferences with and without the regularization prior, respectively.

Birth and death proposals make sampling from a multimodal posterior more efficient, since they result in non-local (i.e., memoryless) changes to the state vector. For the same reason, they accelerate burn-in of the sampler starting at a fair draw from the prior.

SPLITS AND MERGES   At low signal-to-noise, the dominant transdimensional covariance is whether or not features in the data are more consistent with a single element or the background-only hypothesis. However, as the signal-to-noise ratio increases, transdimensional covariances shift to higher numbers of elements, leading to covariances between a single and high-amplitude element or multiple and lower-amplitude elements.

In principle, birth and death proposals span the catalog space, i.e., given infinite resources they
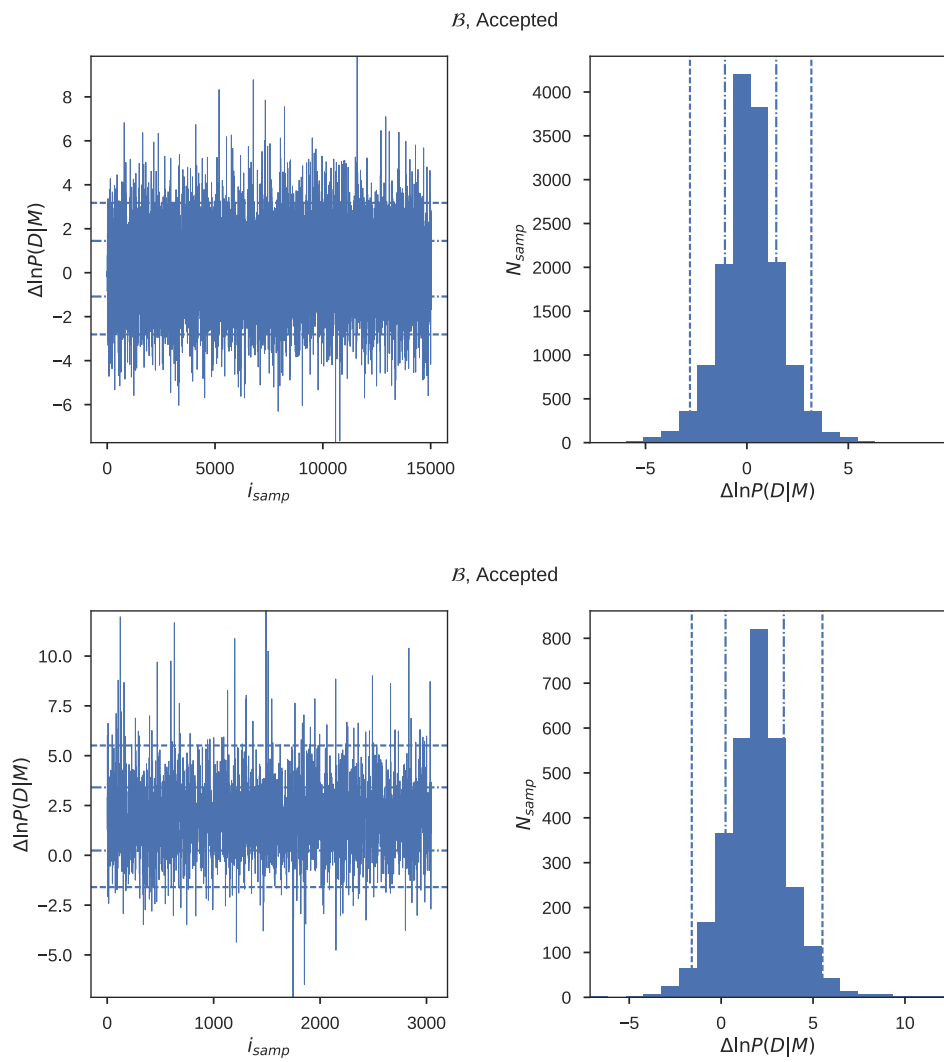
Figure 2.4: Log-likelihood difference during accepted birth proposals.

can be used to probe all transdimensional covariances. However, the exploration of parameter co-variances with only birth and death proposals, can be inefficient for superthreshold elements, because the implementation of birth and death proposals, where the auxiliary parameters are drawn from the prior, is tailored towards exploring subthreshold elements, e.g., between multiplicity of subthreshold elements and the background hypothesis.

Therefore, additional across-model proposals are needed in order to efficiently explore transdimensional covariances at high signal-to-noise. Split and merge proposals, where an element is split into two, and two elements are merged into one, can probe these covariances.

During a split proposal an element with parameters ξ_0 is split into two elements with parameters ξ_1 and ξ_2, such that

$$(\theta_N, u) = (\theta_{N-1}, \xi_0, u) \to (\theta_{N-1}, \xi_1, \xi_2) = \theta'_{N+1} \tag{2.35}$$

where θ_{N-1} is the part of the current state vector that does not change. In what follows, the horizontal position, vertical position, amplitude, and the rest of the parameters of an element will be denoted with ξ_{i,θ_1}, ξ_{i,θ_2}, $\xi_{i,a}$, and $\xi_{i,b}$, respectively, where i is 0, 1, or 2. Hence, the transformation is specified as

$$\xi_{1,\theta_1} = \xi_{0,\theta_1} + u_{\theta_1}(1 - u_a)$$

$$\xi_{2,\theta_1} = \xi_{0,\theta_1} - u_{\theta_1}u_a$$

$$\xi_{1,\theta_2} = \xi_{0,\theta_2} + u_{\theta_2}(1 - u_a)$$

$$\xi_{2,\theta_2} = \xi_{0,\theta_2} - u_{\theta_2}u_a$$

$$\xi_{1,a} = u_a\xi_{0,a}$$

$$\xi_{2,a} = (1 - u_a)\xi_{0,a}$$

$$\xi_{1,b} = \xi_{0,b}$$

$$\xi_{2,b} = u_b$$

(2.36)

where u_{θ_1}, u_{θ_2}, u_a, and u_b are the auxiliary parameters. They represent the horizontal and vertical

splittings, the fraction of the amplitude that is given to the first daughter and the parameters of

the second daughter except its position and amplitude, respectively. Similarly, the transformation

during a merge reads

$$\theta_{N+1} = (\theta_{N-1}, \xi_1, \xi_2) \rightarrow (\theta'_{N-1}, \xi_0, u') = (\theta'_N, u'). \tag{2.37}$$

During a split, the parent is chosen randomly. Therefore, given a state with N elements, there

are N, N and $N + 1$ possibilities, yielding a joint probability of $N^{-2}(N + 1)^{-1}$. During a merge

with $N + 1$ elements initially, the first daughter and the parent are also chosen randomly with a joint
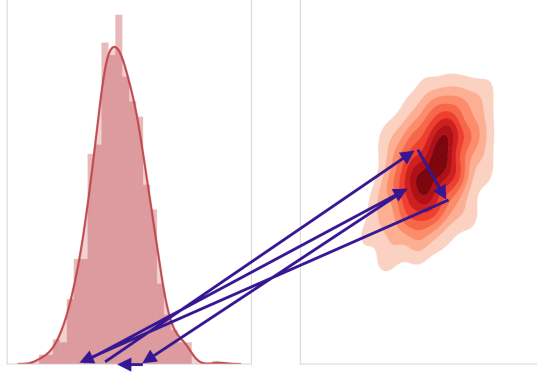
Figure 2.5: Schematic view of split and merge proposals along with the within-model proposals. Arrows indicate the transitions of the MCMC state.

probability of $(N+1)^{-1}N^{-1}$. However, the second daughter is chosen with a probability weighted by a Gaussian in distance from the first daughter and a standard deviation of θ_{sm}. Furthermore, the Jacobian of the split proposal becomes ξ_a. Hence, α_0 for split proposals becomes

$$\alpha_0 = \frac{P(D|N+1, \theta'_{N+1})}{P(D|N, \theta_N)} \times \frac{P(\xi_1)P(\xi_2)}{P(\xi_0)P(u)} \times \exp\left(-\frac{1}{2}N_{ep}\right) \times N \frac{e^{-0.5(\Delta\theta)^2/\theta^2_{sm}}}{\sum_{n=1}^{N-1} e^{-0.5(\Delta\theta)^2/\theta^2_{sm}}} \times \xi_{0,a}$$

(2.38)

The typical transdimensional covariance explored by split and merge proposals is shown in Figure 2.5. Here, a given feature of the data can be fitted with one or two elements. The histogram of the posterior samples from the amplitude of the single element and the amplitudes of the alternative model are shown in the left and right panels, respectively. These histograms are marginalized over all other parameters of the elements such as their positions.

In order for split and merge proposals to be efficient, u_{θ_1} and u_{θ_2} are also drawn from a Gaussian distribution with mean zero and standard deviation θ_{sm}. u_{ξ_a} is drawn from the uniform distribution between 0 and 1, and u_{ξ_b} are drawn from the respective prior distributions. As a result of these choices, the prior ratio, α_u, α_c, and α_j approximately cancel each other, allowing the sampler to visit equally likely configurations in different models.

Figure 2.6 shows $\ln \alpha_j$ for the accepted split proposals. Because α_j is the amplitude, this histogram shifts left or right by a constant depending on the problem. However, its shape is dictated by two competing effects that reduce the acceptance ratio of splits. At high-amplitudes, because elements are more significant, a given spatial splitting causes lower likelihoods. Towards the low-amplitude end, however, daughter elements become more likely to fall below the minimum of the amplitude prior. Therefore, split proposals are more efficient at intermediate amplitudes.

Table 2.1 summarizes the proposal types along with their frequency. In order to ensure that N is not biased, across-model proposals have a fixed probability of proposal, independent of current N.

The state transitions are further illustrated in Figure 2.7, where the sampler transitions between member models using birth, split, death, and merge proposals and explore individual model parameter spaces using within-model proposals.
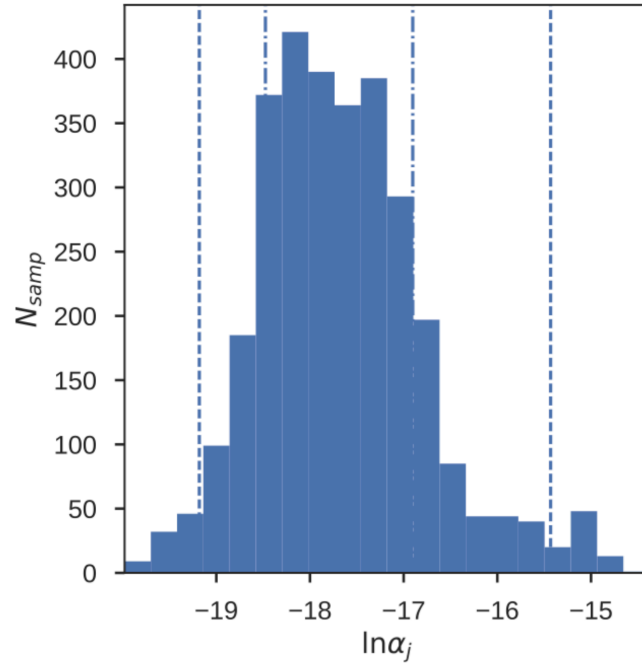
Figure 2.6: Histogram of the logarithm of the Jacobian factor during accepted split proposals.
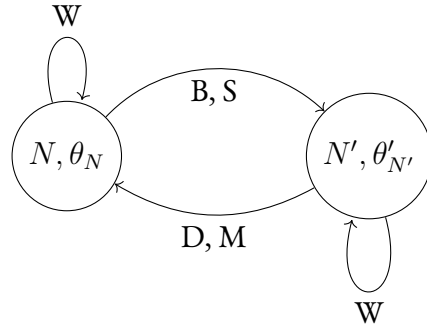
Figure 2.7: State transitions in the Markov chain, showing the birth (B), split (S), death (D), merge (D) and within-model (W) proposals.

Table 2.1: Types of proposals used to explore the catalog space.

Name	Frequency
Within-model	0.6
Birth	0.05
Death	0.05
Split	0.15
Merge	0.15

2.2.1 MODEL SELECTION

The Bayesian evidence for two given models $M_{N'}$ and M_N can be used to find the relative fully marginalized posterior of the models

$$\frac{P(M_{N'}|D)}{P(M_N|D)} = \frac{P(D|M_{N'})}{P(D|M_N)} \frac{P(M_{N'})}{P(M_N)}. \tag{2.39}$$

Because the integral in Equation 2.9 is a high-dimensional integral, its calculation is very noisy. Estimating it with samples from the posterior

$$P(D|M_N) = \frac{N_{samp}}{\sum_a \frac{1}{P(D|N,\theta_{N,a})}} \tag{2.40}$$

or from the prior

$$P(D|M_N) \approx \frac{1}{N_{samp}} \sum_a P(D|N, \theta_{N,a}) \qquad (2.41)$$

where $a = 1, 2, ..., N_{samp}$ is the index of the samples drawn and $\theta_{N,a}$ is the value of the parameter

vector for the ath sample, leads to a large variance due to sampling noise. Stabilizing the estimates

by truncating the samples or sampling from the prior with likelihood thresholds such as in Nested

Sampling[204] may partially alleviate this variance. Nevertheless, the large dimensionality and multi-

modality of the posterior remain intrinsic limitations to estimating the Bayesian evidence.

In probabilistic cataloging, the Bayes factors between member models in Equation 2.39 are in-

stead obtained by the relative frequency of visits to members of the metamodel. Therefore, the

Bayesian evidence of the metamodel, $P(D|\hat{M})$ is not of interest and not calculated. Instead, the

marginalized posterior of the model indicator, $P(N|D)$, directly yields the Bayes factor. Therefore,

detailed balance during reversible jumps implements the parsimony of Bayesian evidence into prob-

abilistic cataloging. This is an advantage of probabilistic cataloging compared to other inference

frameworks that rely on estimating the Bayesian evidences of models and precludes the necessity to

sample from the posterior of individual member models in order to calculate the Bayesian evidences,

i.e., $P(D|M_0), P(D|M_1), ..., P(D|M_N)$.

The parameter vectors of the members of the metamodel contain two distinct classes: the common and element parameters. The common parameters have fixed labels (i.e., interpretations during forward-modeling) that do not change across samples. Therefore, although there is a distinct posterior of the common parameters of each member of the metamodel, $P(N, \vec{\rho}|D)$, a weighted average of these posteriors can yield $P(\hat{M}, \vec{\rho}|D)$.

Unlike the common parameters, the likelihood is invariant to permuting the labels of elements. This is a feature shared by all mixture models[243], where model components are not individually labeled and, hence, can change labels without changing the likelihood. Similarly, in probabilistic cataloging, the labels of elements (and their parameters) can change across samples. As a result, if an infinite number of samples were taken, a weighted average of the posterior of element parameters would become identical. As a result, elements in a given sample from the posterior, cannot be matched to elements in a different posterior sample without breaking the labeling degeneracy.

The degeneracy in the element labeling does not imply any loss of information. Even though the posteriors of a given kind of element parameter, $P(\xi_{n,q}|D)$, are identical for all elements, the information is retained in the posterior of the *count function* of a given element parameter, $P(S(\xi_q > \xi_{q,*})|D)$.

Taking fair samples from the joint posterior of the metamodel is the correct approach for robust uncertainty propagation, which accounts for all within and across-model degeneracies. Breaking of the labeling degeneracy causes some of the information to be lost and introduces a bias into the

posterior due to arbitrary choices made in associating elements across samples. Nevertheless, propagation of uncertainties is a continuous process. Inferences use as prior, the posterior of previous inferences. Therefore, if all inferences had propagated uncertainties by keeping samples from the posterior, the time and space complexity of inference would increase exponentially, making uncertainty propagation impossible after a few subsequent inferences. Therefore, uncertainties should be propagated in the form of posterior samples *only* when, doing so reduces bias and improves information recovery. Even then, the posterior should be dimensionally reduced into a covariance matrix at the earliest possible stage of inference, as soon as uncertainties can be expressed in a different basis without non-Gaussian covariances, i.e., when the covariance matrix can encapsulate the uncertainties.

When samples from the metamodel are to be dimensionally reduced to a covariance matrix of (fixed-label) element parameters, this requires a prescription for breaking the labeling degeneracy. This can be achieved by marginalizing the posterior samples from the catalog space onto the element parameter space and then finding clusters on these projections[180].

2.2.3   Associating a probabilistic catalog with a reference catalog

There are two motivations for associating a probabilistic catalog with a reference (i.e., traditional) catalog. First, using simulated data, the performance of the probabilistic catalog can be studied with respect to the true catalog as in Section 3.1. Second, the associations of the probabilistic catalog with a reference catalog can yield the posterior of parameters that cannot be inferred by using only the data the probabilistic catalog is consistent with, as in Section 3.3.

In this thesis, the elements are associated based on positional proximity. In order to associate a probabilistic catalog with a reference catalog, each posterior sample catalog is associated with the reference catalog by finding the closest sample element inside a circle of maximum radius, θ_{asc}, around every reference element. Then, the posterior of the element parameters that are assigned to reference elements, is calculated.

This approach allows associations between a probabilistic catalog and a reference catalog without breaking the labeling denegeracy of the probabilistic catalog. Furthermore, it yields unbiased uncertainty estimates for the element parameters other then positions (e.g, their amplitudes) at the expense of increasing random associations, i.e., those due to spatial coincidences. However, in order to ensure that these do not bias associations, for all inferences, θ_{asc} is chosen not to be larger than the spatial scale at which elements become covariant.

2.2.4   DECISION LOGIC

Probabilistic cataloging improves inference performance at the expense of increased time complexity. Therefore, a decision logic is required to determine which feature of probabilistic cataloging is justified under what conditions. Figure 2.8 illustrates this decision logic.

2.2.5   IMPLEMENTATION

PCAT is a `Python2.7` implementation of the presented probabilistic cataloging framework and the main product of this thesis. It is a mixture RJMCMC sampler that takes fair samples from the posterior of a metamodel given some data. It is publicly available at https://github.com/
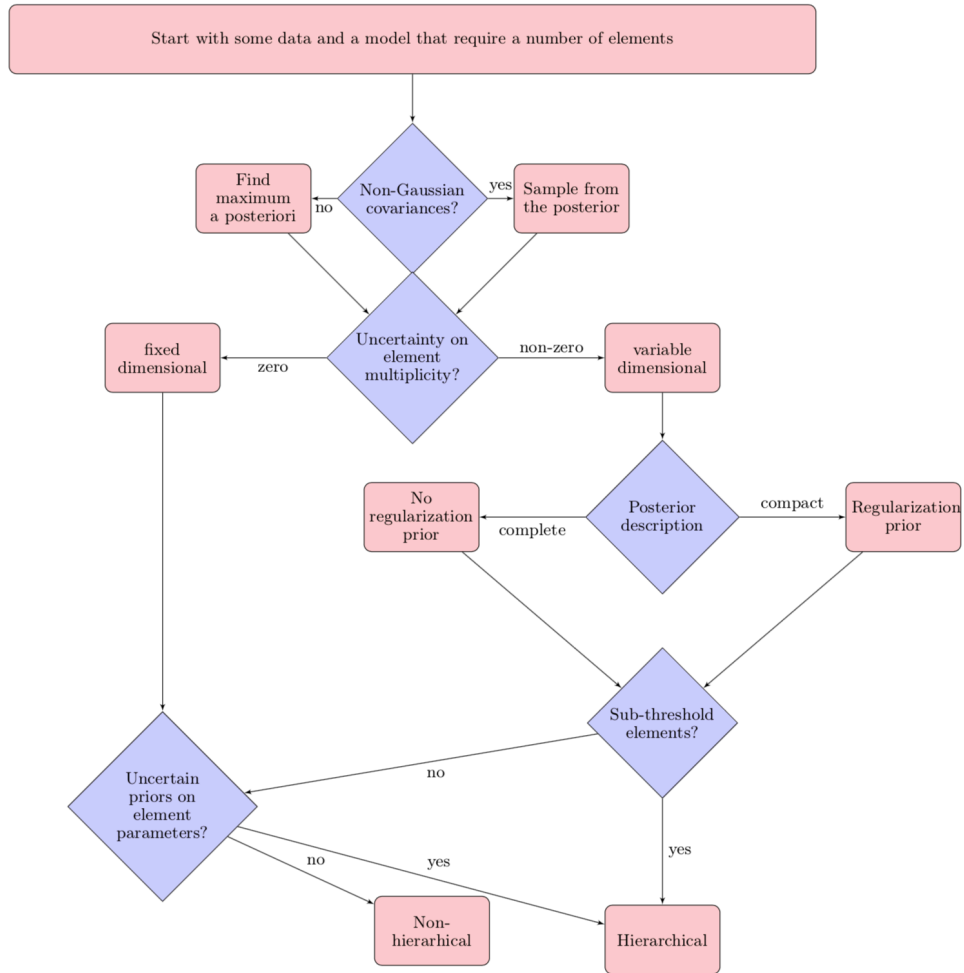
Figure 2.8: Decision logic of how to employ probabilistic cataloging.

`tdaylan/pcat` for the use of the scientific community, with documentation available at `http:`

`//pcat.readthedocs.io`.

2.3 Emission metamodel

Photon collection by a CCD or particle detector, where impinging photons are counted in spatial,

spectral and temporal bins or bands (i.e., bins with transfer functions), is a Poisson process. In what

follows, temporal binning will not be considered, but the data will also be binned in data quality

classes. Furthermore, the observed number of photons in energy band i, pixel j, and data quality

class m will be denoted by k_{ijm}^{D}. Given some data, ideally one would aim to infer the underlying

true generative model, k_{ijm}^{true}, such that k_{ijm}^{D} is a Poisson realization of k_{ijm}^{true}. However, processes in

the Universe cannot be modeled with arbitrary complexity due to unobservable degrees of freedom.

Therefore, k_{ijm}^{D} is instead modeled parametrically with a metamodel whose forward-model, $k_{ijm}^{\hat{\mathrm{M}}}$, is

the Poisson mean of k_{ijm}^{D}, yielding the log-likelihood

$$
\begin{aligned}
\ln P(D|N, \theta_N) &= \sum_{ijm} \ln P(k_{ijm}^{\mathrm{D}}|k_{ijm}^{\hat{\mathrm{M}}}) \\
&= \sum_{ijm} k_{ijm}^{\mathrm{D}} \ln k_{ijm}^{\hat{\mathrm{M}}} - k_{ijm}^{\hat{\mathrm{M}}} - \ln k_{ijm}^{\mathrm{D}}!.
\end{aligned}
\tag{2.42}
$$

The *forward-model*, $k_{ijm}^{\hat{\mathrm{M}}}$, is the photon counts predicted by the metamodel in energy band i, pixel

j and data quality class m, and is the projection of the metamodel onto the data space. It is a deter-

ministic function of the metamodel parameter space, and obtained by the model flux convolved

with the exposure of the detector towards a particular direction on the sky, (θ_1, θ_2), at energy E, $\epsilon(E, \theta_1, \theta_2)$, and the transfer efficiency, $T(E)$,

$$k_{ijm}^{\hat{\mathrm{M}}} = \iiint \mathcal{M}_{\mathrm{m}}(E, \theta_1, \theta_2)\epsilon(E, \theta_1, \theta_2)T(E)dEd\theta_1 d\theta_2, \qquad (2.43)$$

where $\mathcal{M}_{\mathrm{m}}(E, \theta_1, \theta_2)$ is the flux predicted by the metamodel, i.e., number of photons per area, time, energy, and solid angle. When data is collected, there can be discontinuities such as dead CCDs, cosmic rays, leakage from bright sources. The exposure correction allows these pixels to have zero-weight in order to mask out such data. When the number of photon counts is high, as in optical photometry, the data reduces to a Gaussian process.

2.3.1 MODELING PHOTON EMISSION

The forward-model counts in energy bin i and data quality class m, can be written as a sum of several components,

$$\mathcal{M}_{im} = \sum_{w=0}^{N_{dif}-1} \mathcal{D}_{iw} + \mathcal{I}_i + \mathcal{P}_{im}, \qquad (2.44)$$

which contains N_{dif} spatially varying diffuse emission components, $\mathcal{D}_{iw}$, where $w = 0, 1, N_{dif} - 1$, an isotropic emission component, $\mathcal{I}_i$, which models truly isotropic emission, whether of cosmic or instrumental origin, as well as emission from point sources with fluxes below the minimum of the

flux prior. $\mathcal{P}_{im}$ is the contribution of point sources to the emission model.

Incoming direction of photons can only be measured with finite precision. The spatial and spectral shape of this uncertainty is encapsulated in the PSF of the instrument, $\mathcal{F}_{im}$. Therefore, each $\mathcal{D}_i$ is convolved with $\mathcal{F}_{im}$. The emission metamodel contains a common parameter for the normalization of $\mathcal{D}_{iw}$ and $\mathcal{I}_i$ for each i and w. Log-uniform priors are placed on these normalizations and they are denoted with $\vec{A}$.

In general, photon emission from a point source can be modeled as a delta function in position space with an energy spectrum, which are then convolved with the spatial and spectral instrument response, i.e., the point and line spread functions, of the measuring instrument, respectively. The measurement uncertainty in energy is assumed to be much smaller than the size of the energy bins. The delta function at the position of each point source is then convolved with the PSF, $\mathcal{F}_{im}$, which is in units of the fraction of total flux per solid angle.

PCAT allows the PSF to be modeled with Gaussians

$$\mathcal{F}_{im}(\theta_0; \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\theta_0^2}{2\sigma^2}\right), \tag{2.45}$$

where θ_0 is the angular distance from the point source, King functions (Section 3)

$$\mathcal{F}_{im}(\theta_0; \sigma, \gamma) = \frac{1}{\sqrt{2\pi\sigma^2}}\left(1 - \frac{1}{\gamma}\right)\left(1 + \frac{\theta_0^2}{2\sigma^2\gamma}\right)^{-\gamma}, \tag{2.46}$$

which approximates a Gaussian at small θ_0 and a power-law with log-slope γ at large θ_0, or an Airy

pattern (Section 4),

$$\mathcal{F}_{im}(\theta_0) = \left(\frac{2J_1(c\sin\theta_0)}{c\sin\theta_0}\right)^2, \tag{2.47}$$

where c is a constant and J_1 is the Bessel function of the first kind of order one.

In astrophysics, the PSF of a given instrument can be constrained using simulations or calibration data such as bright and isolated or stacked point sources. In that case, priors on the PSF parameters can be made delta functions. However, in certain applications, where the PSF can change over space, energy or time, the uncertainties in the PSF may be comparable to the uncertainties in the fluxes and positions of the point sources. In that case, the uncertainties in the PSF must be marginalized. In yet other applications such as clustering of arbitrary data sets, the PSF is unknown a priori, and must be inferred along with the elements. Parameters that characterize the PSF in the emission metamodel of PCAT are denoted with $\vec{\eta}$.

Hence, when summed over all point sources, the model point source surface brightness, $\mathcal{P}_{im}$, is obtained in energy bin i and PSF class m, in units of photons per area, time, solid angle, and energy,

$$\mathcal{P}_{im} = \sum_{n=0}^{N-1}\mathcal{F}_{im}(\theta_{1,n},\theta_{2,n})f_{ni} \tag{2.48}$$

where f_{ni} is the flux of the nth point source in the ith energy bin.

In the emission metamodel of PCAT, the elements are light point sources and their parameters are the horizontal position, θ_1, vertical position, θ_2, flux, f, and color, s. The spectra of point sources

66

are either modeled as a power-law with a single color, s_n,

$$f_{ni} = f_n \left(\frac{E_i}{E_0} \right)^{-s_n},$$ (2.49)

as in Sections 3.1 and 3.2, or a set of colors, s_{ni}, for each energy bin other than the pivot energy bin, E_0,

$$f_{ni} = f_n \left(\frac{E_i}{E_0} \right)^{-s_{ni}},$$ (2.50)

as in Section 3.3.

A probabilistic graphical model (PGM) of the emission metamodel PCAT is presented in Figure 2.9. In this representation, nodes denote random variables with certain probability distributions, while edges denote conditional dependencies of the probability distributions of the destination nodes. The red, blue, green and yellow nodes represent the metamodel parameters, which are assigned prior probability distributions. In particular, the red nodes are the hyperparameters μ and α, which set the normalization and slope of the point source flux distribution, respectively. The blue node indicates the number of point sources in a member model, N, i.e., is the multiplicity of each green node. Likewise, the green nodes are the element parameters. The yellow nodes are the common parameters of the metamodel. M_D node is the forward-model, i.e., a deterministic function of the metamodel parameters, which has the same dimensions with the data, D. Along with the hyperpriors and priors on the common parameters, the consistency of M_D with D determines the

posterior. The directed edges are color-coded such that black edges denote a probabilistic relation, whereas olive lines show a deterministic relation. Finally, magenta lines imply that the multiplicity of the destination node is set by the origin. The graph does not use plate notation, since the multiplicity itself is a discrete parameter in the metamodel, which admits a hierarchical prior.

2.3.2   Data and metamodel boundaries

In general, the spaces of the data and forward-model do not have to be coincident. A given data set may have features that can only be explained by model components outside the data space. For instance, an image may have features close to its boundaries that have been caused by sources outside the image.

In order to model emission from sources outside the data space, PCAT allows the spatial prior region to be larger than the image by the Full Width at Half Maximum (FWHM) of the PSF. Therefore, the model point sources can move out of the image and probe whether a feature close to the boundary can be fit better by a model point source outside the image. As a result the offset provides a smooth transition from a data-informed region well inside the image to a prior dominated region outside the image, where the posterior asymptotes to the prior.
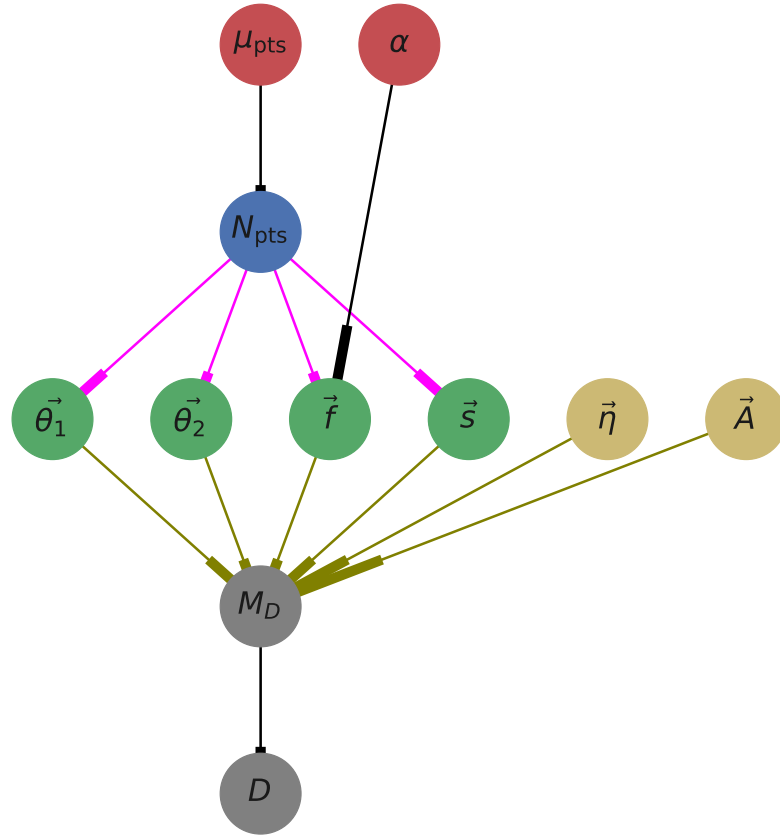
Figure 2.9: The probabilistic graphical model of PCAT emission metamodel. Each colored node in the network corresponds either to a single parameter or a set of parameters (when vectorized) in the metamodel. M_D denotes the forward-model and D stands for the data. Nodes and edges are colored depending on the type of parameter they represent and their conditional dependences, respectively.

3

The gamma-ray sky

Inference of point sources is a fundamental problem in astronomy. Sources may be subthreshold either due to their intrinsic faintness (i.e., photon shot noise), or being fainter than the typical fluctuations of the background emission or nearby brighter sources. As a result, catalogs can become flux-incomplete, and the inferred background emission can be biased high.

In general, there are two inferences that can be performed regarding sources in the model:

- What are the sources above some significance threshold and what are their parameters?

- What are the multiplicity, position, flux and color distribution of the sources?

Traditional catalogs, by construction, can only address the first question, whereas a probabilistic catalog can provide an answer to both questions.

When inferring a catalog from a telescope image, non-Gaussian covariances appear between source parameters (e.g., position, flux, and color parameters) when the number of superthreshold sources per FWHM of the PSF, becomes greater than unity. Therefore, traditional cataloging also mismodels crowded-field images, has lower sensitivity for faint sources, and fails to propagate uncertainties in the source positions, fluxes, and colors to further analyses.

In this section, the inference performance of probabilistic cataloging will be characterized using simulated data. Then, the observed gamma-ray sky towards the northern galactic pole and the inner galaxy will be studied using the emission metamodel of `PCAT`.

3.0.1 Modeling gamma-ray emission

When modeling gamma-ray emission with `PCAT`, it is assumed that the flux of the nth point source in the pivot energy bin, f_n, is distributed as a power-law between f_{min} and f_{max}, with the log-slope $-\alpha$ at the central flux. The value of f_{min} will be different for each section in this chapter, whereas f_{max} will be fixed at 10^{-6} cm^{-2}s^{-1}GeV^{-1}.

Throughout this chapter the gamma-ray data is assumed to have been collected by the Fermi-LAT and is in the form of photon counts binned spatially, spectrally, and in PSF quality classes. The latter allows better modeling of point sources, since photons that convert at the top of the instrument, i.e., front-type events, have smaller angular reconstruction uncertainties. The PSF quality

binning is based on front and back converted events for Pass 7 data (Section 3.2), whereas it is based on the two best PSF classes for the Pass 8 data (Sections 3.1 and 3.3). The energy bins are taken as $0.3 - 1, 1 - 3$ and $3 - 10$ GeV. The pivot energy is taken as $E_0 = 0.5$ GeV for the Sections 3.1 and 3.3 and $E_0 = 1.7$ GeV for the Section 3.2. Furthermore, spatial binning is performed in `HealPix` pixels with base resolution 256.

A weighted sum of two King functions is used to model the PSF[8]. Furthermore, given that the Fermi-LAT energy resolution of $\sim 10\%$ is smaller than the energy bin width, infinite energy resolution is assumed and the line spread function of the detector is neglected.

Given that the PSF of the Fermi-LAT is below 1° in all the energy bands of interest, the radius scale of split and merge proposals, θ_{sm} is taken as 1° in order to optimize the exploration of covariant elements. Point source associations are also performed with an association angle, θ_{asc}, of 1°.

Posteriors in Sections 3.2 and 3.3 are compared to 3FGL to assess the performance of probabilistic cataloging. Thus, the choice of energy binning coincides with that of the 3FGL catalog, in order to facilitate associations in Sections 3.2 and 3.3.

3.1 Simulated data

In order to validate the framework, probabilistic cataloging is first performed on simulated data. Towards this purpose, the posterior of the metamodel is compared with the parameters of the true metamodel, which is forward-modeled to generate the data.

In order to generate the simulated data, a random sample is drawn from the prior of a member

of the emission metamodel with 100 point sources. Then, the sample is forward-modeled to obtain surface brightness. A flat exposure is assumed to convert surface brightness into photon counts. Lastly, in order to account for the photon shot noise, a Poisson realization of the forward-model counts is drawn as the simulated data.

The simulated data count map in the 1 GeV - 3 GeV energy bin is shown in the top left panel of Figure 3.1. When probabilistic cataloging is performed on this data, the posterior of the mean forward-model and residual count maps are obtained as shown in the top right and bottom panels of Figure 3.1. Posterior of the mean residual count maps contain no coherent features and give an averaged χ^2 per degree of freedom of 1.04.

Figure 3.2 shows six fair samples from the posterior of the catalog space superposed on the same data count map. The area of the markers are proportional to the flux of the point source they represent. Sample-to-sample variation of bright model sources are smaller compared to faint model sources. Furthermore, some faint model point sources change labels (i.e., are born, split, killed, or merged) across frames.

Next, the top left panel in Figure 3.3 shows the mean of the posterior (blue) and simulated (green) flux histogram with the 68% credible intervals for the former. The posterior is consistent with the true catalog, although there is some oversliptting at the bright end due to crowdedness of the image. For reference, Figure 3.4, shows the posterior consistent with a similar data set, but one whose dimensions are doubled, i.e., the number of sources per FWHM of the PSF is reduced by a factor of 4.

In crowded fields, a more robust comparison between the true metamodel and the posterior
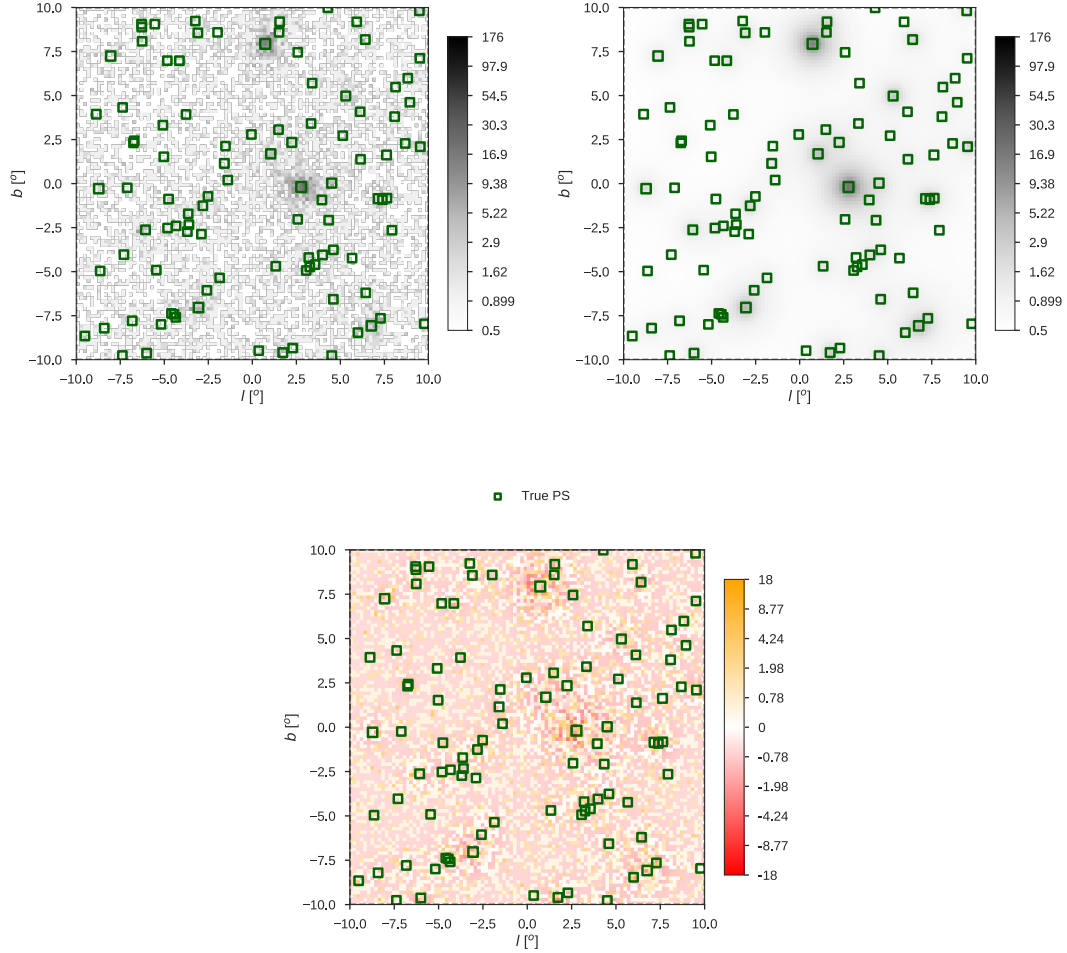
Figure 3.1: The simulated gamma-ray count map (top left), the posterior of the mean forward-model count map (right), and the posterior of the mean residual count map (bottom). The color scales indicate the number of photons per pixel in all panels and are `arcsinh` stretched in order to make faint features more visible. Superimposed with the count maps is the catalogs (green). The sizes of the markers are proportional to the logarithm of the fluxes of the corresponding point sources.
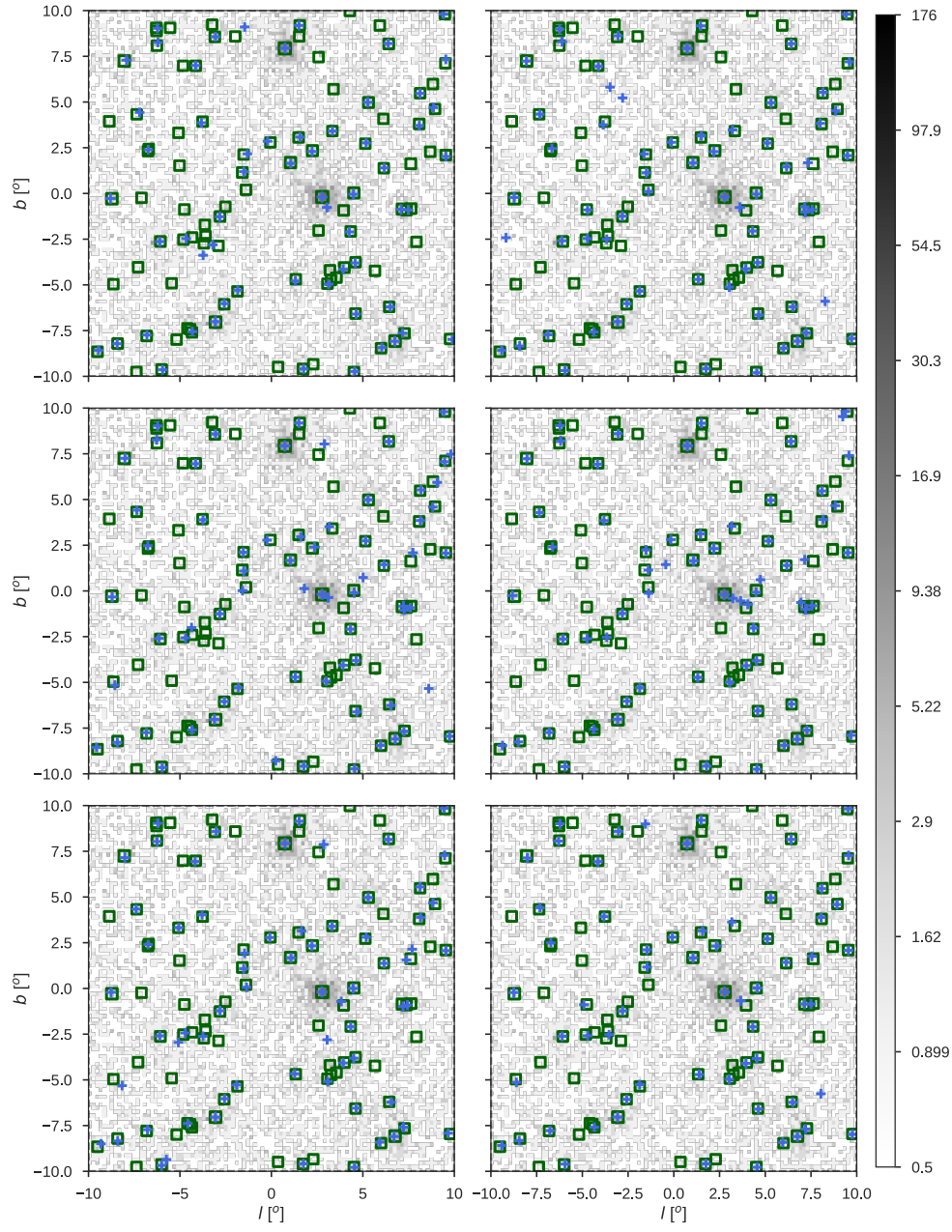
Figure 3.2: Six fair draws from the posterior of the emission metamodel, superposed on the data count map (same as the left panel of Figure 3.1).

can be performed based on the significance. In what follows, significance will be defined as the log-likelihood improvement of adding a given element to the model. The posterior of the color (top right), photon count (bottom left), and significance (bottom right) histograms are shown in Figure 3.3, which show the agreement between the posterior and the true catalog.

Figure 3.3 shows the posterior of the histograms, but this time corrected for the regularization prior. It overplots the transfer function due to the regularization prior in purple with the label PTFN. These are obtained by taking the ratio of the number of true elements in each bin above the log-likelihood penalty (35 for these point sources with 5 degrees of freedom) with the total number of true elements in the bin. This function can be used as a prior in an inference without the regularization prior, or as a correction factor with which the posterior with the regularization prior can be rescaled to undo the effect of the regularization prior, with the caveat that this also increases the posterior uncertainties. In Figure 3.3, the values of the transfer function are plotted after multiplication with 10. Therefore, when the transfer function is 10, that implies that all elements in that bin are above the siginificance scale of the regularization prior.

The former comparison shows that the *distribution* of the posterior and true catalogs are consistent. However, it is also important to check element-wise consistency between the two catalogs. Towards this purpose, the posterior samples from the catalog space are associated with the true catalog as discussed in Section 2.2.3. Figure 3.6 illustrates the resulting correlation. The horizontal axes show the fluxes (top left), colors (top right), photon counts (bottom left), and significances (bottom right) of the true sources, while the vertical axes show the relevant posterior quantities associated to the true point sources. The vertical error bars denote the statistical uncertainty due to marginaliza-
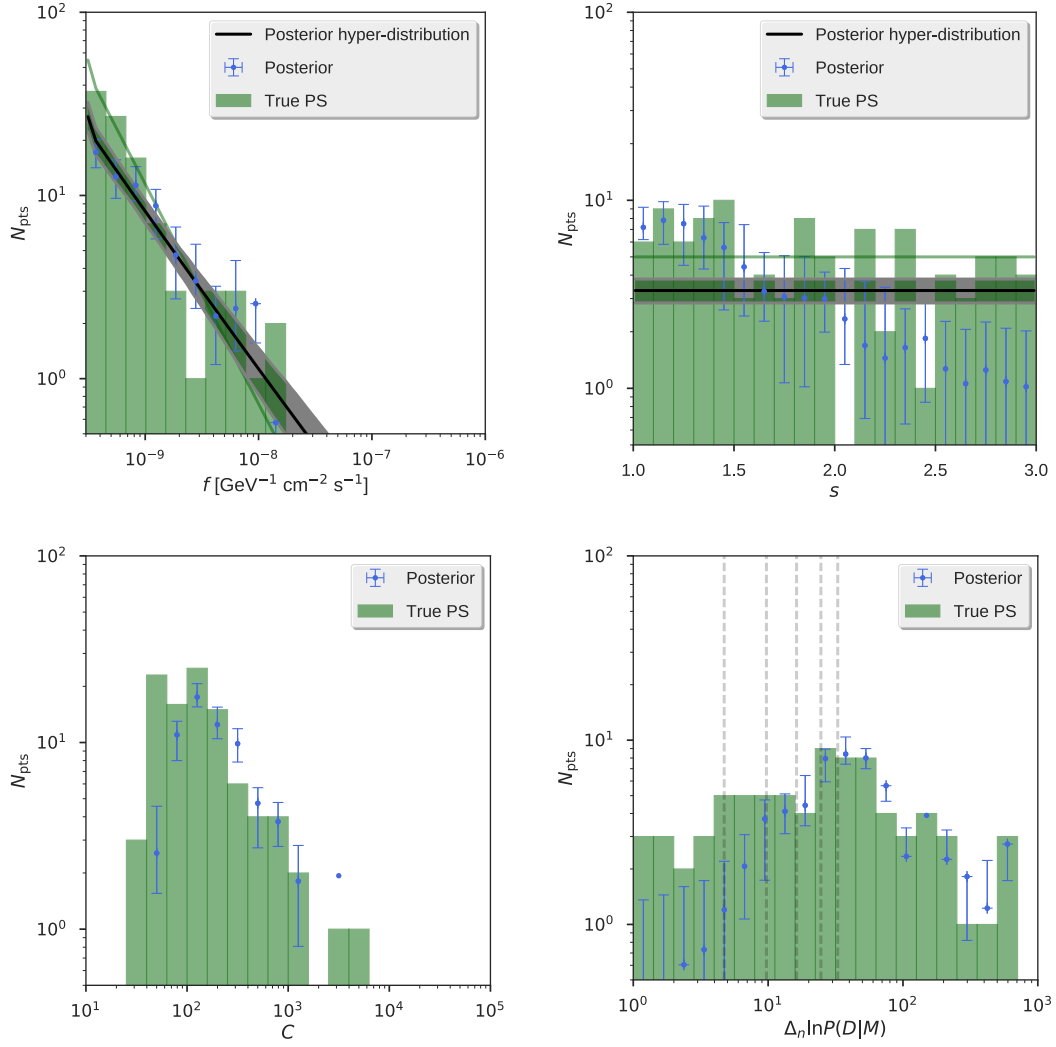
Figure 3.3: Posteriors of the flux (top left), color (top right), photon color (bottom left), and significance (bottom right) histograms, showing the posterior mean, 68% and 95% quantiles (blue). The green histograms show the distribution of simulated point sources.
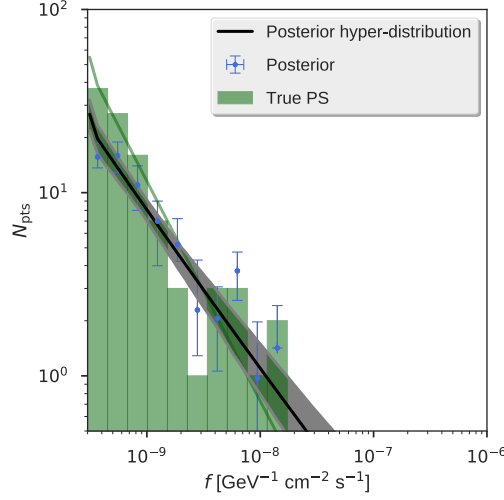
Figure 3.4: Posterior of the flux histogram consistent with an image whose dimensions have been stretched by a factor of 2, reducing the crowdedness.

tion over the posterior of the emission metamodel. Horizontal error bars are not provided, since the true catalog is a simulated catalog without any uncertainties. At the bright end, model point sources are superthreshold and well localized, yielding small vertical error bars. Moving towards the faint end, the correlation broadens due to crowdedness and the model point sources becoming fainter than the Poisson fluctuations of the background. In the extreme faint limit, one expects a given true point source to be associated with a random model point source, completely suppressing the correlation. The spread increases towards low fluxes because of the drop in the significance of the elements, which result in larger flux uncertainties as well as lower completeness.

When generating the simulated data, the colors of the true point sources are chosen such that the significance of the point sources mostly correlate with the flux at the pivot energy bin of 0.3 GeV - 1 GeV. This is a typical problem in traditional cataloging, where the catalogs are more complete in
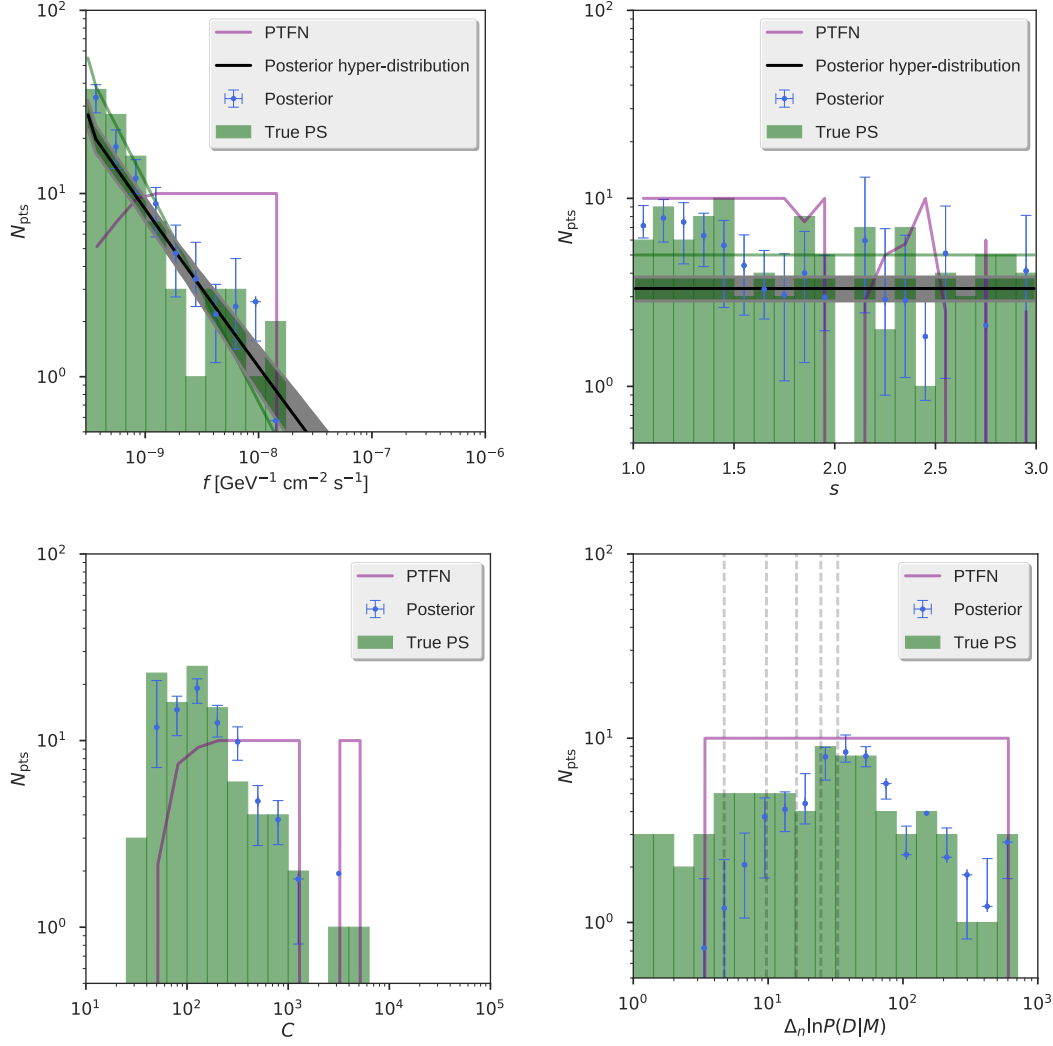
Figure 3.5: Similar to Figure 3.3, but corrected for the effect of the regularization prior. The purple line shows the transfer function of the regularization prior, i.e., in a given bin the ratio between the number of true point sources with significance above 2.5, to the total number of true point sources. The posteriors shown *are* corrected by these transfer functions.
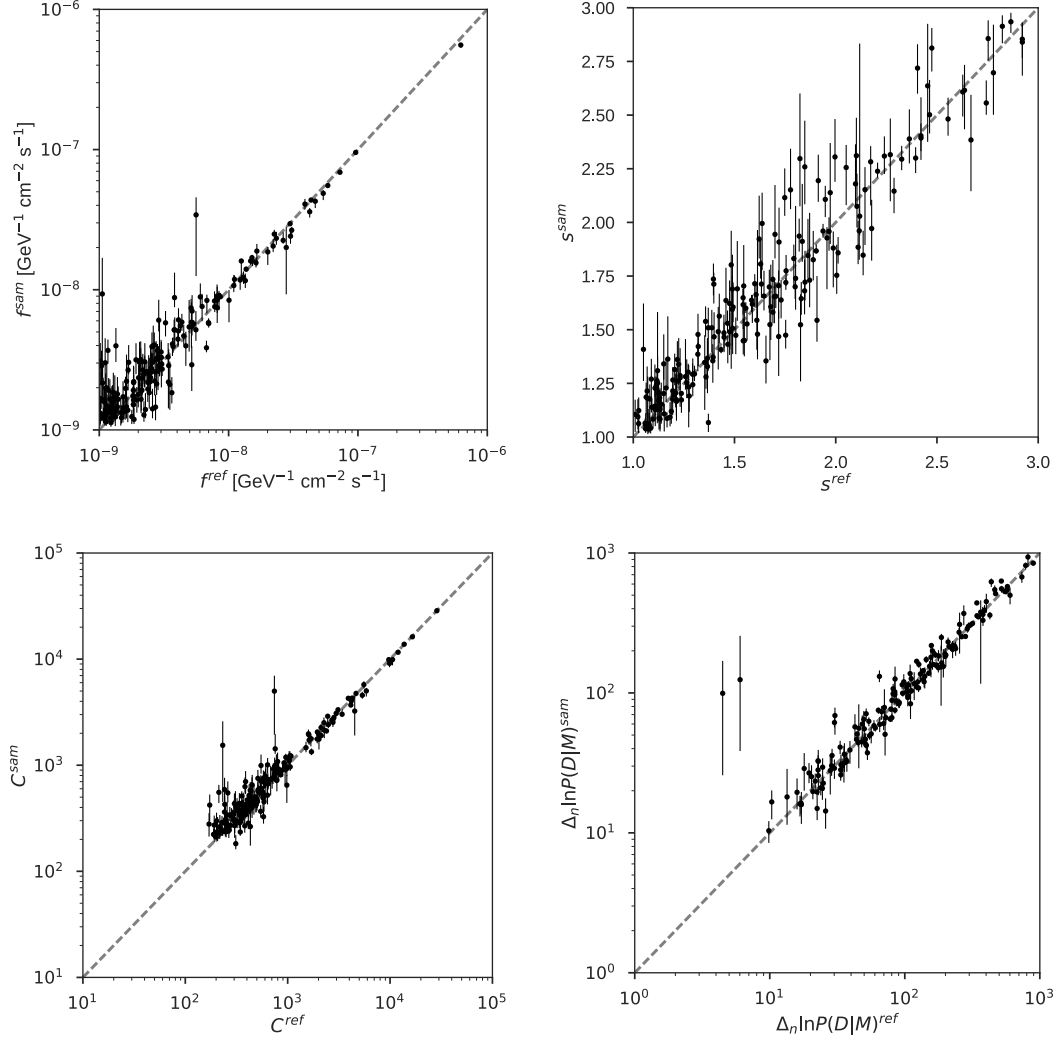
Figure 3.6: Posterior of the flux (upper left), color (upper right), count (lower left), and significance (lower right) of associations to the true catalog. In the absence of Poisson noise and vanishing crowdedness, associations would lie on the diagonal line. The vertical error bars correspond to statistical uncertainty and increase with decreasing significance, i.e., log-likelihood change in adding the point source to the emission metamodel. Outliers are due to crowdedness of the true catalog, where point source confusion causes wrong associations.
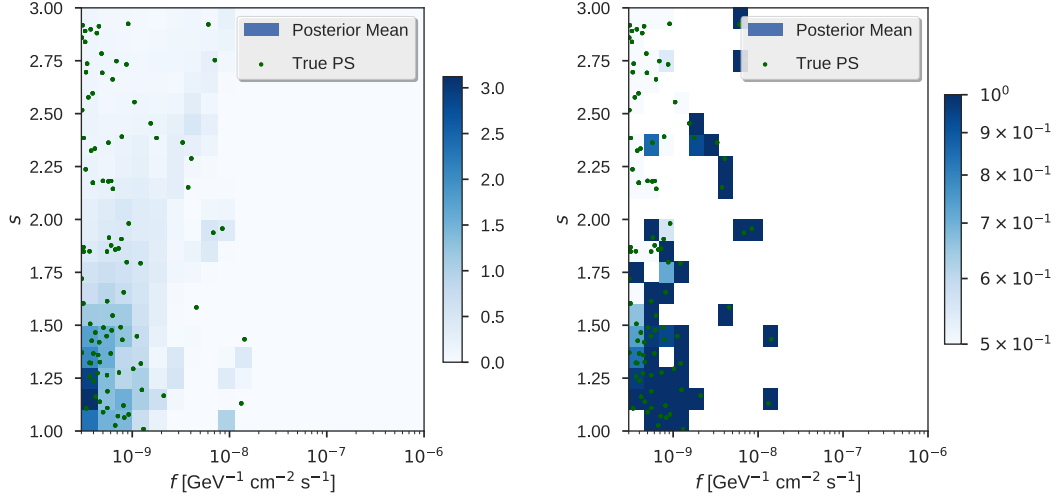
Figure 3.7: Posterior of the joint histogram of the flux and color of model point sources (left) and the completeness as a function of flux and color of true point sources (right).

certain bands than others. Therefore, requirement of detections for further inference on the flux-color space, causes loss of information and bias. Figure 3.7 shows the posterior of the joint histogram of point source fluxes and colors (left) and the completeness as a function of true point source flux and color (right).

In order to further illustrate the effect of the regularization prior on the posterior, the posterior of the model indicator, mean number of point sources, the log-slope of the flux distribution and the background normalization in the middle energy bin are shown in Figure 3.8. The posterior is not consistent with the true values because of the regularization prior, which affects the posterior significantly at low signal-to-noise. Therefore, the posterior must be corrected (or have been sampled with a prior modified by the regularization prior). The posterior of the model indicator shown in the top left panel of Figure 3.8 yields the Bayes factor between the models.
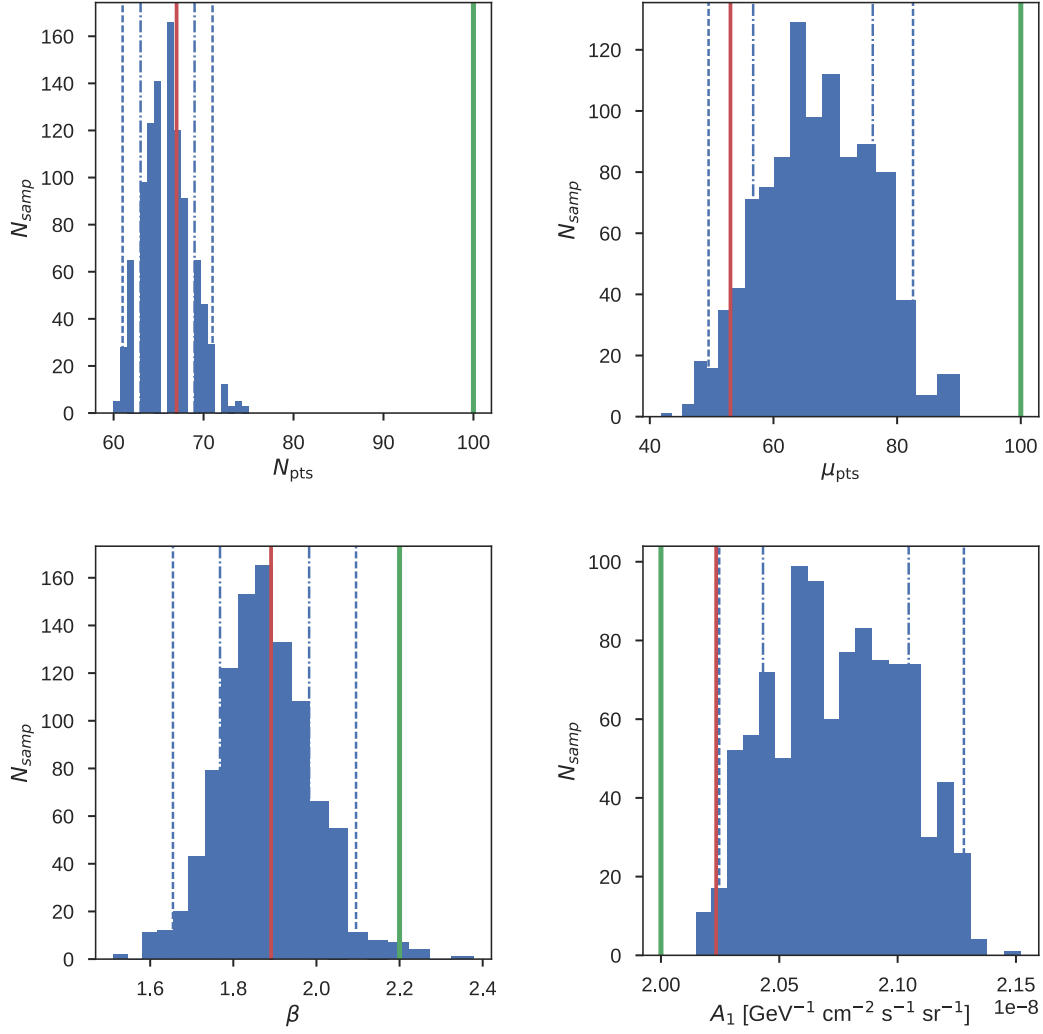
Figure 3.8: Posterior of the number of point sources (top left), mean number of point sources (top right), log-slope of the flux distribution (bottom left), and the background normalization in the energy bin 1 GeV - 3 GeV (bottom right). The true values are denoted with green lines. The value of the maximum-likelihood sample is shown with the red line.

Another approach to probing subthreshold point sources is the fluctuation analysis where the 1-point probability distribution of the emission is used to estimate the contribution of subthreshold point sources to the total emission[194]. By modeling the tail of the distribution of deviations due to subthreshold point sources below the detection threshold, fluctuation analysis can distinguish truly isotropic emission from subthreshold point sources. The method has been considered across the whole electromagnetic spectrum, e.g., in radio[228,61], far-infrared[100], X-rays[90,236] and gamma-rays[93,241,143].

Probabilistic cataloging does not rely on this approach, but can produce the marginal posterior of the 1-point function of counts per pixel. Figure 3.9 shows the 1-point function of two components, i.e., the isotropic (left), and the subthreshold point source (right). The true histogram and the posterior are overplotted with green and blue along with the 1-point function of the data (black). The posterior distribution of the subthreshold point sources significantly differ from the posterior of the isotropic emission component, indicating the existence of subthreshold point sources.

3.2   Northern Galactic Polar Cap

In this section, the observed gamma-ray data in the direction of the North Galactic Pole (NGP) is probabilistically cataloged and the posterior is compared with the 3FGL.

Reprocessed Pass 7, source class gamma-ray data inside a $40° \times 40°$ square around the NGP, i.e., the Northern Galactic Polar Cap (NGPC), is binned spatially, spectrally and in PSF quality classes (i.e., front and back converted events) provided by the Fermi-LAT collaboration (i.e., events that
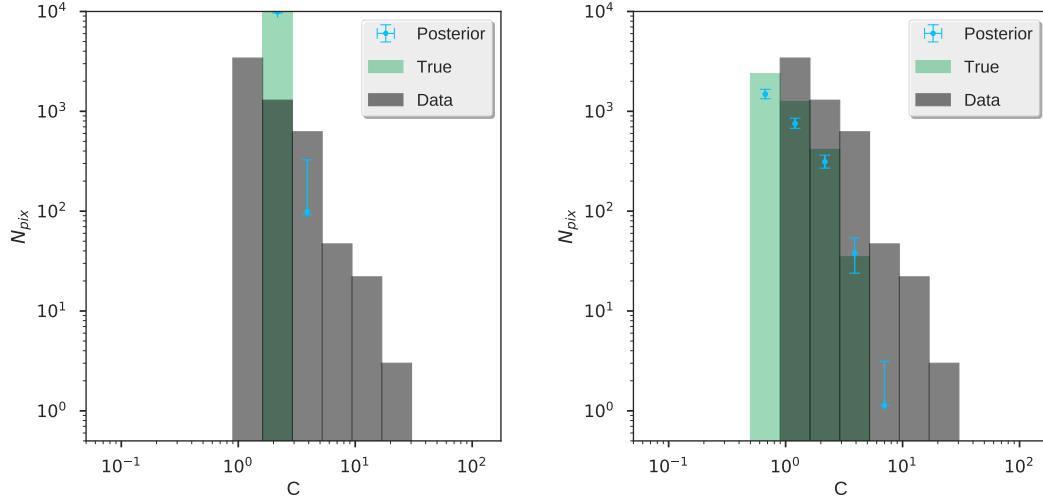
Figure 3.9: Posterior of the histogram of number of counts per pixel, shown with blue, . The data is shown with the black.

convert in the front and back of the instrument). Furthermore, since most of the sources at high galactic latitudes are time-variable blazars, the data in the same time interval as that used to construct the 3FGL catalog, i.e., weeks 9 through 217, is used. The exposure map of the Fermi-LAT for the same data type and acquisition interval is used. Importantly, in this section, regularization prior is not applied.

At $\sim$ GeV energies, modeling the spectra of the point sources using a power-law is not an accurate description of the data. Galactic gamma-ray emitters such as pulsars are known to exhibit an exponential cutoff in their spectra at $\sim 1 - 10$ GeV. Nevertheless, most of the sources in the NGPC are expected to be extragalactic. Moreover, extragalactic gamma-ray sources such as blazars and radio galaxies have curved spectra, i.e., a power-law with a running index, due to their broadband inverse Compton emission on the external radiation field of the Active Galactic Nucleus (AGN). To address

this concern the energy span is restricted to 1.5 decades, i.e., between 0.3 GeV and 10 GeV, where curved spectra can be approximated using a single power-law.

Two background emission components are used, i.e., the diffuse emission model[5] and an isotropic emission component for each energy bin. At high galactic latitudes such as in the NGPC, the diffuse model is approximately equal to the interstellar medium dust column density[7,196].

Figure 3.10 tiles together fair samples from the posterior showing the number of data counts in each pixel, i.e., similar to the left panel in Figure 3.2, but showing observed NGPC data instead. The grid illustrates the typical evolution of the MCMC state, where bright true point sources have an associated model point source with precise spatial and spectral localization, whereas faint 3FGL point sources are associated with model sources in some fraction of samples. Because some of the flaring 3FGL sources are detected over a small period of time, not all 3FGL sources are above detection threshold in this data set.

Figure 3.11 shows the posterior of the flux and color histograms consistent with the NGPC. That the mean of the posterior color distribution is 2.2 is consistent with majority of the point sources being blazars, where the lower tail is dominated by Flat-spectrum Radio Quasars (FSRQs) and BL Lacs make up the harder subpopulation[79].

Next, samples from the posterior catalog space are associated with the 3FGL. Figure 3.12 shows the posterior of the fluxes associated with the 3FGL. Posterior of the point source fluxes associated with the 3FGL agree with the 3FGL fluxes, with an increasing uncertainty towards the dim end. The correlation is stronger at larger fluxes, where covariances with the background normalization and the PSF do not significantly affect the posterior on the fluxes. The level of agreement decreases towards
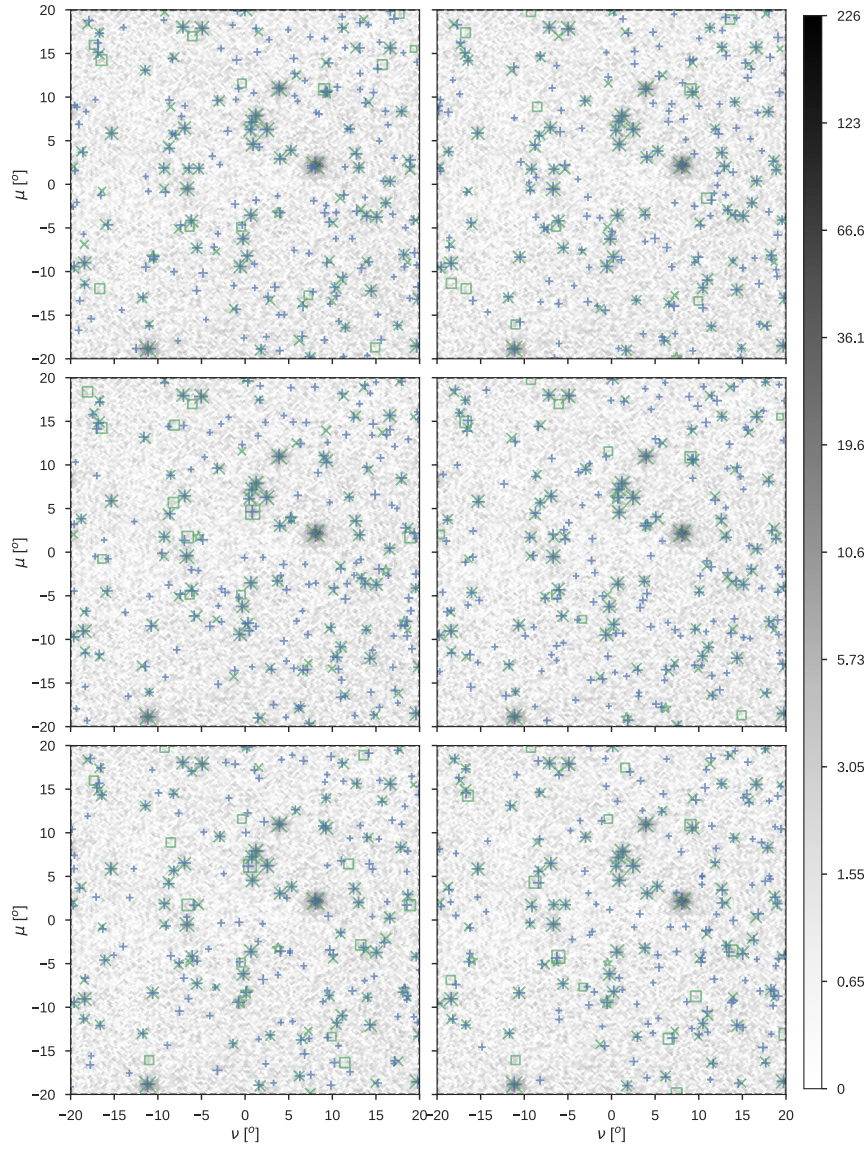
Figure 3.10: The NGPC photon count map as measured by the Fermi-LAT (color scale), the 3FGL sources in the region (green markers) and six fair samples from the catalog space (blue pluses). The image is centered at the NGP and the axes correspond to a Cartesian projection about the NGP. The size of the markers are proportional to the logarithm of the flux of the point source. The color scale corresponding to the number of photon counts per pixel has been `arcsinh` stretched in order to emphasize faint features. The 3FGL sources are marked with a green square if the sample catalog does not have a model point source within $0.5°$. Otherwise they are marked with a green X, indicating a hit.
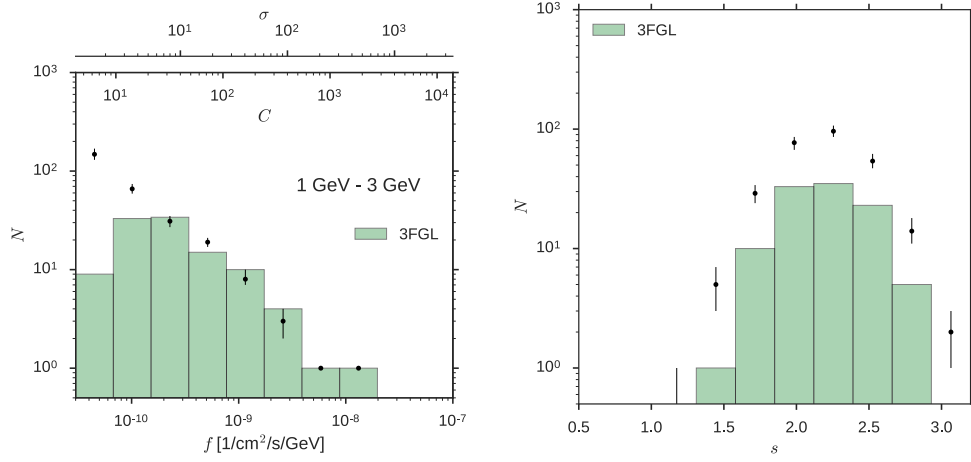
Figure 3.11: Posterior (black) of the flux (left) and color (right) histograms consistent with the NGPC and the histograms of the 3FGL (green) fluxes and colors. In the left panel, the top axes show the corresponding photon counts (C) and the significance (σ).

lower fluxes, which is partially due to the different spectral modeling used by the two catalogs.

By summing fluxes of the point sources in the posterior catalog samples, the posterior contribution of the point sources to the total emission in the NGPC can also be inferred. The median spectra of the point sources, isotropic, and diffuse components are given in Figure 3.13. The diffuse model is observed to be the dominant component, accounting for $57\%^{+6}_{-5}$ of the total emission. The isotropic component and the point sources account for the rest in roughly equal amounts, i.e., $25\%^{+4}_{-3}$ and $18\%^{+2}_{-2}$, respectively. However, the partitioning between the point sources and the isotropic component is set by the choice of the minimum allowed point source flux. If the minimum flux allowed for the point sources is lowered, then the relative contribution of the point sources would account for some of the isotropic component, since emission from faint point sources are nearly degenerate with the emission of the isotropic component. Therefore, the inference of how
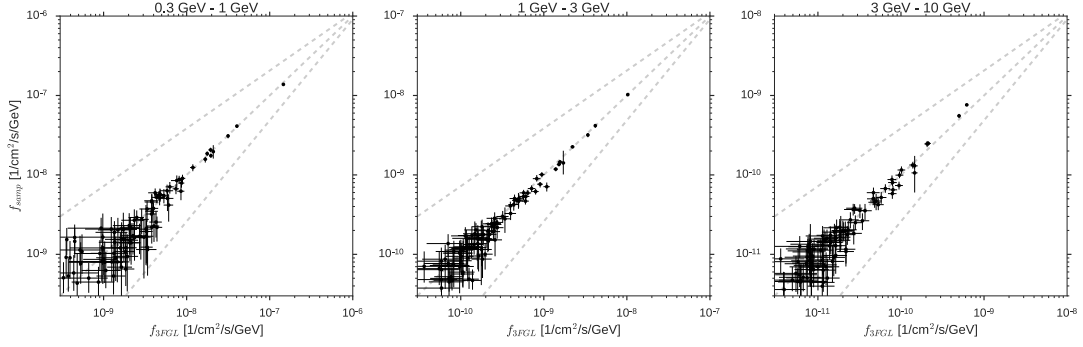
Figure 3.12: Posterior of the fluxes associated with the 3FGL in all energy bins.

much of the emission is accounted for by the point sources should be addressed with reference to a particular minimum point source flux.

Posterior of the hyperparameters provides another handle on the population characteristics. Figure 3.14 shows the posterior of the log-slope of the flux distribution consistent with the NGPC.

Posterior of the power-law slope is inferred to be $-1.92^{+0.07}_{-0.05}$. This is smaller than the expectation from a uniform distribution of equally bright blazars, i.e.,

$$\frac{dN}{df} = \frac{dN}{dr}\left(\frac{df}{dr}\right)^{-1} \propto r^2 \times \left(\frac{1}{r^3}\right)^{-1} = f^{-5/2}. \tag{3.1}$$

Previously,[3] found that the source count function has a slope of -2.6 ± 0.2 at the bright end, which hardens to -1.6 ± 0.1 at the faint end. Because a single power-law is used in this inference, the resulting posterior converges to an intermediate value. Moreover, given that the inference is based not on the full-sky, but only on the NGPC, the posterior of the number of point sources is subject to shot noise around the mean number of true point sources.
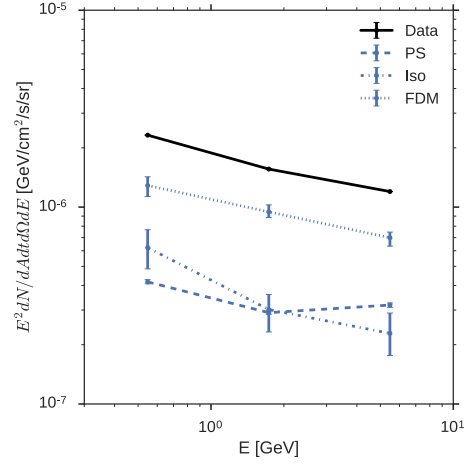
Figure 3.13: Spectra of emission correlated with the Fermi diffuse model (dotted), isotropic model (dot-dashed) and the total emission from point sources (dotted) averaged over the region.
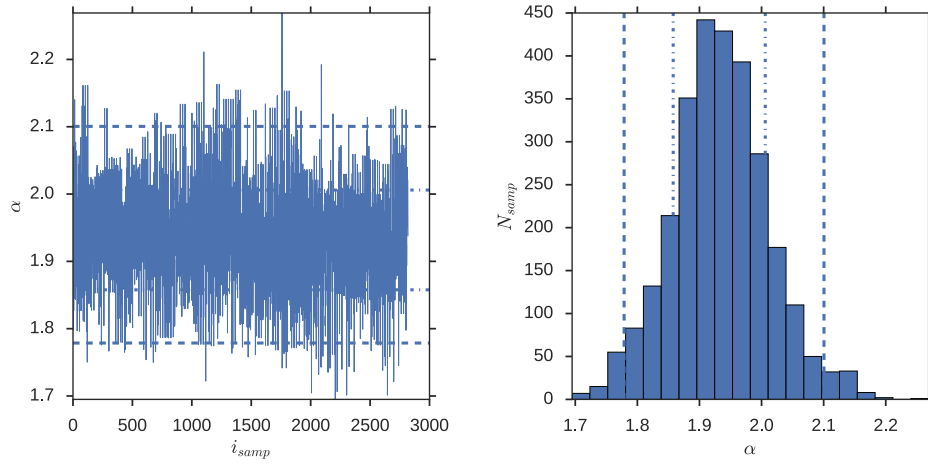
Figure 3.14: Posterior of the power-law slope of the flux distribution consistent with the NGPC.
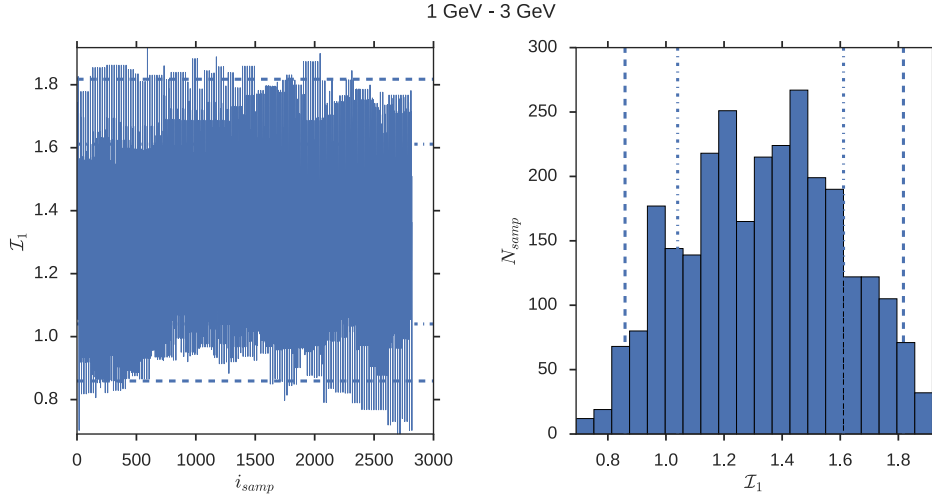
Figure 3.15: Posterior of the normalization of the isotropic emission component consistent with the NPGC.

Figures 3.15 and 3.16 show the normalization of the isotropic and diffuse components for each energy bin, respectively. Median of the posterior isotropic and diffuse background normalizations are inferred to be larger than unity by a factor of ~ 1.3 and ~ 1.1. Moreover, because the diffuse model is relatively featureless in the NGPC, the two normalizations have a large covariance.

Figure 3.17 shows the posterior of the number of point sources. It is inferred that there are 270^{+30}_{-10} point sources in the region above 3×10^{-11} cm^{-2}s^{-1}GeV^{-1} in the $1-3$ GeV energy bin.

3.3   Inner galaxy

The GeV excess can be the emission from subthreshold point sources[144,23,2,126,240,40,178] in the inner Milky Way, cosmic ray energy injection from the supermassive black hole at the center of the Milky Way[177,50] or WIMP annihilations. A hypothesis test regarding the GeV excess that precedes this

1 GeV - 3 GeV
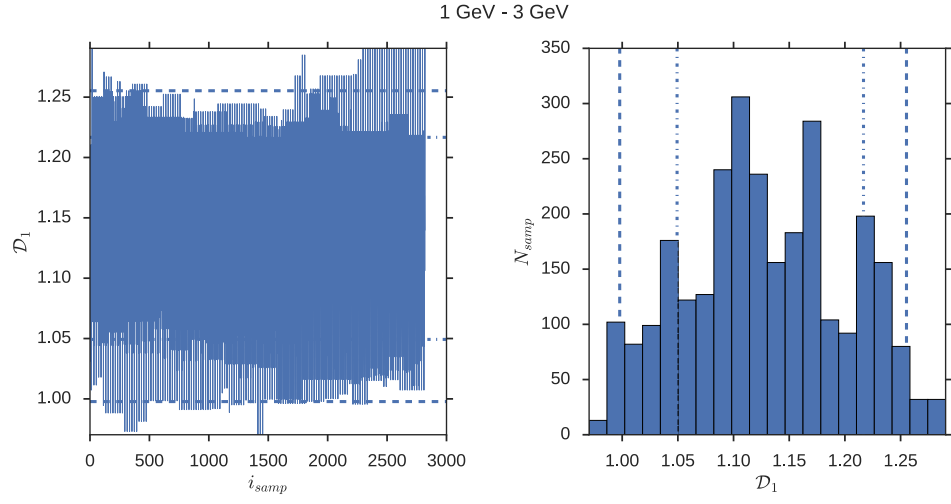
Figure 3.16: Posterior of the normalization of the diffuse emission component consistent with the NGPC.
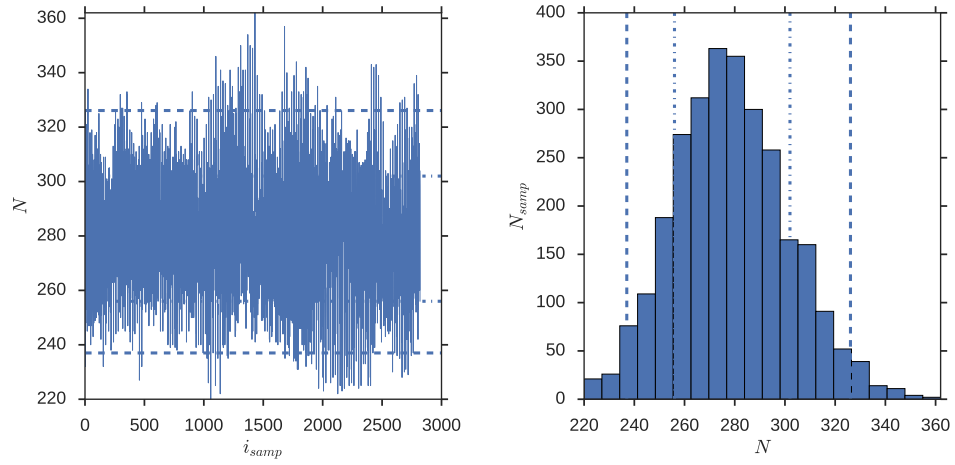
Figure 3.17: Posterior of the number of point sources consistent with the NGPC.

hypothesis test, however, is whether or not it is of diffuse or point source origin. Previously, the GeV excess has been studied with the non-Poissonian template fitting[144] and wavelet decomposition[23] and substantial evidence has been found for the subthreshold point source origin.

The inner Milky Way is a crowded field of gamma-ray point sources and its modeling can be improved with probabilistic cataloging. Furthermore, probabilistic cataloging can probe the properties of subthreshold point sources in this region.

Towards this purpose, 8 years of Pass 8, source class Fermi-LAT data inside $\pm 10°$ of the galactic center with standard data quality cuts, is binned spatially, spectrally and, in PSF quality classes. Then, using this data, probabilistic cataloging is performed using the emission metamodel of PCAT.

The set of background emission components include, in addition to the isotropic and diffuse emission components, the squared and projected NFW profile, which is referred to as the *dark* emission component.

The observed photon count map in the energy bin 1 GeV - 3 GeV and the posterior of the mean residual count map are shown in Figure 3.18.

Figure 3.19 shows the data and posterior of various emission components spatially averaged inside $\pm 5°$ without (left) and with (right) the regularization prior. When the regularization prior is applied nominally, the posterior of the dark emission component is inferred to be higher than that of the subthreshold point sources in the 1 - 3 GeV and 3 - 10 GeV energy bins. When the regularization prior is not applied, however, the posterior of the dark emission component is consistent with zero in the 0.3 GeV - 1 GeV and 1 GeV - 3 GeV energy bins, and a factor of 3 smaller than the contribution of the subthreshold point sources.
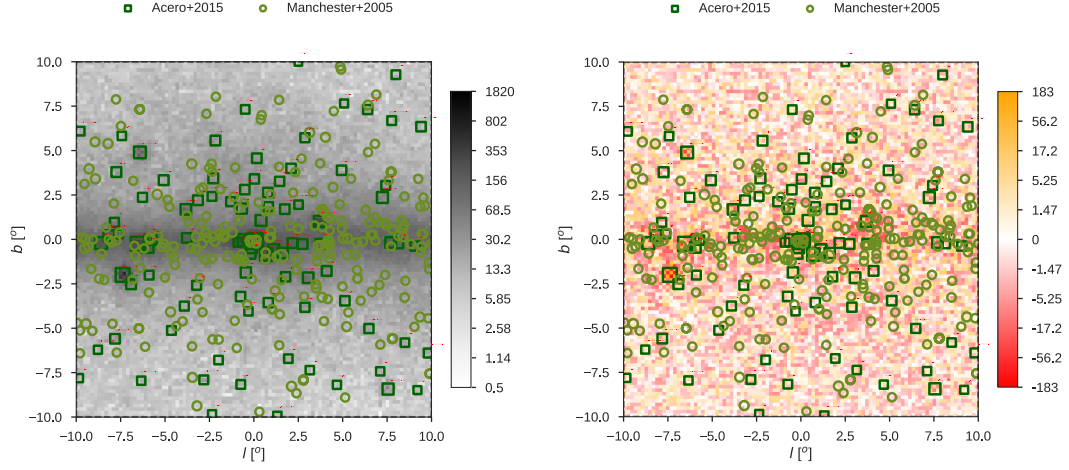
Figure 3.18: Left: the photon count map observed by the Fermi-LAT. Right: posterior of the mean residual count map.
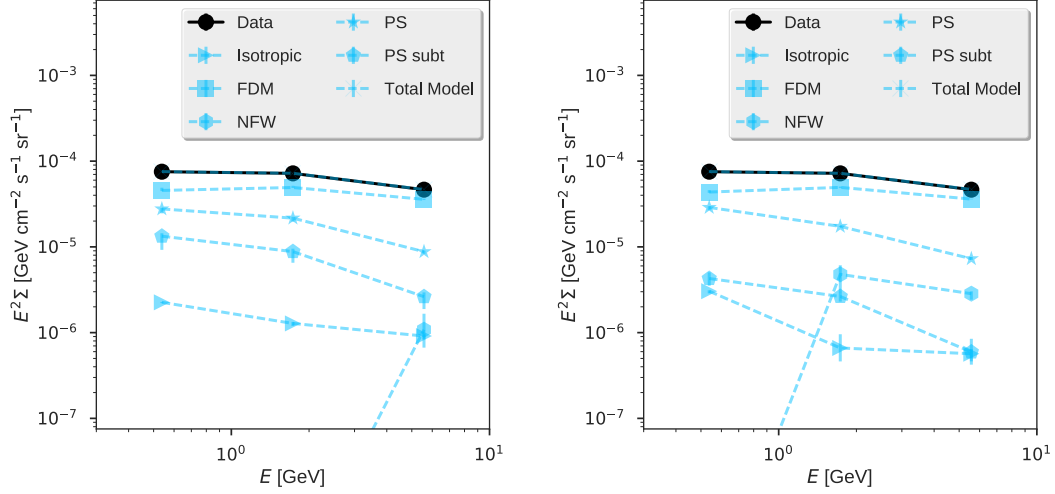
Figure 3.19: Posterior of the spectra of various components in the inference of the emission metamodel consistent with the gamma-ray emission in the inner Milky Way, spatially averaged over the inner $\pm 5°$ without (left) and with (right) the regularization prior.

Figure 3.20 shows the posteriors of other inferences where the strength of the regularization prior is varied as in Equation 5.9. In general, the stronger the regularization prior becomes, the larger the posterior significance of the dark emission component becomes. In particular, the bottom right panel of Figure 3.20 corresponds to an inference, where essentially all subthreshold point sources are removed from the posterior. In that limiting case, the dark emission component can fully explain the GeV excess. However, as the strength of the regularization prior is reduced, the significance of the dark emission component diminishes, and becomes insignificant when the regularization prior is completely removed.

Furthermore, Figures 3.21 and 3.22 show the posterior of the joint histogram of the point source positions when the regularization prior is omitted and applied, respectively. In the former case, subthreshold point sources can explain the GeV excess, whereas in the latter, the GeV excess is mostly explained by the dark emission component.

Figure 3.23 also shows the mean of the posterior of the joint histogram of the point source positions when the regularization prior is applied, but the dark emission component is not in the model. In that case, the GeV excess is forced to be explained by subthreshold point sources and Figure 3.23 yields a map of the locations of the most significant subthreshold point sources that could be contributing to the GeV excess.

Sample catalogs from the posterior can be associated with external reference (i.e., traditional) catalogs to extend the posterior onto quantities about which there was no information in the probabilistically cataloged data set. Figure 3.24 presents the posterior of the pulsar periods assigned to the samples from the catalog space by associating the samples with the ATNF catalog. It is inferred that
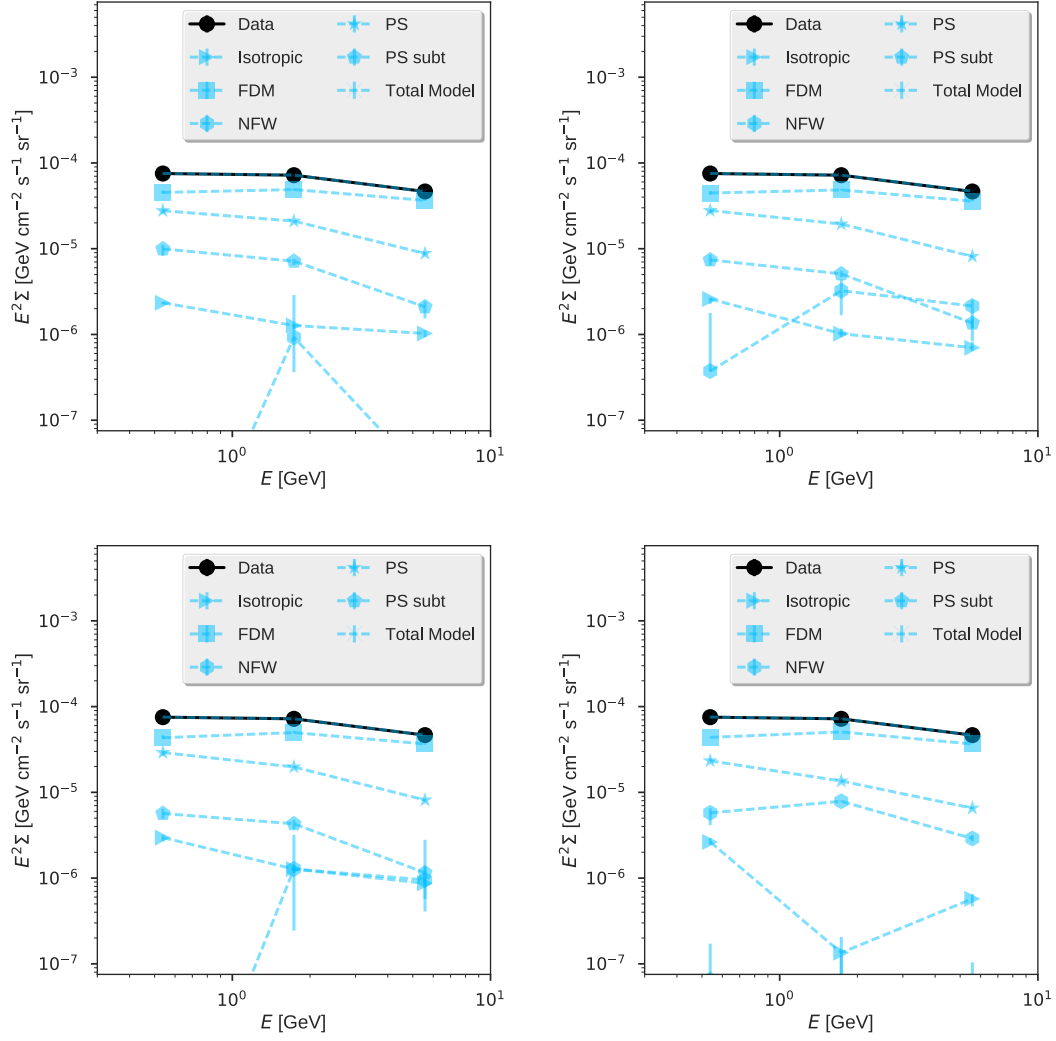
Figure 3.20: Similar to Figure 3.19, but with $\alpha_p = 0.25$ (top left), $\alpha_p = 0.5$ (top right), $\alpha_p = 0.75$ (bottom left), and $\alpha_p = 10$ (bottom right).
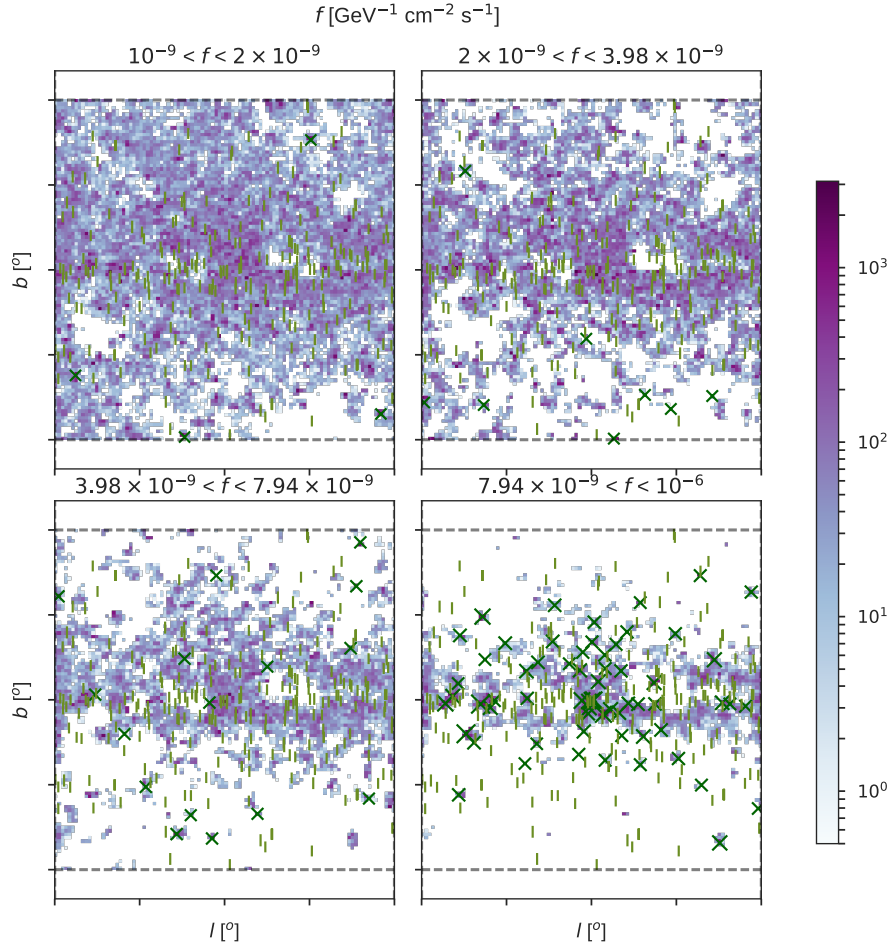
Figure 3.21: Mean of the posterior of the galactic longitude and latitude histogram of model point sources. The color scale gives the number of element samples in a given pixel.
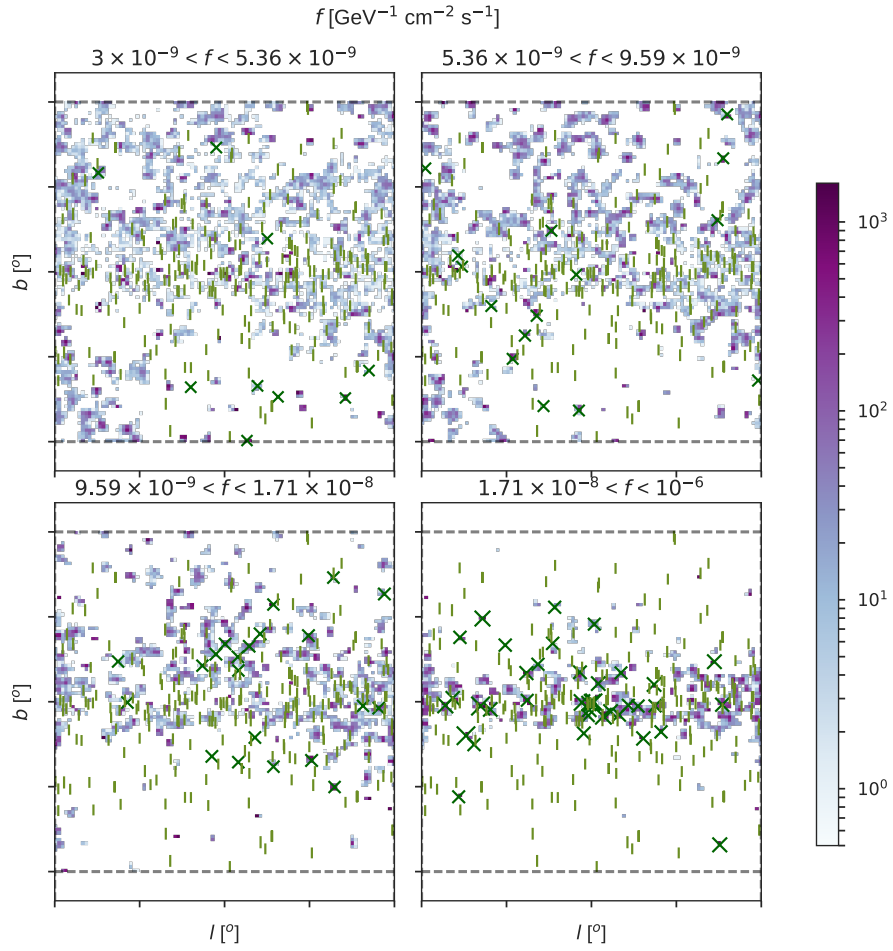
Figure 3.22: Similar to Figure 3.21, but with the regularization prior.
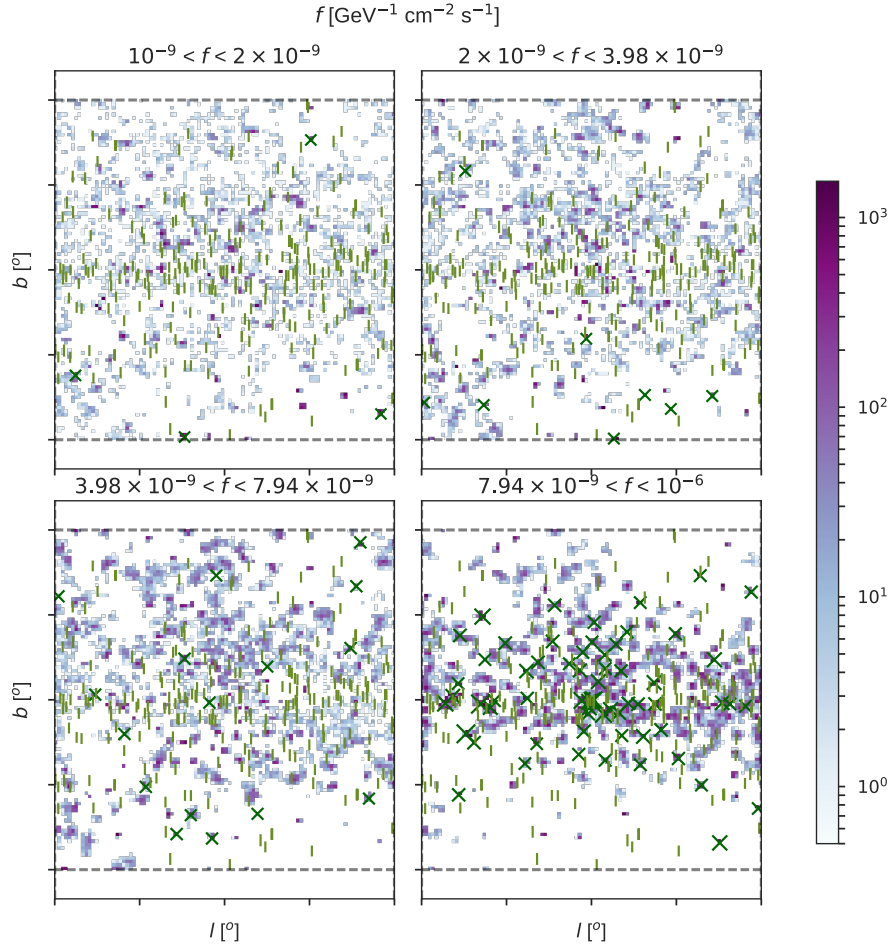
Figure 3.23: Similar to Figure 3.21, but for an inference where the regularization prior is applied, but there is no dark emission component to account for the GeV excess.
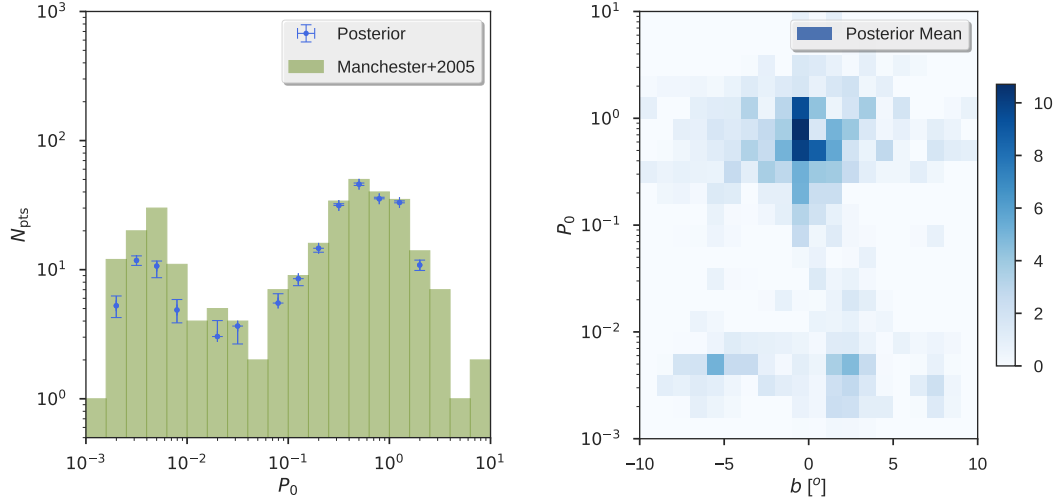
Figure 3.24: Left: Posterior of the histogram of the ATNF periods associated to the catalog samples. Right: Mean of the posterior of the joint histogram of the galactic latitude and the pulsar period of the ATNF source associated to the model point sources.

the posterior catalog space consistent with the gamma-ray data is associated with a larger fraction of young pulsars compared to the millisecond pulsars. This is likely due to the fact that point sources in the catalog samples fit the GeV excess and the residuals along the galactic plane. Because these regions have a higher fraction of young pulsars, these posterior samples consistent with gamma-rays preferentially correlate with young pulsars. Potential discovery of the milliseconds in the region can reduce this correlation.

The hypothesis test between a diffuse and subthreshold point source origin for the GeV excess can also be studied using the 1-point function of photon counts per pixel, after marginalizing over the rest of the emission metamodel. Figure 3.25 shows the 1-point functions of the photon counts in the 1 GeV - 3 GeV energy bin of three components: all point sources (top), subthreshold point sources (middle), and the dark emission component (bottom) for two inferences with (left) and

without (right) the regularization prior. In the right column, the absence of the parsimony prior makes the dark emission component consistent with zero. However, even in the left column, the contributions of the subtreshold point sources and the dark emission component are comparable at the bright end.

Even though the inference without the regularization prior can fully explain the GeV excess, it is important to emphasize that this is achieved by point sources with individual significances much below 1σ. This is shown in Figure 3.26, where most of the posterior volume as a function of flux and significance come from low-significance sources.

The resulting posterior of the flux, significance and color histograms consistent with the gamma-ray emission in the inner galaxy are presented in Figure 3.27. Because the pivot energy bin is 0.3 GeV - 1 GeV, the colors belong to the 1 - 3 GeV and 3 - 10 GeV energy bins.

When sample catalogs from the posterior are associated with the 3FGL, the posterior of the parameters associated with the 3FGL point sources can be compared with the 3FGL parameters. Figure 3.28 shows this comparison, where it is observed that the catalog samples and the 3FGL mostly agree, although the two make significantly different predictions for some sources. However, these discrepancies can be due to the fact that the two catalogs are consistent with data sets with different time intervals and reconstruction algorithms.

It is a well-motivated hypothesis that there are subthreshold millisecond pulsars in the inner Milky Way. Probabilistic cataloging of the gamma-ray emission in the Milky Way shows that the GeV excess is consistent with a subthreshold point source origin and that the latter is favored against a diffuse origin when no regularization is applied. However, the latter statement is partially prior-
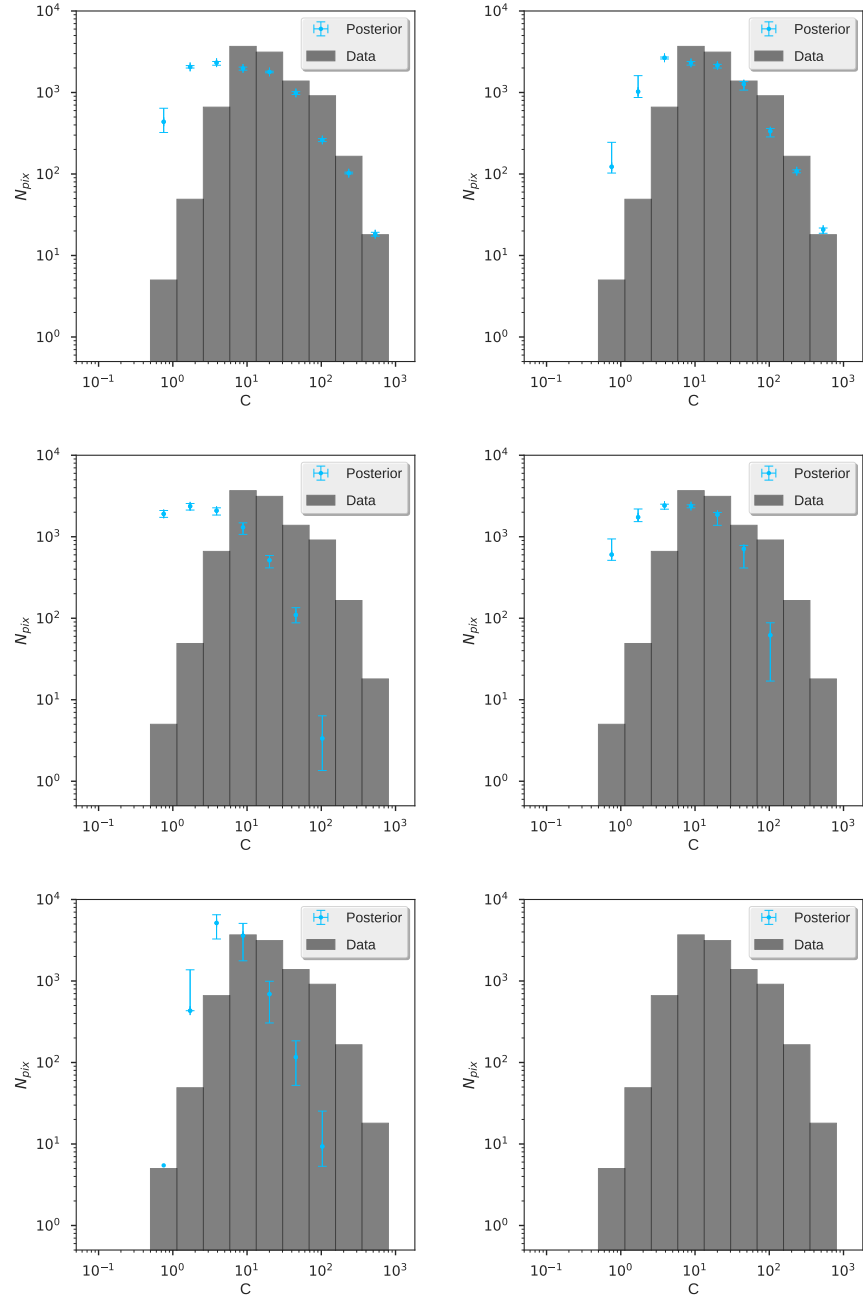
Figure 3.25: Similar to Figure 3.9, but for the posterior consistent with the inner galaxy, showing the 1-point functions for all point sources (top), subthreshold point sources (middle), and the dark emission component (bottom) for two inferences with (left) and without (right) the regularization prior.
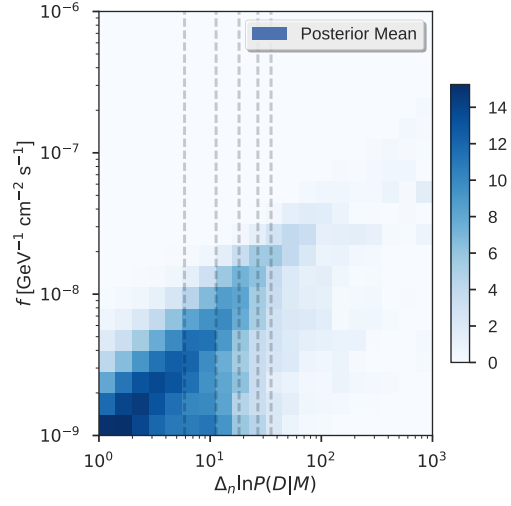
Figure 3.26: Posterior of the histogram of the fluxes and significance of the point sources.

driven, because a compact description of the data with subthreshold point sources above $\sim 1\sigma$

cannot explain the GeV excess better than a dark emission template.
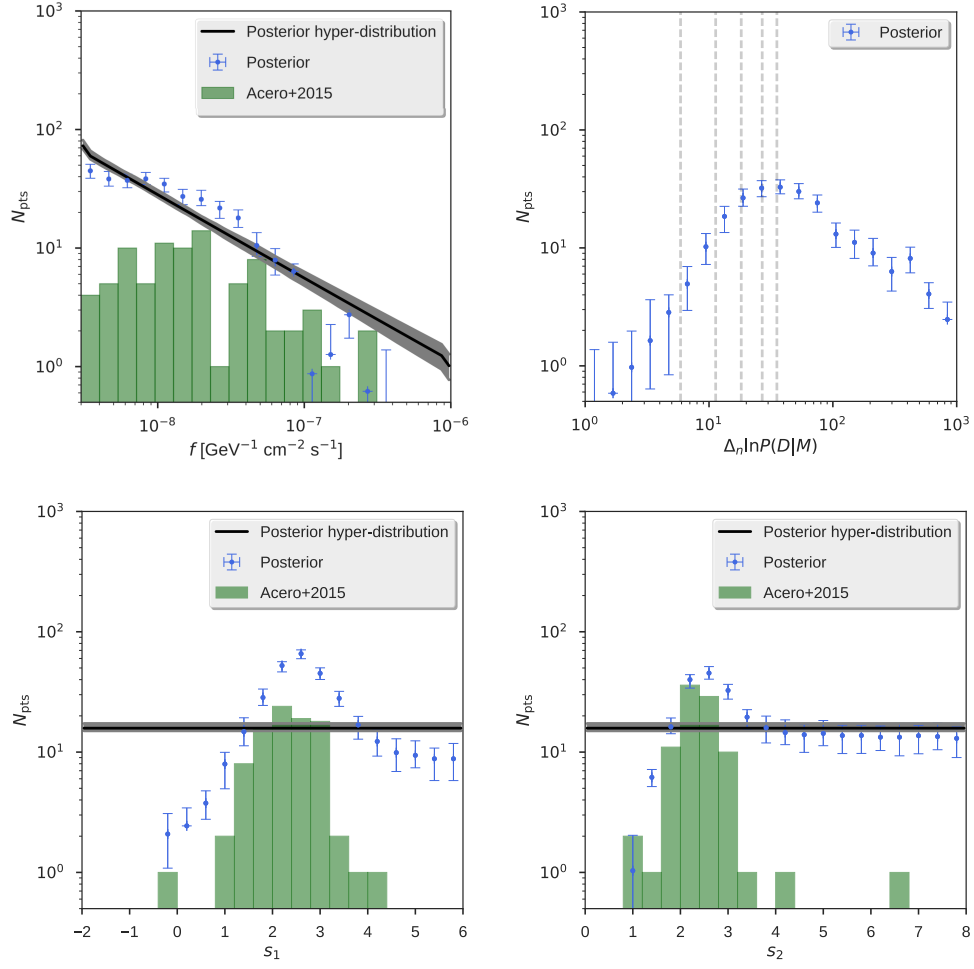
Figure 3.27: Posterior of the flux (top left), significance (top right), and color (bottom) histograms (blue), and the histograms of the associated quantities of the 3FGL point sources (green). The mean (black line) and the 68% credible region (grey shade) of the posterior of the hyperdistribution of the quantities are also shown.
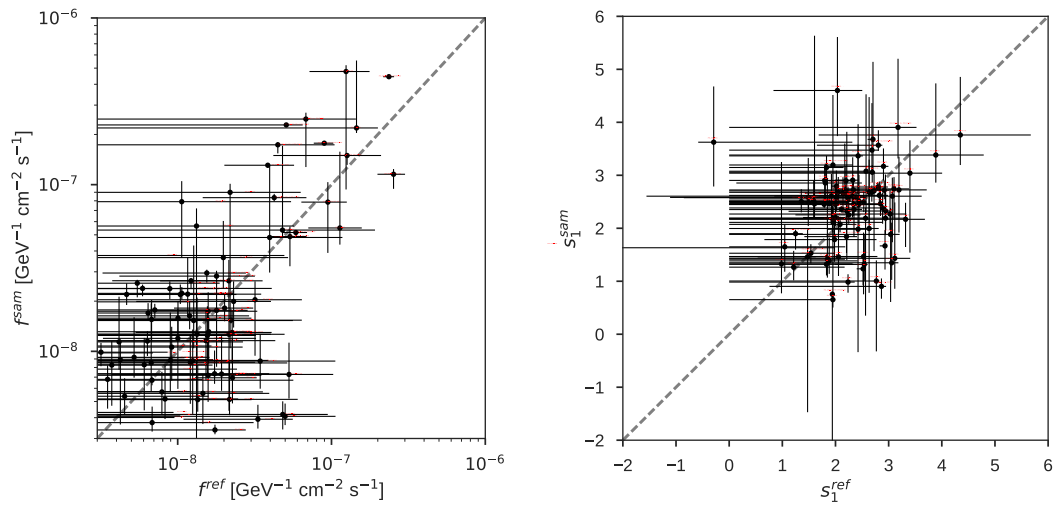
Figure 3.28: Posterior of the flux (left) and the 1 GeV - 3 GeV color (right) of associated to the 3FGL catalog.

4

Gravitationally lensed optical images

Small-scale structure of dark matter can, in principle, be probed by the light emission from faint dwarf galaxies in the Local Group and beyond. However, star formation in these low-mass subhalos is suppressed due to reionization, interaction with their host, and self-quenching[95,181]. Consequently, most of the subhalos with virial halo mass $\lesssim 10^8 M_\odot$ should be dark, making gravitational lensing[85] the only observable probe of their properties. Therefore, gravitational lensing provides a unique

window into the small-scale structure of dark matter in the Universe. In particular, galaxy-scale strong lensing systems in which a background source (e.g. galaxy or quasar) is multiply-imaged by a foreground massive galaxy provide an informative observable to probe the dark matter distribution on subgalactic scales. High-resolution (i.e., with PSF FWHM ~ 0.1 arcsec) optical images of these strong lenses yield a rich data set that can be used to constrain the foreground mass model.

Existence of small-scale structure on the mass plane causes smooth mass models to fail in describing the lensed emission, resulting in anomalies in the flux ratios of multiple images. Observations of these flux-ratio anomalies in strongly-lensed quasars[149,156,53,155,67,68,137,154,91,168,109,165,166,108,128] have been used to study the abundance of substructure within lens galaxies. In particular,[67] used the observation of seven radio-loud lensed quasars to put a constraint on the substructure mass fraction of $0.006 < f_{\mathrm{sub}} < 0.07$ at 90% confidence level. More recently, gravitational imaging of mass substructure[140,223,121] close (in projection) to magnified arcs and Einstein rings has allowed the inference of the small-scale distribution of dark matter. This method fits the observed strong lenses with a smooth mass distribution, i.e. without substructure, and then adds pixel-based corrections[226,227] or a number of linearized subhalos[122] to the mass model with the requirement that the additional degrees of freedom improve the goodness-of-fit above some detection threshold, e.g., greater than $\sim 10\sigma$. For instance,[226,227] detected substructure with masses above $10^8 M_\odot$ at high significance using optical images of two strong lens systems. Similarly, a $10^9 M_\odot$ substructure was detected by[122] in an interferometric data set at a significance of 6.9σ in the strong lens SDP.81[229]. Constraints on the mass fraction in substructure and on the subhalo mass function can, in principle, be placed using these detections as well as nondetections in other lens systems[224,225,145,122,146].

Linking a detection of mass substructure into constraints on the subhalo mass function and their internal properties (e.g., concentration and truncation) is nontrivial. Degeneracies in the lensing model can make the interpretation of a detection ambiguous[159]. For instance, covariances between the source, smooth lens, and substructure models, can lead to biases and underestimated errors in the inferred substructure parameters if such uncertainties are not propagated properly. Furthermore, the nonlinearity of the lensing magnification implies that the likelihood improvement resulting from the addition of a subhalo to the lens model is not independent of the presence of other sub-halos within the lens. In addition, subhalo models also have transdimensional (i.e., across model) covariances since the observed arcs can be fitted with approximately equal likelihoods, using multiple subhalos at various separations and a single more massive subhalo. Properly taking into account these degeneracies is important in obtaining robust uncertainties on subhalo population parameters.

With the exception of several studies[28,120,65,29], which aim to infer subhalos below the detection threshold *statistically*, current inferences[225,145,122] place constraints on the subhalo population *after* detecting (or not detecting) mass substructure within lens galaxies. There are two main problems with this approach. First, by using a reduced form of the data (i.e., position and mass of detected substructure), these inferences cannot fully propagate degeneracies within and across lens models of different complexity. Second, relying on detections causes subthreshold information in lensing data features (e.g., arcs) to be discarded.

In this chapter, the emission metamodel of PCAT is extended to include lensed emission, as well as elements that represent the mass distribution on the lens plane. The resulting metamodel is referred to as the *lensed emission* metamodel, which is the union of lens models with different num-

bers of dark matter subhalos. Therefore, the parameter space of the elements constitute the catalog space of subhalos on the lens plane. The common parameters of the metamodel contain the macro lens model parameters in addition to the parameters describing the PSF and background emission. The observed data is a series of images (i.e., photon counts) of a strong-lens in different energy bands. In principle, the framework can also incorporate other types of auxiliary data such as time delays of the multiple images when analyzing quasar background sources[214], or spectroscopic data of the foreground galaxy to constrain its mass budget[151] as well, by imposing informative priors on the relevant metamodel parameters.

Unlike traditional cataloging of subhalos, probabilistic cataloging generates an ensemble of samples from the joint posterior of the macrolens model and the subhalo catalog space. It also does not require model subhalos to improve the goodness-of-fit above the detection threshold, which reduces mismodeling of the data and recovers more information. Furthermore, it bases the inference on the observed photon count maps directly, as opposed to derived, potentially biased estimates such as inferred astrometric perturbations or relative magnifications of the multiple images of the background source, which have significant non-Gaussian covariances. The joint posterior of the lensed emission metamodel can be marginalized to obtain the marginal posterior of the lens model and subhalo parameters. This allows one to fully propagate within and across model covariances and uncertainties in the observed data to the parameter of interest, e.g., the subhalo mass function.

To illustrate the features of probabilistic cataloging, the focus in this chapter is on simple smooth lens models and source configurations, and results are presented for simulated data sets, which allows one to study potential biases and systematics. Analyses of observed data would require a more

complex lens and source plane model[216]. While this additional complexity can increase the time complexity of inference, the presented framework can be extended to incorporate them.

4.1 Lensed emission metamodel

The lensed image, $\tilde{f}_{\text{src}}$, can be written as the source brightness profile, f_{src}, evaluated on the image plane,

$$\tilde{f}_{\text{src}}(\theta_1, \theta_2) = f_{\text{src}}(\theta_1 - \alpha_{\theta_1}, \theta_2 - \alpha_{\theta_2}), \tag{4.1}$$

where $\vec{\alpha} = (\alpha_{\theta_1}, \alpha_{\theta_2})$ is the deflection vector at the image position, $\vec{\theta} = (\theta_1, \theta_2)$. θ_1 and θ_2 denote the horizontal and vertical coordinates on the two dimensional image plane and $\vec{\tilde{\theta}}$ is the position on the source plane such that

$$\vec{\tilde{\theta}} = \vec{\theta} - \vec{\alpha}(\vec{\theta}), \tag{4.2}$$

which is the lens equation. Furthermore, the deflection field $\vec{\alpha}$ can be written as the gradient of the Newtonian gravitational potential integrated along the line-of-sight,

$$\vec{\alpha} = \frac{2}{c^2} \vec{\nabla}_\theta \int d\chi \frac{\chi_{\text{src}} - \chi}{\chi \chi_{\text{src}}} \Psi(\vec{r}), \tag{4.3}$$

where χ is the comoving distance along the line-of-sight, χ_{src} is the comoving distance to the source and $\Psi(\vec{r})$ is the gravitational potential at the three dimensional position vector, $\vec{r}$.

Because the lensing halo is much smaller than other distance scales in the problem, in the thin-lens approximation, this Equation (4.3) reduces to

$$\vec{\alpha} = \frac{2}{c^2}\frac{D_{\mathrm{LS}}}{D_{\mathrm{S}}D_{\mathrm{L}}}\vec{\nabla}_\theta \int dz \Psi(\vec{r}),\tag{4.4}$$

where $t\vec{heta}$ is the angle subtended by the halocentric three-dimensional vector $\vec{r}$ at a line-of-sight distance of z, D_{L}, D_{S}, and D_{LS} are the angular diameter distances from the observer to the lens and sources planes, and from the lens to the source plane, respectively. Analytic expressions for the line-of-sight integral of various foreground mass components can then be used to calculate the deflection field[136].

4.1.1 Deflection field

Because of the linearity of the Poisson equation, the deflection field, $\vec{\alpha}$, can be expressed as the superposition of deflections due to several components. In a typical strong lens system, the dominant contribution comes from the smooth mass distribution of the main foreground galaxy (hereafter the host halo), $\vec{\alpha}_{\mathrm{hst}}$, and the external shear due to the large-scale structure, $\vec{\alpha}_{\mathrm{ext}}$. Subhalos perturb this deflection field at the percent level with individual contributions to the deflection field, $\vec{\alpha}_n$, where the subscript $n = 0, 1, ..., N-1$ is the index of a subhalo. As a result, the total deflection field is

given by

$$\vec{\alpha} = \vec{\alpha}_{\text{ext}} + \vec{\alpha}_{\text{hst}} + \sum_{n=0}^{N-1} \vec{\alpha}_n. \tag{4.5}$$

SMOOTH HALO   Cosmological N-body simulations of ΛCDM [105] suggest that the equilibrium distribution of collisionless dark matter particles is approximately spherically symmetric about the dynamical center of mass of the self-gravitating halo of dark matter particles. Furthermore, the resulting virialized halos have a universal radial profile that is described by the NFW profile, i.e., a broken power-law in the radial halo-centric distance, r. However, the host galaxy also possesses baryonic matter, which dominates the mass budget in the bulge. Taking the mass budget of the host halo to be the sum of the baryonic and dark matter components, it has been found that the mass density profile of Early Type Galaxies (ETGs) can be fitted by Singular Isothermal Ellipsoids (SIEs), which is known as the bulge-halo conspiracy [220] since neither dark nor baryonic matter density distribution is individually isothermal. As a result, the mass density profile of the host halo is taken as

$$\rho_{\text{hst}}(r) = \frac{\sigma_{\text{hst}}^2}{2\pi G r^2}, \tag{4.6}$$

where σ_{hst} is the (constant) velocity dispersion and G is the gravitational constant. In general, galaxies also possess some ellipticity due to the existence of a baryonic disc or as artifacts of recent gravitational interactions. By taking into account the ellipticity of the isothermal gas, and assuming that the contours of equal surface mass density are in the form of concentric ellipses, the deflection is

given by[136,134],

$$\vec{\alpha}_{\text{hst}} = \frac{\theta_{\text{E,hst}} q_{\text{hst}}}{\sqrt{1 - q_{\text{hst}}^2}} \left(\tan^{-1}\left(\frac{\sqrt{1 - q_{\text{hst}}^2} \theta'_{1,\text{hst}}}{\omega}\right) \hat{\theta}'_{1,\text{hst}} + \tanh^{-1}\left(\frac{\sqrt{1 - q_{\text{hst}}^2} \theta'_{2,\text{hst}}}{\omega}\right) \hat{\theta}'_{2,\text{hst}} \right), \quad (4.7)$$

where $\omega \equiv \sqrt{q_{\text{hst}}^2 \theta'^2_1 + \theta'^2_2}$, $q_{\text{hst}} \equiv 1 - \epsilon_{\text{hst}}$ is the minor to major axis ratio, and ϵ_{hst} is the ellipticity. The primed quantities, i.e, $\theta'_{1,\text{hst}}$, $\theta'_{2,\text{hst}}$, $\hat{\theta}'_{1,\text{hst}}$, and $\hat{\theta}'_{2,\text{hst}}$, denote the angular coordinates and the associated unit vectors along the minor and major axes of the elliptical host halo, respectively. If one uses ϕ_{hst} to denote the angle between the major axis of the ellipse and the horizontal axis of the image, θ_1, then the primed coordinates are obtained by translating the image by $(\theta_{1,\text{hst}}, \theta_{2,\text{hst}})$ and then rotating it by ϕ_{hst}.

Furthermore, the lensing strength of the isothermal deflector is parametrized using the projected Einstein radius, $\theta_{E,\text{hst}}$, i.e., the radius of the circular image that would be produced if the lens were spherical and aligned with the background light source. In general, $\theta_{E,\text{hst}}$ can be defined as the radius of a circle that has an area equal to the area inside the critical curve, which holds true even for non-spherical lenses.

Note that many observed strong lenses require departures from an isothermal model, where the preferred inner log-slope of the three dimensional density distribution can be significantly different from -2. However, since the presented method is independent of the details of the lens model, a relatively simple mass model is assumed for the smooth halo.

Subhalos ΛCDM predicts that, with cosmological time, more density peaks collapse to form new self-gravitating halos, while halos that formed earlier are accreted into more massive ones. Because hierarchical structure formation is a continuous and ongoing process, at any given time some fraction of matter is expected to be tied in halos that have been accreted into more massive halos. These dark, gravitationally bound satellites inside the virial extent of the host, are referred to as *subhalos*.

Subhalos grow by accretion and lose mass via tidal stripping, collisions and evaporation. The relative efficiency of these processes determines the level of substructure in a host halo of mass M_{hst} at redshift z_{hst}. N-body simulations can predict the abundance and properties of these subhalos, although the finite mass resolution of simulations can preclude robust conclusions[222]. Nevertheless, it is expected that subhalos have a truncated mass profile due to tidal stripping by the host galaxy. Therefore, the three dimensional subhalo density profile of the nth subhalo, $\rho_n(r)$, is taken to be the NFW profile with a power-law truncation as[22]:

$$\rho_n(r) = \frac{M_{0,n}}{4\pi r_{\text{s},n}^3} \frac{1}{(r/r_{\text{s},n})(1 + r/r_{\text{s},n})^2} \frac{\tau_n^2}{(r/r_{\text{s},n})^2 + \tau_n^2}, \tag{4.8}$$

$$M_{\text{c},n} = M_{0,n} \frac{\tau_n^2}{(\tau_n^2 + 1)^2} \left[(\tau_n^2 - 1) \ln \tau_n + \pi \tau_n - (1 + \tau_n^2) \right]. \tag{4.9}$$

where $M_{\text{c},n}$ is the truncated mass of the subhalo. In this parametrization, most of the mass of a subhalo is contained inside the scale radius, $r_{\text{s},n}$, which would be the subhalo-centric distance at which the density profile steepens from a power-law with index -1 to a power-law with index -3 if it were not tidally truncated. Moreover, the central mass concentration is encoded in $\tau_n = r_{\text{c},n}/r_{\text{s},n}$ similar to the concentration parameter $c = r_{\text{vir},n}/r_{\text{s},n}$ in the non-truncated case, where $r_{\text{vir},n}$, $r_{\text{c},n}$ and

$r_{\mathrm{s},n}$ are the virial, tidal cutoff, and scale radii of a subhalo. $M_{0,n}$ is a mass scale relevant to the nth subhalo. Subhalos with large τ_n represent old, highly concentrated subhalos that orbit the host at large halocentric distances. In contrast, those with low τ_n typically correspond to subhalos that are close to the center and are being tidally stripped by the host. The subscript n is used to refer to a parameter of the nth subhalo when the host halo has the same type of parameter.

It is also expected that subhalos near the dynamical center of the host halo, and closest to disruption, should be tidally elongated. Nevertheless, because current data sets do not have the signal-to-noise necessary to constrain departures from spherical symmetry, any triaxiality in the subhalo shapes is neglected. Furthermore, because virialized halos are generally found to have concentrations that reflect the average background density of the Universe at the time of their collapse, earlier halos are expected to have higher concentrations. Upon accretion into the host halo, their high concentrations can help them remain intact despite strong tidal stretching and evaporation by the host halo. Motivated by recent N-body simulations[209], substructure inside subhalos is neglected and it is assumed that the subhalos of the host galaxy have smooth mass distributions.

Although the mass of a non-truncated NFW profile is logarithmically divergent with halocentric radius, tidal truncation makes the total mass $M_{\mathrm{c},n}$ finite[22] such that

$$M_{\mathrm{c},n} = M_{0,n} \frac{\tau_n^2}{(\tau_n^2 + 1)^2} \left[(\tau_n^2 - 1) \ln \tau_n + \pi \tau_n - (1 + \tau_n^2) \right] . \qquad (4.10)$$

This finite mass is plotted in Figure 4.1 for a subhalo in units of $M_{0,n}$. It changes by about an order of magnitude in the relevant region of $1 \lesssim \tau_n \lesssim 10$. Below $\tau_n \sim 1$, the subhalo is likely

to be completely disrupted by evaporation and tidal stripping. Truncated masses are reported when referring to subhalos.

The simulated images are approximately centered at the dynamical center of the host halo, and have a size comparable to its Einstein radius, which is much smaller than its virial radius. When projected along the line-of-sight, subhalos in the field of view have a large spread in their three dimensional distance from the dynamical center of the host. As a result, there should be a large spread in their deflection profile. This is taken into account approximately by drawing the projected scale and cutoff radii of the simulated subhalos from a uniform distribution. The limits of these distributions are given in Table 4.1. The resulting ratios between the projected cutoff and scale radii, τ_n, fall between 3 and 25.

It is important to emphasize that these simulated subhalos provide a test bed for introducing probabilistic cataloging of subhalos and do not fully reflect their rich dynamics. Furthermore, although simulations and analytical treatments of ΛCDM can predict the scale and cutoff radii of a given subhalo given its mass and distance from the center of the host galaxy[156], implying that predictions can be folded into the inference as a joint prior distribution on the projected scale and cutoff radii, the fact that strong lens images are not informative about the three dimensional distance of a subhalo from the center of its host means there is a degeneracy in this description.

The deflection field due to the nth subhalo, α_n, is parametrized using the angular distances corresponding to the scale and cutoff radii, $\theta_{s,n} = r_{s,n}/D_L$ and $\theta_{c,n} = r_{c,n}/D_L$, and the normalization of the deflection profile, $\alpha_{s,n}$, which is referred to as the deflection strength.

When projected onto the lens plane, the deflection field due to an untruncated, spherically sym-
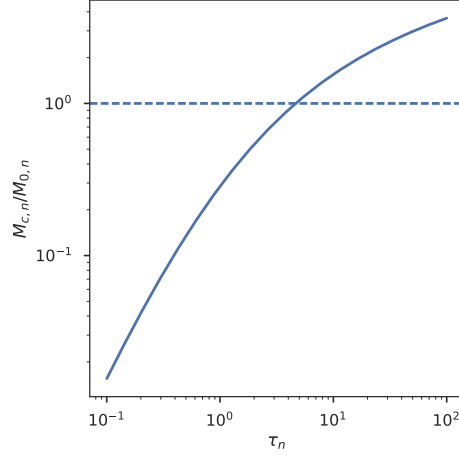
Figure 4.1: Truncated mass of a subhalo, $M_{c,n}$, normalized by $M_{0,n}$ as a function of the ratio between its cutoff and scale radii, τ_n.

metric NFW subhalo becomes azimuthally symmetric and can be described by a radial profile[110]

$$\alpha_n \equiv \theta \frac{\bar{\Sigma}_n(\theta)}{\Sigma_{\text{crit}}} = \frac{M_{0,n}}{\pi D_{\text{L}}^2 \Sigma_{\text{crit}} \theta_{\text{s},n}} \frac{1}{\theta'} \left(\ln \frac{\theta'}{2} + F(\theta') \right)$$
$$= \frac{\alpha_{\text{s},n}}{\theta'} \left(\ln \frac{\theta'}{2} + F(\theta') \right),$$
(4.11)

where $\bar{\Sigma}_n(\theta)$ is the mean surface mass density of the nth subhalo inside a subhalo-centric circle of

radius θ, and

$$\Sigma_{\text{crit}} = \frac{c^2}{4\pi G} \frac{D_{\text{S}}}{D_{\text{LS}} D_{\text{L}}}$$
(4.12)

is the critical surface mass density on the lens plane, which has a value of $3.1 \times 10^9 M_\odot \text{ kpc}^{-2}$ for

the simulated data. For notational brevity, the rescaled angular distance, $\theta' \equiv \theta/\theta_{\text{s},n}$, is used. The

function $F(\theta')$ is given by

$$
F(\theta') = \begin{cases} \dfrac{\arccos(1/\theta')}{\sqrt{\theta'^2 - 1}} & \text{if } \theta' > 1 \\[4mm] 1 & \text{if } \theta' = 1 \\[4mm] \dfrac{\operatorname{arccosh}(1/\theta')}{\sqrt{1 - \theta'^2}} & \text{if } \theta' < 1. \end{cases} \qquad (4.13)
$$

When the truncation is taken into account, the deflection profile becomes[22]

$$
\alpha_n = \frac{\alpha_{\mathrm{s},n}}{\theta'} \frac{\tau_n^2}{(\tau_n^2 + 1)^2} \Bigg(\Big(\tau_n^2 + 1 + 2(\theta'^2 - 1) \Big) F(\theta') + (\tau_n^2 - 1) \ln \tau_n
$$
$$
+ \pi \tau_n + \sqrt{\tau_n^2 + \theta'^2} \Big(\frac{\tau_n^2 - 1}{\tau_n} L(\theta') - \pi \Big) \Bigg), \quad (4.14)
$$

where

$$
L(\theta') = \ln \left(\frac{\theta'}{\tau_n + \sqrt{\tau_n^2 + \theta'^2}} \right). \qquad (4.15)
$$

In the limit of large $\tau_n = \theta_{\mathrm{c},n}/\theta_{\mathrm{s},n}$, this expression reduces to Equation (4.11). Hence, the three parameters used to describe the resulting deflection profile of the nth subhalo are the normalization $\alpha_{\mathrm{s,n}}$, projected scale and cutoff radii $\theta_{\mathrm{s},n}$, $\theta_{\mathrm{c},n}$.

The deflection as a function of radial distance in three dimensions from the dynamical center of mass of the subhalo is proportional to the integrated mass up to that radius. Therefore, it initially rises because of the shallow log-slope of the NFW profile, turns over at the scale radius and further
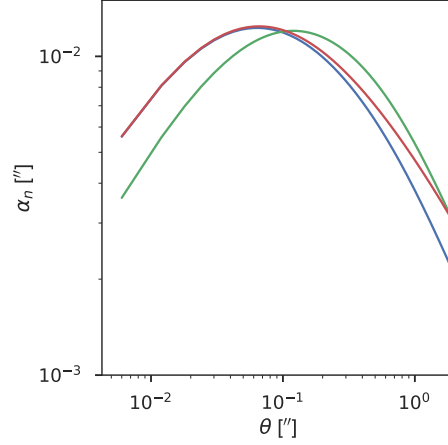
Figure 4.2: Deflection as a function of angular distance from a given subhalo in arcsec. All three profiles have a subhalo deflection strength of $\alpha_{s,n} = 0.1$ arcsec. The blue, green and red profiles correspond to subhalos with ($\theta_{s,n} = 0.05$ arcsec, $\theta_{c,n} = 1$ arcsec), ($\theta_{s,n} = 0.2$ arcsec, $\theta_{c,n} = 1$ arcsec) and ($\theta_{s,n} = 0.05$ arcsec, $\theta_{c,n} = 2$ arcsec), respectively.

steepens beyond the cutoff radius. The projected deflection profile of a subhalo is given in Figure 4.2.

ΛCDM predicts that the variance of the matter density fluctuations in the Universe monotonically increases as one averages the density field inside successively smaller spheres. Therefore, the number of collapsed halos (and hence of subhalos) increases steeply as a power-law toward low masses with a log-slope of ~ -1.9[209]. When ΛCDM is modified at small scales via self-interactions, free-streaming at kinetic decoupling or by the Heisenberg uncertainty in the case of wave-like dark matter[129], this is no longer the case and one expects a mass scale below which the subhalo mass function is suppressed.

Since subhalo deflection strengths are proportional to mass, it is assumed that $\alpha_{s,n}$ are also distributed as a power-law between some minimum and maximum value. However, the lower limit of

this prior is not motivated by a physical cutoff in the mass distribution. It is rather a computational convenience, since probabilistic cataloging of very low mass subhalos result in little information gain at the expense of significantly increasing computational complexity. Furthermore, the pixel size and the PSF of the photon collecting device sets an angular distance scale below which deflections cannot significantly affect the image. This scale also depends on the level of background emission and the flux of the lensed light source. For the simulated images of the Wide Field Camera 3 / Ultraviolet Visible (WFC3/UVIS) camera on board the HST, whose pixel size is 0.04 arcsec, this scale roughly corresponds to $\sim$ 0.01 arcsec. Therefore, the minimum allowed value of $\alpha_{s,n}$ is chosen as 0.01 arcsec, in order to allow low-significance subhalos into the posterior. The maximum of $\alpha_{s,n}$ is chosen as 1 arcsec. However, this maximum value does not significantly affect the posterior, since most of the prior volume is contained at small $\alpha_{s,n}$ due to the steepness of the power-law. The normalization of this power-law is then given by μ_{sub}, i.e., the Poisson mean of the number of subhalos.

Because of the hierarchical prior in probabilistic cataloging, the negative of the log-slope of the $\alpha_{s,n}$ distribution, β, is also allowed to vary. Therefore, the sampler visits all prior configurations consistent with the data and generates the posterior of β. Although β parametrizes the prior, it is treated similar to other parameters in the model, and is referred to as a *hyperparameter*. It is the prior on this *hyperparameter* that specifies the prior belief on the distribution of $\alpha_{s,n}$. This prior is set such that β itself is Gaussian distributed with a mean and standard deviation of 1.9 and 0.5, respectively. The mean reflects the value motivated by ΛCDM, whereas the spread accounts for any deviations and allows the sampler to visit subhalo configurations that deviate from a power-law with index 1.9.
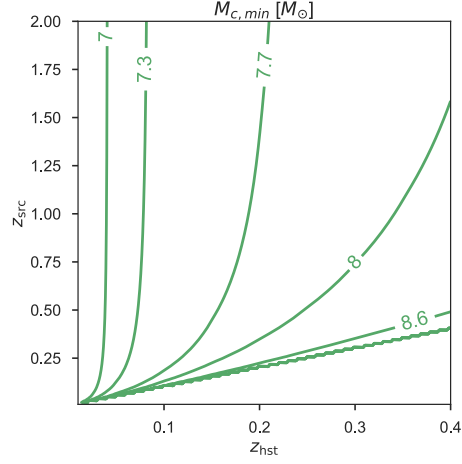
Figure 4.3: Contours of minimum truncated subhalo mass allowed in the metamodel, i.e., lower limit of the truncated subhalo mass prior, as a function of host and source galaxy redshifts, assuming $\theta_{s,n} = 0.1$ arcsec and $\theta_{c,n} = 1$ arcsec.

The deflection strength, $\alpha_{s,n}$, projected scale and cutoff radii, $\theta_{s,n}$ and $\theta_{c,n}$, of a subhalo uniquely determine its truncated mass, $M_{c,n}$. Therefore, given the above priors, there is no hard boundary on the prior distribution of the subhalo truncated masses. However, most of the prior volume falls in the interval $10^8 - 10^9 M_\odot$ for the host and source redshifts of $z_{hst} = 0.2$ and $z_{src} = 1$ simulated in this chapter. This interval changes as a function of angular diameter distances to the host and source planes. To put the prior subhalo mass interval into cosmological context, Figure 4.3 shows the lower bound of the prior on the truncated mass of a subhalo, if it had $\theta_{s,n} = 0.05$ arcsec and $\theta_{c,n} = 1$ arcsec.

Last, a spatially uniform prior is adopted on the projected surface density of subhalos. The prior structure of the lensed emission metamodel is summarized in Tables 4.1 and 4.2.

EXTERNAL SHEAR   Lensing galaxies are usually found in galaxy groups or clusters, whose overall contribution adds angular structure to the deflection field in the vicinity of the lensing galaxy. Even in the case of an isolated galaxy, matter distribution in the foreground of the lensing galaxy can introduce additional multipole terms to the deflection field. This external deflection is parametrized by a reduced shear field γ_{ext} oriented at an angle ϕ_{ext} with respect to the longitudinal axis[135]. The field is traceless and invariant under a rotation of $180°$.

$$\vec{\alpha}_{\mathrm{ext}} = \gamma_{\mathrm{ext}} \begin{pmatrix} \cos 2\phi_{\mathrm{ext}} & +\sin 2\phi_{\mathrm{ext}} \\ \sin 2\phi_{\mathrm{ext}} & -\cos 2\phi_{\mathrm{ext}} \end{pmatrix} \vec{\theta}. \tag{4.16}$$

In this formalism, only the reduced shear, $\gamma_{\mathrm{ext}} = \gamma'_{\mathrm{ext}}/(1 - \kappa_{\mathrm{ext}})$, is modeled, where κ_{ext} represents a potentially underlying, spatially uniform convergence in the field of interest. This is because adding a constant mass sheet to the model, $(1 - \kappa_{\mathrm{ext}})$, reproduces the same lensed image by simultaneously rescaling the source position and flux or the deflector mass[92]. These degenerate mass models are linked by the transformations

$$(1 - \kappa_{\mathrm{ext}}) \rightarrow \lambda\kappa_{\mathrm{ext}},$$
$$\gamma_{\mathrm{ext}} \rightarrow \lambda\gamma_{\mathrm{ext}}. \tag{4.17}$$

An independent measurement of the deflector mass, e.g., using stellar kinematics, could constrain λ[197,39]. Alternatively, because the magnification field and light travel time change with λ, knowledge

of the non-lensed fluxes of the background sources or having multiple images of the same system taken at different times could break this degeneracy. Specifically, the knowledge of time delays between multiple images of a background quasar and the time-delay distance allows one to constrain the gravitational potential of the main deflector independent of the astrometric information. Because only a single photometric exposure of the strong lens system is used, all convergences are determined only up to the transformation in Equation 4.17 and no attempt is made to lift the degeneracy.

The inclusion of only a quadrupolar shear field is motivated by the assumption that the lensing galaxy is isolated and gravitationally relaxed. If there are nearby galaxies, higher order multipoles would be needed.

4.1.2 Light emission from the source

In general, the background light source in a strong lens system could be a quasar, a blue, young galaxy with bright Lyman-α emission, or a red ETG similar to the lensing galaxy considered in this thesis. When the background light source is a quasar, i.e., a point source without any spatial feature, the resulting multiple images are not extended and can be used to probe subhalos only over a small region on the lens plane depending on the mass of the subhalo. Therefore, galaxy-galaxy lenses are considered instead, where light from a background galaxy is strongly deflected by a giant elliptical in the foreground. In these systems multiple images are more extended and a higher fraction of the pixels are informative on the subhalo parameters.

It is assumed that the emission from both the foreground and the background galaxies have Sérsic

profiles, i.e., the surface brightness is given by[199]

$$f_{\text{gal}}(\theta) = f_{\text{e,gal}} \exp\left(-b_n\left((\theta/\theta_{\text{e,gal}})^{1/n} - 1\right)\right), \qquad (4.18)$$

where the subscript gal refers to both lensing (foreground) and lensed (background) galaxies, i.e., gal={hst, src}. In this relation b_n is a coefficient that depends on the index n[55],

$$b_n = 2n - \frac{1}{3} + \frac{4}{405n} + \frac{46}{25515n^2} \qquad (4.19)$$

which controls the level at which the inner and outer slopes anti-correlate. Furthermore, $\theta_{\text{e,gal}}$ is the projected distance within which half of the total emission is contained and $f_{\text{e,gal}}$ is the surface brightness at this radius. It is assumed that $n = 4$, which corresponds to the de Vaucouleurs profile[77]. For this profile, the surface brightness at the half-light radius, $f_{\text{e,gal}}$, is related to the flux of the galaxy, I_{gal}, by $I_{\text{gal}} \approx 7.2\pi\theta_{\text{e,gal}}^2 f_{\text{e,gal}}$.

Modeling the source plane emission with the surface brightness of a single galaxy is a simplified description of observed lenses. In particular, emission or absorption regions in the source galaxy can significantly change the appearance of the resulting arcs and arclets. However, the main purpose is to construct a simple test bed to compare probabilistic cataloging to mainstream analysis tools of strong lensing.

4.1.3   Light emission from the foreground galaxy

The same surface brightness profile in Equation 4.18 is used to represent emission from the host galaxy, but with different half-light radius and surface brightness, $\theta_{e,hst}$, and $f_{e,hst}$.

Because the scattering cross section between dark and baryonic matter is not necessarily zero, in principle, there can be a spatial offset between the surface brightness and mass density of a host galaxy that has not been gravitationally relaxed[203]. Such unrelaxed systems are not modeled and it is assumed that the smooth mass distribution and the light emission of the host galaxy are cocentric.

Furthermore, an estimate of foreground galaxy or isotropic emission is not subtracted from the observed photon count map. This is because subtracting a best-fit emission component from the photon count map is equivalent to placing a delta function prior at the associated value, or equivalently, assuming that the parameter has vanishing covariance with all other parameters. Instead, samples are taken from the joint posterior of all parameters including those of the host galaxy and the nuisance parameters are marginalized out.

It is assumed that the background emission is spatially uniform. Therefore, a single isotropic emission component is allowed in the lensed emission metamodel.

4.1.4   Instrumental PSF

In the absence of an intervening atmosphere and in the limit of distortion-free optics, the PSF of the HST is limited by diffraction, although it has diffraction spikes due to the support vanes of the secondary mirror. Given the circular aperture of the HST, the PSF in band i can be modeled as an

Airy pattern whose first zero-crossing occurs at a radius of $\sigma_{\text{psf},i}$.

In general, uninformative priors can be placed on the parameters that characterize the PSF shape, e.g., $\sigma_{\text{psf},i}$. This would allow PSF parameters to be inferred along with the lensed emission meta-model. However, in the absence of a bright point source in the image, the PSF cannot be constrained. Therefore, Gaussian priors are adopted on the PSF parameters, whose means and standard deviations are provided by external calibration. This allows propagation of uncertainties due to potential covariances between the PSF and model parameters.

4.1.5 MODEL IMAGE

Given the above ingredients, the model image prior to the PSF convolution can be expressed as the sum of

- isotropic (i.e., spatially uniform) emission component to model detector background and isotropic sky emission, f_{bac},
- emission from the host galaxy, f_{hst},
- gravitationally lensed emission of a background light source due to a foreground host galaxy and a variable number of its subhalos, $\tilde{f}_{\text{src}}$.

The total model emission, f_{m}, is then obtained by convolving this image with the PSF, $\mathcal{F}$, such that

$$f_{\text{m}} = \mathcal{F} * (f_{\text{bac}} + f_{\text{hst}} + \tilde{f}_{\text{src}}). \tag{4.20}$$

Tables 4.1 and 4.2 list the metamodel parameters. The row group at the top contains the hyper-parameters, which parametrize the conditional distribution of subhalo properties. These subhalo parameters are shown in the row group at the bottom of the list. The remaining row groups are the PSF, background, and lens parameters from top to bottom. The lens parameters are further divided into three subgroups, separately showing the foreground host lens, background light source, and external shear parameters.

The third and fourth columns indicate the minimum and maximum when the associated distribution is uniform, log-uniform or power-law, and show the mean and standard deviation when the distribution is Gaussian. The fifth column indicates the parameters of the metamodel from which the simulated data sets are drawn. Simulated data generation will be discussed in Section 4.2.1.

A probabilistic graphical model of the lensed emission metamodel PCAT is presented in Figure 4.4, which uses the same notation as Figure 2.9.

4.2 INFERENCE ON THE LENS PLANE

In order to validate the inference performance of PCAT, it is run on simulated data. These simulated count maps are generated as Poisson realizations of the photon count maps obtained by forward-modeling a generative (*true*) metamodel. Since the parameters of the true metamodel are known, the posterior can be compared with the underlying true parameters. This approach yields full control over systematic errors, by allowing us, for instance, to fit a data set that has been drawn from a different metamodel.

name	prior	min/mean	max/std	true	unit
μ_{sub}	uniform	0	100	-	
β	Gaussian	1.9	0.5	1.9	
σ_{psf}	Gaussian	0.087	0.01	0.087	arcsec
A	log-uniform	10^{-8}	10^{-6}	2×10^{-7}	erg/cm^2/s/Å/sr
$\theta_{1,\text{src}}$	uniform	-2	2	RC	arcsec
$\theta_{2,\text{src}}$	uniform	-2	2	RC	arcsec
I_{src}	log-uniform	10^{-20}	10^{-15}	10^{-18}	erg/cm^2/s/Å
$\theta_{e,\text{src}}$	log-uniform	0.1	2	0.5	arcsec
ϵ_{src}	uniform	0	0.3	R	
ϕ_{src}	uniform	0	2π	R	radian
$\theta_{1,\text{hst}}$	uniform	-2	2	RC	arcsec
$\theta_{2,\text{hst}}$	uniform	-2	2	RC	arcsec
I_{hst}	log-uniform	10^{-20}	10^{-15}	10^{-16}	erg/cm^2/s/Å
$\theta_{e,\text{hst}}$	log-uniform	0.1	2	1	arcsec
$\theta_{E,\text{hst}}$	log-uniform	0.5	2	1.5	arcsec
ϵ_{hst}	uniform	0	0.5	R	
ϕ_{hst}	uniform	0	2π	R	radian
γ_{ext}	uniform	0	0.3	R	
ϕ_{ext}	uniform	0	2π	R	radian
N_{sub}	Poisson	μ_{sub}	$\sqrt{\mu_{\text{sub}}}$	25	
$\vec{\theta}_1$	uniform	-2	2	R	arcsec
$\vec{\theta}_2$	uniform	-2	2	R	arcsec
$\vec{\alpha}_{\text{s}}$	power-law	0.01	1	R	arcsec
$\vec{\theta}_{\text{s}}$	uniform	0	0.1	R	arcsec
$\vec{\theta}_{\text{c}}$	uniform	0	2	R	arcsec

Table 4.1: The parameter list of the PCAT lensed emission metamodel. The letter R under the "true" column implies that the associated parameter is randomly sampled from the prior. The letters RC, on the other hand, mean that the parameter is drawn from a Gaussian with mean 0 and standard deviation 0.04 arcsec.

name	explanation
μ_{sub}	mean number of subhalos
β	slope of the deflection strength distribution
σ_{psf}	radius of the Airy disk that represents the PSF
A	normalization of the isotropic emission
$\theta_{1,\mathrm{src}}$	horizontal coordinate of the background source
$\theta_{2,\mathrm{src}}$	vertical coordinate of the background source
I_{src}	flux of the background source
$\theta_{e,\mathrm{src}}$	half-light radius of the background source
ϵ_{src}	ellipticity of the background source
ϕ_{src}	azimuthal orientation of the background source
$\theta_{1,\mathrm{hst}}$	horizontal coordinate of the host galaxy
$\theta_{2,\mathrm{hst}}$	vertical coordinate of the host galaxy
I_{hst}	flux of the host galaxy
$\theta_{e,\mathrm{hst}}$	scale size of the Sérsic profile of the host galaxy
$\theta_{\mathrm{E,hst}}$	Einstein radius of the host galaxy
ϵ_{hst}	ellipticity of the host galaxy
ϕ_{hst}	azimuthal angle of the ellipticity of the host galaxy
γ_{ext}	amplitude of the external shear
ϕ_{ext}	azimuthal angle of the external shear
N_{sub}	number of subhalos in a model
$\vec{\theta_1}$	horizontal coordinate of the nth subhalo
$\vec{\theta_2}$	vertical coordinate of the nth subhalo
$\vec{\alpha_{\mathrm{s}}}$	deflection strength of the nth subhalo
$\vec{\theta_{\mathrm{s}}}$	projected scale radius of the nth subhalo
$\vec{\theta_{\mathrm{c}}}$	projected cutoff radius of the nth subhalo

Table 4.2: The parameter list of the PCAT lensed emission metamodel, indicating the meaning of the parameter.
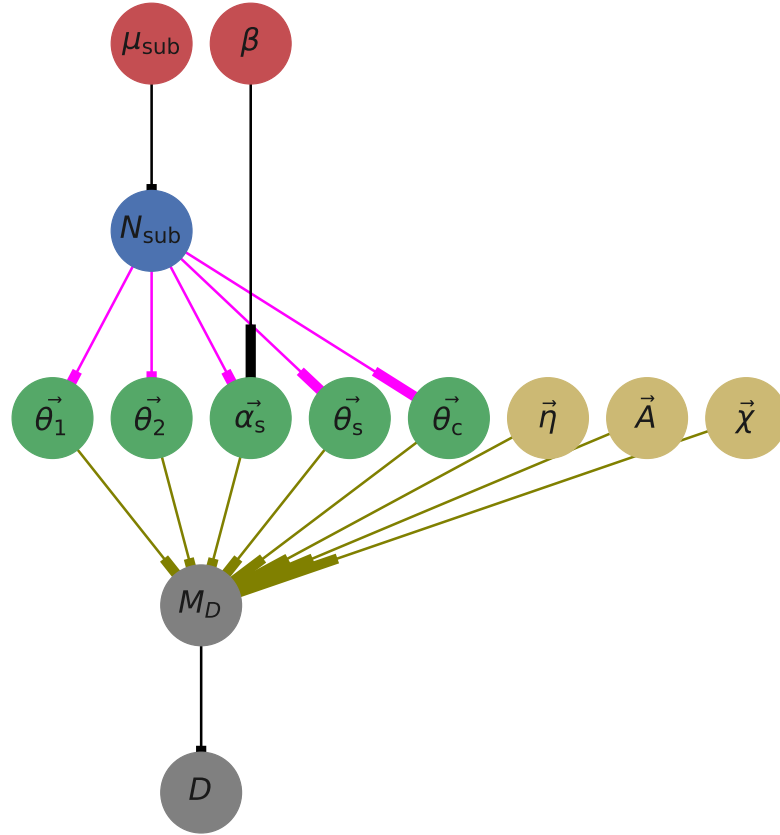
Figure 4.4: The probabilistic graphical model of PCAT lensed emission metamodel. The notation is the same as that in Figure 2.9. The lensed emission metamodel has massive subhalos as elements and another class of common parameters, χ, to model the macro lens.

4.2.1 Simulated data

When simulating an image, it is assumed that the data is taken using the WFC3/UVIS detector on the HST. This choice is intended to provide an example, since probabilistic cataloging can be applied to any photometric data set such as ground-based high-resolution images that use adaptive optics. It is further assumed that the F814W filter is used to collect the photons. For this HST band, the forward-modeled surface brightness, i.e., in units of $\mathrm{erg\,cm^{-2}\,s^{-1}\,\mathring{A}^{-1}\,sr^{-1}}$, is multiplied by $6.1 \times 10^{18}\ \mathrm{erg^{-1}\,cm^2\,s\,\mathring{A}}$[188], as well as the pixel area in order to obtain the forward-modeled photon counts per pixel. It is assumed that the photometric data is collected in a single exposure and that no drizzling is applied. Furthermore, the observation time is fixed at 1000 seconds, which is roughly equivalent to a half orbit of HST. Choosing the simulated galaxy brightnesses low enough allows one to ensure that the CCD does not saturate over the selected exposure time. Finally, defining signal-to-noise as the ratio of the lensed surface brightness with the square root of the total surface brightness, the nominal simulated data set has a maximum per-pixel signal-to-noise of 9, which quickly decays away from the multiple images of the background source.

The parameters of the true metamodel are sampled randomly from the prior, unless stated otherwise in Table 4.1. In particular, the minimum of the true distribution of $\alpha_{s,n}$ is chosen to be 0.003 arcsec, and the number of true subhalos is fixed at 25. These two choices are made so that the resulting true subhalo mass fraction averaged within 0.1 arcsec of the critical curve is $\sim 2\%$[101]. Furthermore, deterministic values are assigned to some parameters as indicated under the *true* column of the table. The letter combination RC under this column indicates that the associated parameters

(galaxy coordinates) are drawn from Gaussians at the center of the image with a standard deviation equal to a pixel size (0.04 arcsec). Likewise R denotes that the parameter is drawn randomly from the prior. Once the true metamodel parameters are determined, the predicted image is calculated and a Poisson realization of the map is drawn to obtain the simulated photon count map.

The true subhalos in this simulated data set are mostly low-significance subhalos. Out of the 25 true subhalos, only 3 improve the goodness-of-fit, $\log P(D|\Omega_{\hat{M}})$, by more than 35 when added to the model, meaning that the rest are below the detection threshold (i.e., 5σ) for 5 degrees of freedom.

4.2.2 Nominal Results

Figure 4.5 shows the nominal simulated photon count map (repeated across panels) and six fair samples from the ensemble of subhalo catalogs, represented with blue pluses. Also shown with the green markers is the true catalog of subhalos.

The multiple images in the simulated data set are formed by a background galaxy near the fold caustic of an elliptical foreground galaxy. The green right triangle and the blue left triangle show the position of the true and model background source, respectively. Similarly, the green diamond and the blue filled square indicate the position of the true and model host galaxy on the lens plane, respectively. The positions of the true and model galaxies are overlapping in all samples, although the positions of the blue markers have unnoticeable variations across samples due to the rather small uncertainties in the posteriors of the source and host galaxy positions.
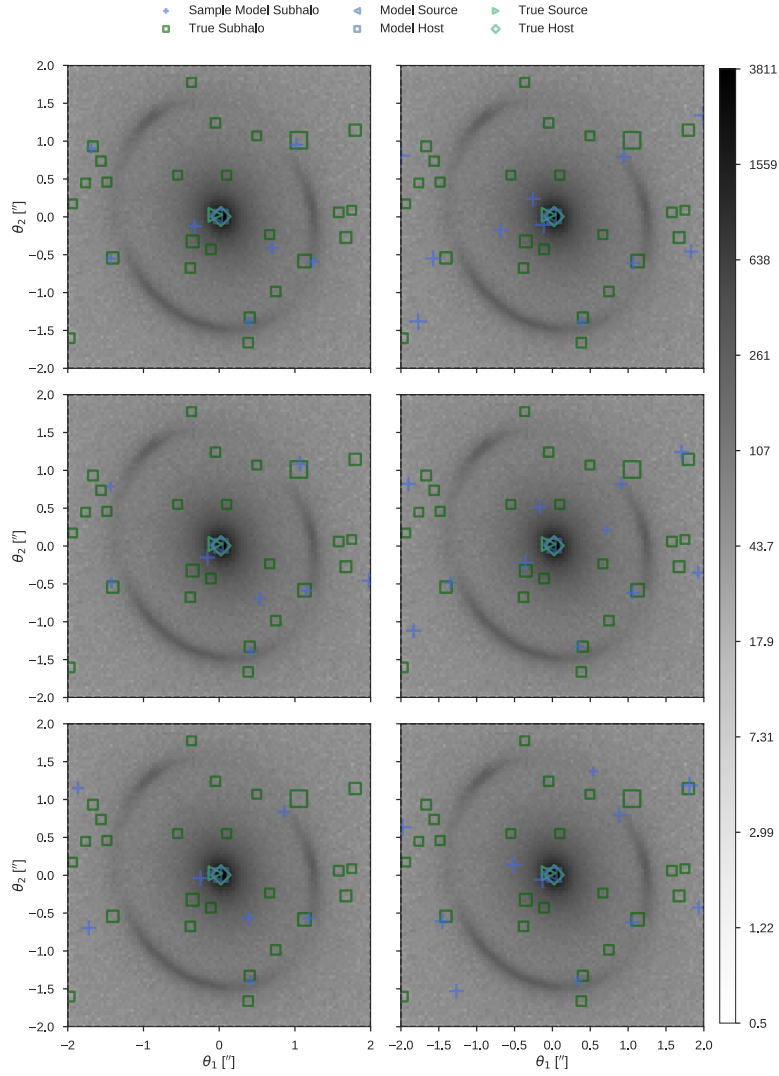
Figure 4.5: Fair samples from the posterior of the catalog space of subhalos. The gray scale shows the number of photons per pixel using a $\log$ stretch. Superposed on the simulated image, the panels show 6 out of 1000 fair samples from the posterior of the subhalo catalog space, using blue pluses. Posterior samples shown in the left column have the regularization prior, which effectively reduces the number of subhalos. The samples on the right have the default (i.e., uniform $P(\log \mu)$) prior. Also superposed on all panels is the true catalog of subhalos, shown with green squares. The sizes of the markers are proportional to the square root of the deflection strengths of the subhalos. The green diamond and blue square denote the positions of the true and sample host galaxy, respectively. Similarly, the green right triangle and the blue left triangle show the position of the true and sample background source, respectively. Note that the macro lens model also changes from panel to panel.

MACRO LENS MODEL   Subhalos have a perturbative effect on the deflection field otherwise dominated by the host galaxy. Therefore, any bias in the host galaxy parameters, as well as other macro lens parameters such as those of the background galaxy and the external shear, can potentially leak into constraints on the subhalo properties. Hence, the posterior of the macro lens model parameters are discussed first.

In general, it is found that the best constrained macro lens parameters are the positions of the host and source galaxies and the Einstein radius of the host galaxy. They generally have posterior 68% credible intervals that are 6×10^{-4} arcsec, 5×10^{-3} arcsec, and 0.01 arcsec wide, respectively. However, it is also found that the Einstein radius of the host halo correlates with its ellipticity, whose amplitude and orientation angle are constrained inside 68% credible intervals that are 0.01 and $2°$ wide, respectively. The anti-correlation can be attributed to the fact that the critical curve is constrained more strongly along one axis of symmetry than the other.

As an example, Figure 4.6 shows the distribution of posterior samples on the horizontal position of the host halo. The green vertical line indicates the true value of the parameter, whereas the red vertical line shows the value of the parameter in the maximum likelihood sample.

The host and source galaxy fluxes, on the other hand, are constrained at the $\sim 2\%$ and $\sim 10\%$ level, respectively. Specifically, the prior knowledge of the morphology of the galaxies, i.e., that they are Sérsic profiles of known indices, reduces the uncertainty on the fluxes. If the source plane emission had been allowed to vary from pixel to pixel subject to some regularization, the uncertainties on the background emission would be larger. Nevertheless, since the posterior is marginalized over the emission on the source plane, this propagates any uncertainties to the marginal posterior of the
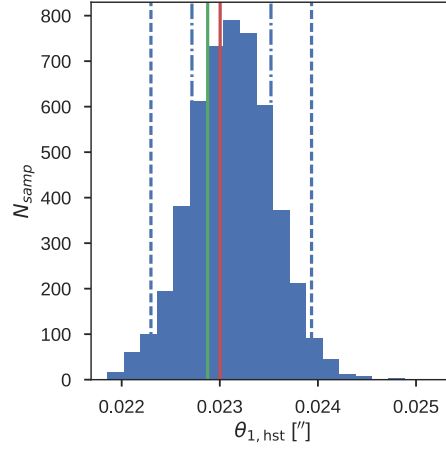
Figure 4.6: Histogram of samples from the posterior of the horizontal position of the host halo shown with the blue histogram. The green vertical line indicates the true value of the parameter, whereas the red vertical line shows the value of the parameter in the maximum likelihood sample. The dot-dashed and dashed vertical lines denote the following percentiles: 2.5, 16, 84, and 97.5.

subhalo properties. The choice of Sérsic profile also introduces a strong covariance between the luminosity and half-light radius of the galaxies, which can have correlation coefficients as large as 0.9.

The external shear is the least strongly constrained component of the macro lens model with typical uncertainties of $\sim 10\%$ on the magnitude of the shear field and $\sim 20°$ uncertainty on its direction. The large uncertainty in the alignment is due to the degeneracy of the quadrupole shear term with the overall deflection field due to subhalos. Similarly, the posterior of the lensed emission metamodel is marginalized over the shear field and its uncertainties are propagated to the uncertainties on the subhalo properties.

Finally, an uncertainty of $\sim 0.1\%$ is obtained on the isotropic emission component and the prior uncertainty on the PSF is recovered in the posterior. The latter allows one to marginalize over uncertainties due to the lack of perfect knowledge of the PSF and is not intended to gain information
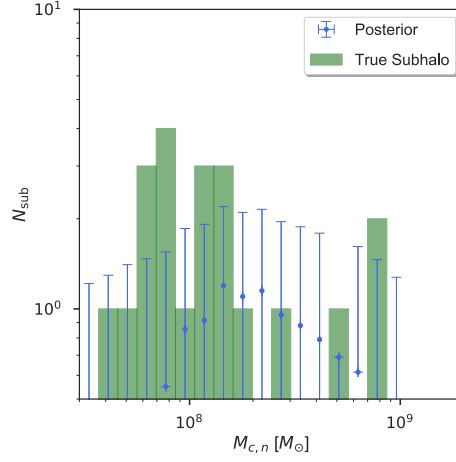
Figure 4.7: Histogram of the tidally truncated subhalo masses in the true model (green) and the posterior (blue). The central point and the error bars show the 16th, 50th and 84th percentiles of the histogram samples from the posterior, respectively. Posterior histogram agrees with the true histogram even though the majority of the subhalos are below detection threshold. It is important to note that the inferred posterior depends on the choice of $\alpha_{s,min}$, i.e., the posterior would extend to lower masses (and be more prior-dominated) if $\alpha_{s,min}$ is lowered. Therefore, the main inference of interest here is the *normalization* of the posterior subhalo mass distribution *given* a choice of $\alpha_{s,min}$ and a broad Gaussian prior on the log-slope, β.

about the PSF using the image.

SUBHALO MASS FUNCTION   The posterior truncated mass distribution of the subhalos is presented

in Figure 4.7. The figure shows the histogram of the truncated masses of the true subhalos in green.

Overplotted with blue error bars is the median histogram of the truncated masses of the subhalos

in the posterior samples. These blue points with error bars are obtained by first calculating the his-

togram of the truncated masses for each posterior sample from the metamodel. Then, 16th, 50th,

and 84th percentiles in each bin are plotted as the lower cap, central point, and the upper cap of the

error bar, respectively.

In Figure 4.7, mass bins that contain a true subhalo may have a median posterior other than one,

and the posterior uncertainties go above zero when there is no true subhalo. This is a result of accounting for transdimensional (across-model) covariances. Thanks to transdimensional proposals, a true subhalo can be fitted both with multiple and lower-mass subhalos and a single, more massive subhalo. It is important to emphasize that such transdimensional covariances exist even for true subhalos with high significance. Therefore, even though the log-likelihood difference between two models with and without a given subhalo may be well above 35, it may be as low as a few when comparing two models where the subhalo exists as a single, high-mass clump and as two, smaller-mass clumps. Furthermore, it is also true that a high-mass subhalo can be far from the multiple images, causing the posterior subhalo mass function to be uninformed of its existence.

The roll-off at small masses is due to the choice of the minimum deflection strength allowed in the model, $\alpha_{s,\min}$. If this value is decreased, the posterior extends to lower masses. Therefore, the constraints derived on the subhalo mass function must be quoted for a given $\alpha_{s,\min}$. Given the choice of the simulated data set and the simulation redshifts, the posterior subhalo mass function becomes prior-driven below $\sim 10^{8} M_{\odot}$. This is illustrated in Figure 4.8, which shows the truncated mass and significance of subhalos in a different simulated data set with a higher number of subhalos.

As for the deflection profile parameters $\theta_{s,n}$ and $\theta_{c,n}$, it is found that the projected cutoff radius is better constrained than the projected scale radius. WFC3's pixel size of 0.04 arcsec is comparable to the projected scale radius of an NFW subhalo with $r_{s,n} = 200$ pc at the simulation redshift of $z = 0.2$. This implies that current optical photometry is insensitive to the scale radii of most subhalos. However, more massive subhalos close to the multiple images can still be distinguished based on their projected scale radii, which motivates the choice to make $\theta_{s,n}$ subject to inference.
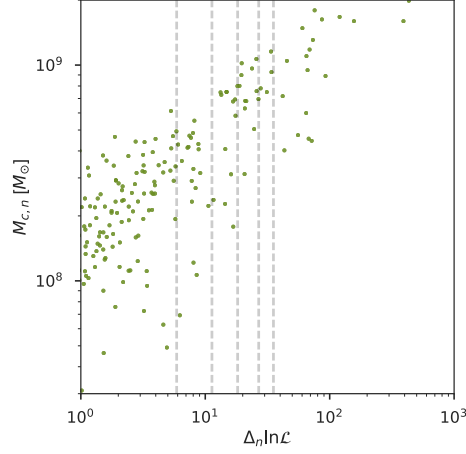
Figure 4.8: Truncated mass vs. significance (i.e., maximum log-likelihood difference of adding the subhalo to the model) of simulated subhalos in a different simulation with a higher number of subhalos. The vertical lines indicate 1σ through 5σ for 5 degrees of freedom. For this simulation, subhalos fall below 1σ at a mean truncated mass of $2 \times 10^{8} M_{\odot}$.

The mass of a subhalo does not have a one-to-one correlation with its significance. In a photometric inference problem, how the flux of a given point source compares to the background emission inside the FWHM of the PSF, largely determines its significance[180]. Therefore, isolated, bright light sources are expected, on average, to be more statistically significant than fainter ones. However, how strongly a subhalo can be constrained in the lensing problem, depends on its mass as well as how its deflection field is oriented with respect to the lensed emission. Since the deflection profile of a subhalo drops beyond its projected scale radius, in order to have non-negligible effect on the log-likelihood, a subhalo needs to be close to an already lensed emission on the image plane *as well as* having a large enough mass.

The leading order contribution of a given subhalo with index n to the model image can be obtained by Taylor expanding Equation (4.1). Recasting this equation in terms of the photon count

map lensed by the host halo, external shear, and all subhalos except the nth subhalo, $\bar{k}_m$, the lensed count map

$$\tilde{k}_m(\theta_1, \theta_2) = \bar{k}_m(\theta_1, \theta_2) + \vec{\alpha}_n \cdot \vec{\nabla} \bar{k}_m(\theta_1, \theta_2) + \mathcal{O}(\vec{\alpha}_n^2), \qquad (4.21)$$

is obtained.

The second term is negative over some regions of the lensed emission and positive in other regions. $\left\langle \left| \vec{\alpha}_n \cdot \vec{\nabla} k_m \right| \right\rangle$ instead better encapsulates the level of perturbation a given subhalo introduces to the image already lensed by all other mass components. Therefore, it is a relatively more accurate estimator of the significance of a given subhalo compared to $M_{c,n}$. This quantity is referred to as the subhalo *relevance*. Subhalos with low relevance cannot be constrained using the observed photometric data. An example can be seen in Figure 4.5, where the true subhalo at (1.8, 1.1) arcsec is relatively less constrained despite having a mass above $10^8 M_\odot$.

Figure 4.9 shows the correlation between the relevance of a subhalo and the log-likelihood increase when including the subhalo into the lens model while leaving all other parameters fixed. The correlation is not one-to-one, however, due to the fact that the host galaxy and the external shear can dominate the deflection field in some regions more than others, which causes subhalos with equal relevance to have different significances. Therefore, in general significant subhalos have high relevance, but high relevance subhalos can have very low significance.
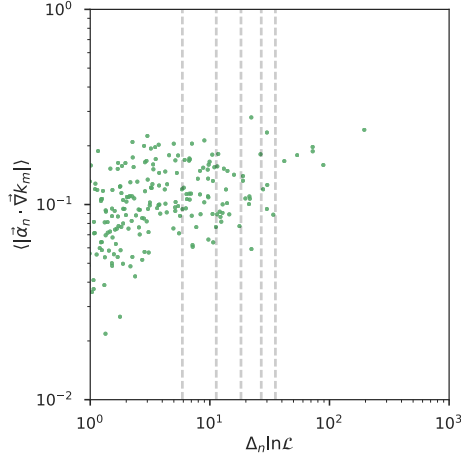
Figure 4.9: Correlation between the relevance of a simulated subhalo population and the log-likelihood improvement in adding each subhalo to the model, $\Delta_n \ln \mathcal{L}$. Vertical dashed lines indicate $1\sigma, 2\sigma, 3\sigma, 4\sigma$, and 5σ contours for 5 degrees of freedom. These subhalos are independently drawn and do not correspond to the true subhalos in the simulated strong lens shown in Figure 4.5.

Subhalo mass fraction   The shape of the subhalo mass function is poorly constrained given the small number of strongly lensed systems analyzed so far. Nevertheless, the fraction of mass tied in subhalos in the vicinity of the critical curve of the host halo, can be accurately measured for high signal-to-noise strong lenses. Therefore, the subhalo mass fraction provides an observational summary statistic of the subhalo mass function.

In order to determine the posterior fraction of mass locked in subhalos, the quantity is first calculated for each sample and then the moments of the posterior are determined. The subhalo mass fraction for a given sample from the metamodel is defined as the ratio between the mean surface mass density due to all subhalos and that due to the host halo inside an annuli centered at the host
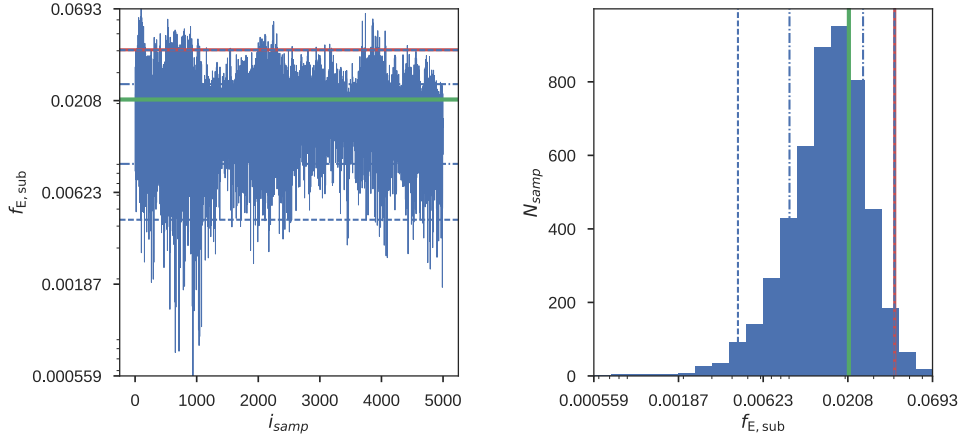
Figure 4.10: Histogram of samples from the posterior of the subhalo mass fraction (right). The MCMC time evolution of the quantity (left). The legend is similar to that in Figure 4.6. The red (maximum likelihood sample), and the dashed blue (97.5th percentile) lines overlap.

position with radius $\theta_{\mathrm{E,hst}}$ and thickness 0.2 arcsec,

$$f_{\mathrm{E,sub}} \equiv \frac{\sum_n \int_{\mathrm{ann}} \kappa_n d^2\theta}{\int_{\mathrm{ann}} \kappa_{\mathrm{hst}} d^2\theta}. \tag{4.22}$$

where "ann" represents the mentioned annulus around the host halo. The resulting posterior of

the subhalo mass fraction at the Einstein radius is shown in Figure 4.10 along with its evolution as a

function of MCMC time. Note that a prior probability distribution on the subhalo mass fraction,

$f_{\mathrm{E,sub}}$ is not imposed directly. The figure shows that the posterior of $f_{\mathrm{E,sub}}$ agrees with the true value

and that it is sensitive to the existence of subhalos below detection. The green line shows the true

subhalo mass fraction.

Last, Figure 4.11 shows the median of the posterior deflection due to various components. In or-
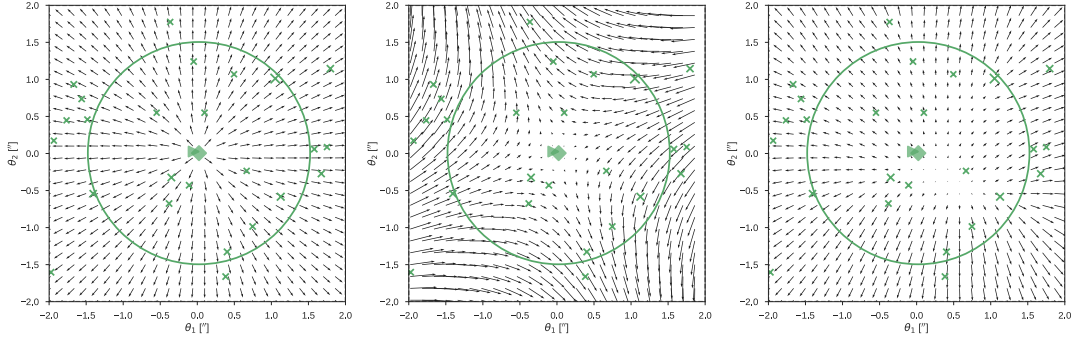
Figure 4.11: Median of the posterior deflection fields due to various components. The arrows indicate the local direction of the deflection field and have lengths that represent the magnitude of the deflection field. However, the arrows in the left and right panels have been scaled by 0.1 and 10, respectively, for better visibility. The positions of the true subhalos are shown with green Xs, whose area are proportional to the masses of the subhalos. The radii of the green and dashed blue circles are the Einstein radii of the true and sample macro lens models, respectively. They are only drawn to guide the eye in the absence of the observed image.

der to compute these maps, the horizontal and vertical components of the associated vector field

are first calculated for each fair sample from the posterior of the lensed emission metamodel. Then,

for each pixel, the 50th percentiles (medians) are determined. Figure 4.11 then illustrates the deflec-

tion field separately for the host halo (left), external shear (center), and subhalo population (right).

The radius of the green and blue dashed lines, which are also mostly overlapping, denote the circles

whose radii coincide with the Einstein radii of the true and model host halos, respectively. Since the

host galaxy is elliptical, however, the circles do not correspond to the critical curve and are intended

to guide the eye. The area of a circle is proportional to the host mass contained inside the Einstein ra-

dius. The true Einstein radius is 1.5 arcsec, yielding a mass inside the Einstein radius of $5 \times 10^{11} \ M_\odot$.

Posterior total deflection field is found to be constrained at the percent level across most of the

image. However, this translates to $\sim 100\%$ uncertainty in the deflection field due to subhalos. Pos-

terior deflection field due to subhalos is only constrained well near the multiple images, where the
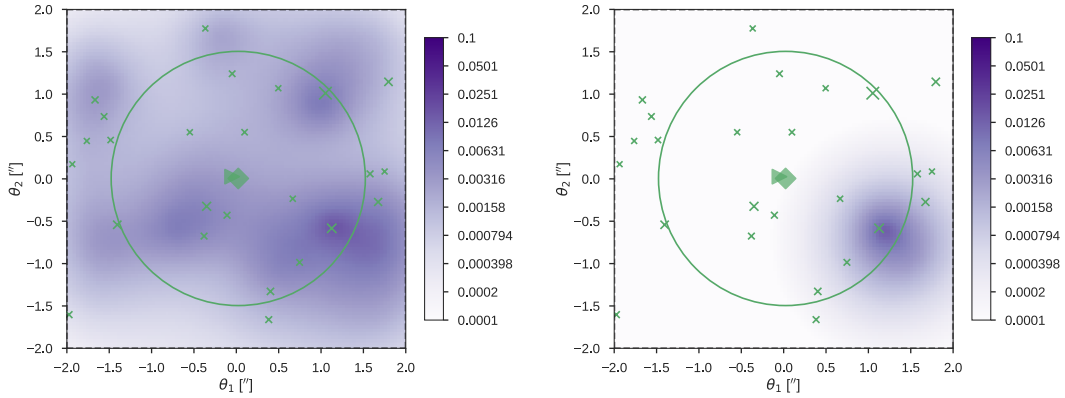
Figure 4.12: Mean of the posterior convergence obtained from the transdimensional (left) and one-subhalo (right) inferences. Green markers and the circle have the same meaning as in Figure 4.11.

uncertainties can be as low as 20%.

4.2.3 THE BIAS IN A ONE-SUBHALO LENS MODEL

Fixed-dimensional lens modeling requires iteratively testing lens models with a fixed number of subhalos (or a fixed amount of any other form of additional model complexity) in addition to the smooth lens model. In the simplest approach, a single subhalo can be added to the smooth lens model to fit an observed photon count map. In this section, the previous results from transdimensional sampling are compared to such a case, where the lens model contains a single subhalo. To make a robust comparison, the same simulated photon count map is used as shown in Figure 4.5. Then, instead of allowing a variable number of subhalos in the lens model, the transdimensional proposals in PCAT are turned off and the number of subhalos is fixed to 1. The latter scheme is still Bayesian and probes all within-model covariances, but neglects across-model covariances.

Figure 4.12 illustrates the median of the posterior convergence obtained in the transdimensional (left) and one-subhalo (right) inferences. These posterior convergence maps are obtained similarly to the posterior deflection maps, i.e., the convergence maps are calculated for each sample from the posterior by numerically differentiating the deflection field, and then the median convergence in each pixel is presented as the median of the posterior convergence map.

In the left panel of Figure 4.12, transdimensionality allows subhalos to be born and die across the image, probing the goodness-of-fit of all combinations of subhalo multiplicities, positions and deflection profiles allowed by the prior. Such configurations mostly include mild likelihood improvements below the detection threshold assumed in traditional cataloging. The median of the posterior convergence reveals low-significance diffuse features in the underlying subhalo convergence field as well as constraining superthreshold true subhalos. In contrast, the median of the posterior convergence obtained from the one-subhalo inference on the right contains a single overdensity, which is also inferred by the transdimensional inference. However, the transdimensional approach also (partially) deblends subhalos in the crowded region to the extent allowed by the information available in the data and degeneracies in the likelihood, despite the fact that most sample subhalos presented in Figure 4.5 do not indicate associations with the true subhalos.

The failure of the one-subhalo inference to reveal the remaining features in the underlying true convergence field is due to the absence of additional model subhalos which causes the macro lens model and the single subhalo to be biased to increase the likelihood. Without proposals to probe transdimensional covariances, the sampler explores a mode of likelihood of the one-subhalo lens model. The inference of lower-significance features in the convergence map does not imply that
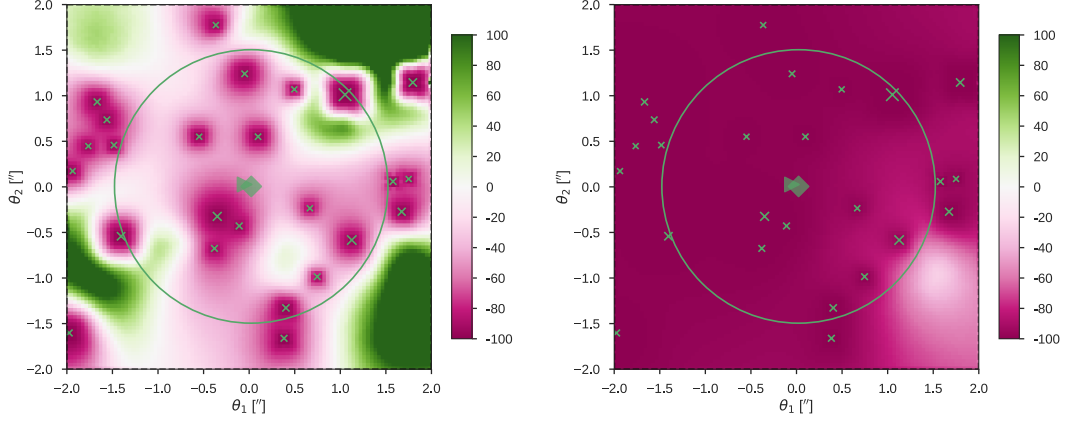
Figure 4.13: Median of the posterior residual between the model and true convergence in the transdimensional (left) and one-subhalo (right) inferences, given as percentage of the true subhalo convergence map. The divergent color scale is linear and saturated at ± 100. Green markers and the circle have the same meaning as in Figure 4.11.

the additional subhalos are formally detected. The increase in the goodness-of-fit between the two

approaches is $\Delta \log P(D|\Omega_{\hat{M}}) \sim 10$. In order to detect these model subhalos at 5σ, they would be

required to individually improve the goodness-of-fit above $\Delta \log P(D|\Omega_{\hat{M}}) \sim 35$ given that they

have 5 degrees of freedom.

The difference between the two inference schemes becomes more evident when the posterior

predictions are compared with the underlying true convergence field due to subhalos. Figure 4.13

illustrates the residual between the true subhalo convergence map and the median of the posterior

convergence obtained in the transdimensional (left) and one-subhalo (right) inferences, respectively.

The disagreement is significantly smaller in the case of a transdimensional approach, which correctly

predicts the mean subhalo convergence field over most of the lens plane and have smaller disagree-

ment close to the subhalos. However, note that even the transdimensional approach cannot fully

constrain the true subhalo convergence field in the vicinity of most subhalos. This is expected, be-

cause the simulated photon count map contains very little information on their properties.

Figure 4.14 shows the posterior of the position and deflection strength of the model subhalo in the one-subhalo inference. The joint distribution is observed to be highly skewed with heavy tails. Furthermore, the median of the posterior mass of the subhalo, $4.5 \times 10^8 M_\odot$, is biased low compared to the mass of the closest true subhalo at (1.2, -0.5) arcsec, $8 \times 10^8 M_\odot$.

Another problem with fitting the lensed image using a fixed number of subhalos is that the posterior of the macro lens model can be significantly biased because of the missing modeled subhalos. This is because the macro lens model compensates for the absence of subhalos below the detection threshold, resulting in a seemingly high goodness-of-fit that mismodels the underlying true lensed emission metamodel. An example is shown in Figure 4.15, where the posterior of the host halo ellipticity is plotted for the cases of nominal (transdimensional) and one-subhalo inferences using a similar, but independently drawn data set. One-subhalo inference is observed to generate multimodality in the posterior of the ellipticity of the host.

4.2.4 Extending the metamodel to lower subhalo masses

The lower limit of the deflection strength prior can be made smaller at the expense of making the posterior prior-dominated (i.e., resulting in little information gain) and significantly increasing the computation time required for convergence. Therefore, given the observed data, there is a reasonable lower limit to the deflection strength, e.g., 0.01 arcsec, which can balance information gain and computational complexity.

Next, results from a different inference is presented, where there is no mismodeling at the low
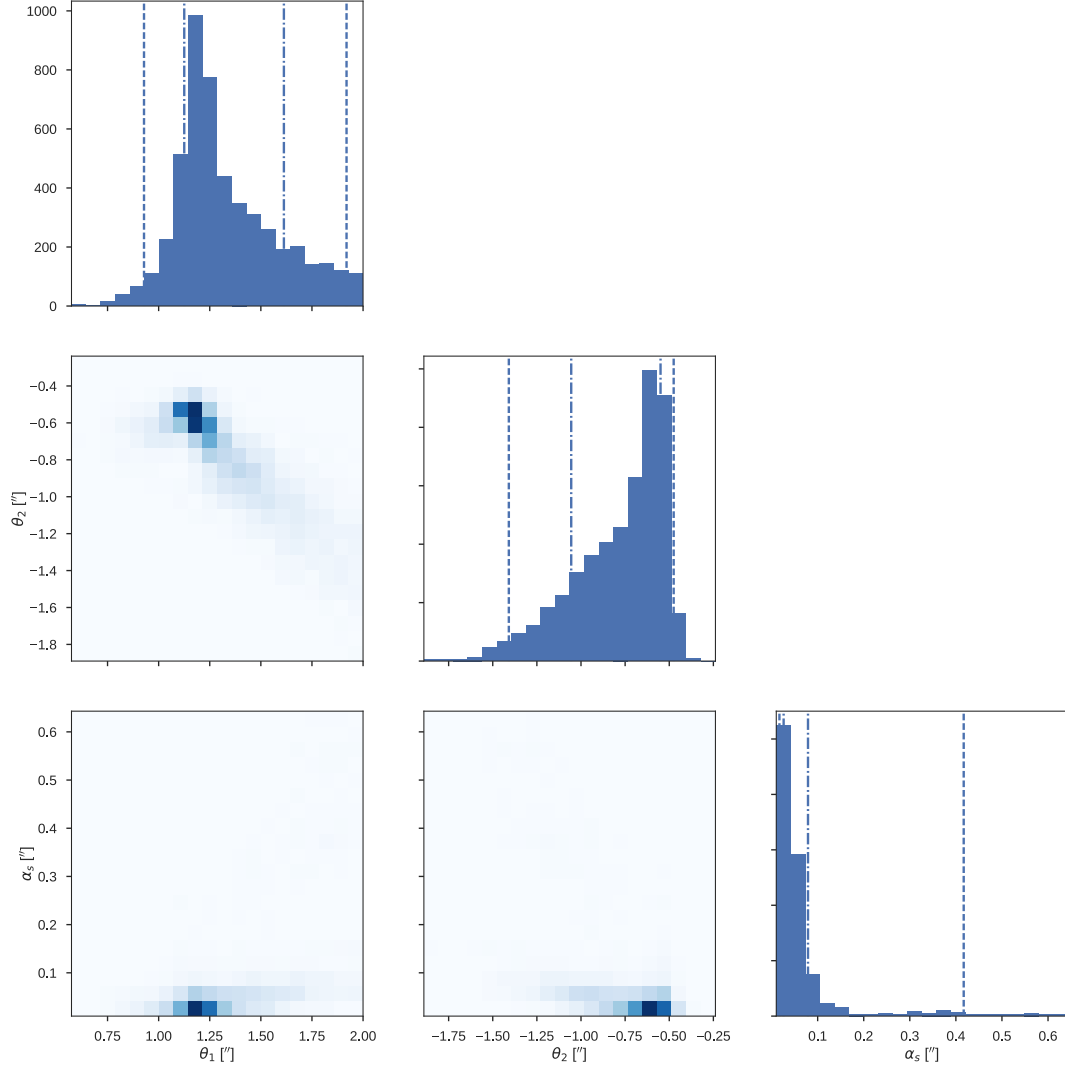
Figure 4.14: Joint posterior of the horizontal and vertical positions and the deflection strength of the subhalo in the one-subhalo inference. Dashed lines have the same meaning as in Figure 4.6. For comparison, the position of the nearest true subhalo is (1.2, -0.5) arcsec and it has a deflection strength of 0.05 arcsec.
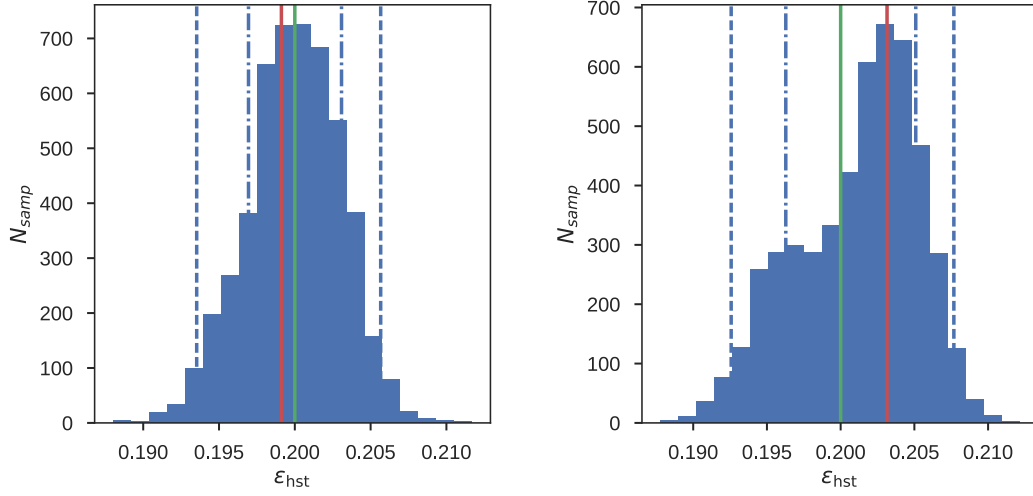
Figure 4.15: Histogram of samples from the posterior of the host halo ellipticity. The legend is same as that in Figure 4.6.

end of the subhalo mass function, unlike the nominal results in Section 4.2.2, where the minimum subhalo deflection strength allowed in the fitting metamodel is a factor of 3 higher than the true minimum subhalo deflection strength. Using the same data set as shown in Section 4.2.2, Figure 4.16 compares the posterior of the number of subhalos in the nominal results (left) and in a run, where $\alpha_{s,\min} = 0.003$ arcsec (right). Given a lower $\alpha_{s,\min}$, the run shown in the right panel explores a higher number of subhalos whose masses are lower on average. The two posteriors are consistent with each other. Given thata a scale-free prior exists on the mean number of subhalos, μ_{sub}, the fact that there is a preferred scale in the number of subhalos and that it is consistent with the true number of subhalos, implies that the posterior is informed by the underlying subhalos below the detection threshold.
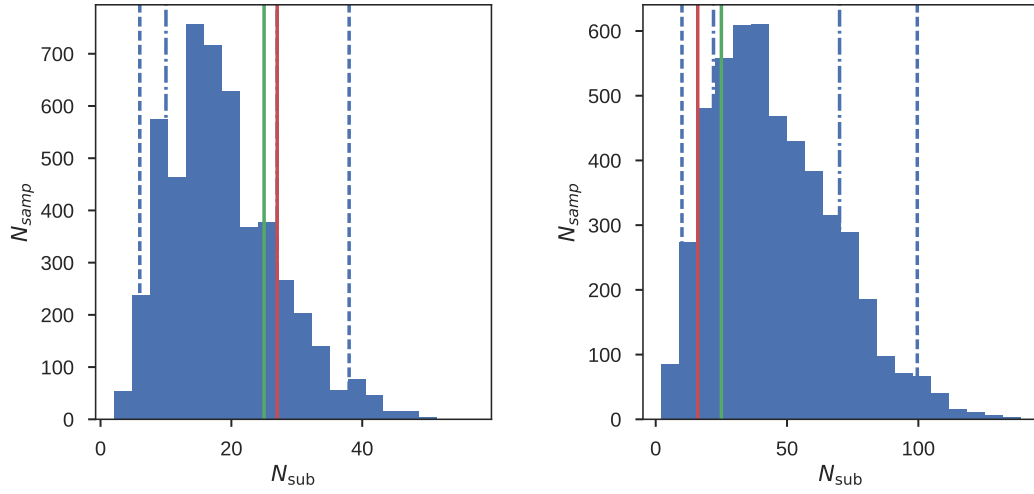
Figure 4.16: Histogram of samples from the posterior of the number of subhalos shown with blue. The legend is same as that in Figure 4.6.

4.3 No-signal test

Any inference framework is expected to produce null results when subject to an input data set that does not contain the signal of interest. Therefore, the inference framework is next tested on a simulated data set that was not affected by any subhalo. For this purpose another simulated image is generated as a Poisson draw of an image forward-modeled by the macro lens model, background emission and PSF (same as that used to generate the main data set presented in the thesis), but without any subhalo. Then, PCAT is run on this image. Figure 4.17 shows the histogram of the number of subhalos in the posterior, which is consistent with the true value of 0. Furthermore, none of the subhalos in the posterior ensemble of catalogs is more significant than 4σ. The fact that there is a non-zero number of subhalos in the posterior is largely a consequence of the degeneracy between

the mass of the main deflector and the total mass of subhalos inside the critical curve. The total mass in the subhalos, as shown in Figure 4.18, is also mostly zero, although the posterior does contain some mismodeled subhalo mass due to degeneracies in the lensing problem. The 84th percentile (vertical blue dashed-dotted line) extends to about $3 \times 10^8 M_{\odot}$, which coincides with the maximum likelihood sample (vertical red line), indicating that these few samples overfit the image.

4.4 Sensitivity to the prior choice

An important nuisance parameter of the model is the minimum deflection strength allowed by the prior. The reason for its criticality is that the subhalo mass function is expected to be steep, causing the posterior total number of subhalos to sensitively depend on the minimum deflection strength
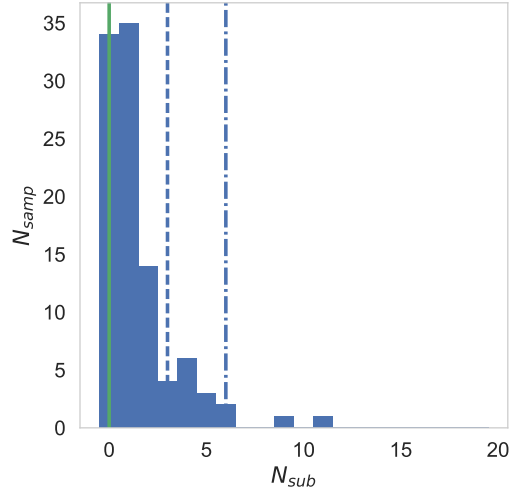
Figure 4.17: Histogram of the number of subhalos in the posterior when there is no true subhalo that affects the simulated image. The legend is sames as that in Figure 4.6.

allowed by the metamodel. The lower $\alpha_{s,min}$ is allowed to go, the more prior dominated the posterior becomes. Conversely, taking $\alpha_{s,min}$ too high causes mismodeling, since this removes from the metamodel those subhalos with marginal significances that could fit subtle features in the data.

Figure 4.19 illustrates the effect of varying this nuisance parameter away from the nominal value of 0.01 arcsec. In the left panel, $\alpha_{s,min}$ is reduced to 0.005 arcsec, whereas in the right panel, it is increased to 0.02 arcsec. Although the posterior in the left panel agrees better with the ground truth at the low mass end, this happens at the expense of oversplitting the most massive subhalos. Furthermore, the lower end of the posterior subhalo mass function becomes prior-dominated, and could miss a potential cutoff or flattening required by the data. The 0.02 arcsec run on the right performs comparably at the high masses, but fails to explore the subhalo mass function at the low-mass end.
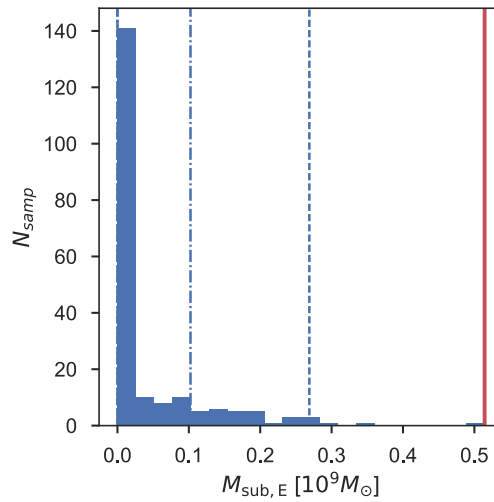
Figure 4.18: Posterior distribution of the total mass in subhalos inside the Einstein radius of the host galaxy, when there is no true subhalo that affects the simulated image. The legend is sames as that in Figure 4.6.
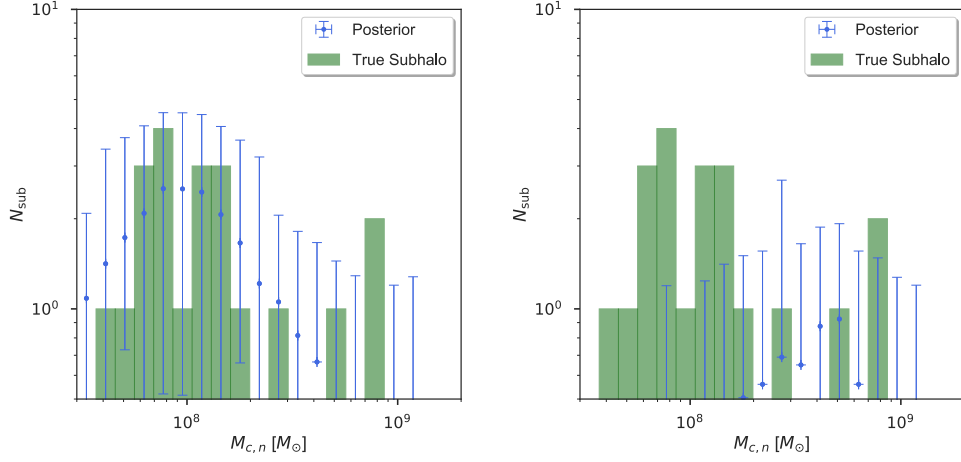
Figure 4.19: Posterior histogram (blue) of the subhalo masses, when $\alpha_{\mathrm{s,min}}$ is 0.005 arcsec (left) and 0.02 arcsec (right). The green histograms show the mass function of the true subhalos.

These results based on simulated data show that probabilistic cataloging is an effective tool to probe lens models with a variable number of deflectors. Probabilistic cataloging differs from traditional cataloging, because it does not discard information or bias the posterior by imposing a detection threshold on model subhalos, i.e., requiring that model subhalos improve the goodness-of-fit above some threshold. By subsequent marginalization over the catalog space of subhalos, it obtains robust estimates of the underlying mass model and population characteristics of subhalos, without individually detecting them. This contrasts with traditional cataloging, where detections and nondetections are used to constrain the subhalo mass function. However, because these detections are not fair draws from the underlying subhalo mass function, this approach cannot account for covariances in the subhalo catalog space, potentially yielding biased results with underestimated uncertainties. It is shown that, by not including these small subhalos in the lens model, fixed-dimensional inference methods can significantly mismodel the data. Furthermore, the subhalo mass function can be

probed even when many subhalos in the sample catalogs are individually below the detection threshold and would be absent in a traditional catalog.

Currently, the most extensive and homogeneous collection of optical images of galaxy-galaxy type strongly lensed systems is the SLACS collection[35,34]. These targets have been spectroscopically selected from the Sloan Digital Sky Survey (SDSS) data set conditional on encountering multiple strong emission lines at a higher redshift than that of the target galaxy. They were later followed up by the Advanced Camera for Surveys (ACS) on the HST. However, the limited depth of the HST only allows the most massive subhalos, if any, to be constrained[226,227]. Next generation, wide field, high angular resolution telescopes such as WFIRST and EUCLID are expected to revolutionize studies of dark matter substructure. The number of high signal-to-noise strong lens images is expected to go up to thousands[220,170]. Furthermore, greater exposure depth should allow constraints on multiple subhalos per strong lens, making it possible to constrain the subhalo mass function to lower masses. These infrared telescopes will also carry spectrometers, which, in synergy with the photometric redshifts provided by the Large Synoptic Survey Telescope (LSST), will allow simultaneous constraints on the redshifts of the model sources and deflectors. Follow up of high signal-to-noise systems with the James Webb Space Telescope (JWST) and interferometry at submillimeter wavelengths, e.g., Atacama Large Millimeter Array (ALMA), may allow constraints on the subhalo properties of systems at high redshift.

The presented inference can be improved in several ways. First, multi-band images can be used by introducing color parameters for the source and host emissions. This would make available information contained in the different colors of sky background, source, and foreground galaxies.

Furthermore, the source plane emission can be parametrized on an adaptively refined grid, instead of imposing that the source plane emission is from a galaxy with a perfect Sérsic profile. This would yield a more realistic model of the source plane emission, and a more principled method to marginalize over uncertainties in the emission from the source plane. This is especially relevant for late type background galaxies, where emission has strong spatial features such as outflows, bright spots or dust obscured regions on the disc.

Probabilistic cataloging can also be extended to use temporal information by sampling from the posterior of a lensed emission metamodel given exposures of the strongly lensed system taken at *different* times. Because lensing is also sensitive to the angular diameter distances to the lens and source planes, this would allow an independent measurement of the Hubble constant, H_0, as in [214], but with constraints that have been marginalized over dark subhalos that can potentially bias the macro lens parameters. Also, combining multiple exposures may improve sensitivity to substructure via time-delay lensing[138,65], although microlensing could potentially bias this inference[218].

Furthermore, probabilistic cataloging provides a principled way to gather information from *independent* strong lens systems, when constraining the posterior of the subhalo parameters. Given images of different strong lenses, the catalog space consistent with these images can be jointly sampled with hyperparameters characterizing the population characteristics of the subhalos. This is a significant improvement with respect to traditional cataloging, where incorporating nondetections into the inference may become ambiguous, since fixed-dimensional approaches do not propagate transdimensional covariances

Inference of mass substructure on the lens plane also has strong degeneracies with the inference

of light substructure on the source plane. In particular, the light substructure in the source plane can affect the lensed image similar to the mass clumps in the lens plane. Such degeneracies can be partially broken by exploiting a data set such as velocity-binned ALMA images, that are informative on the line-of-sight velocities of the light substructure, thus allowing component separation in the source plane.

5

Probabilistic Cataloger

PCAT is the software implementation of the probabilistic cataloging discussed in this thesis. It was initially designed to sample from the posterior of a gamma-ray emission metamodel consistent with the Fermi-LAT data[74]. Later, it was generalized to include lensing in the emission metamodel. It has been used in two publications[74,73]. Its design philosophy is to fill the gap between the statistically unsophisticated users and data sets that require a transdimensional and covariant modeling. A typi-

cal user experience is the preparation of the data in the form that can be fed into `PCAT`, choosing the metamodel, element kernel type, and priors on the element parameters as well as common parameters of the metamodel. The outputs of the framework are the HDF5 files that contain the chain and plots that visualize the posterior of the metamodel in various projections. It's main dependencies are `numpy`, `scipy`, `matplotlib` and `multiprocessing` modules.

In addition to the gamma-ray and lensed emission metamodels presented in this thesis, `PCAT` has two more metamodels that have been implemented in the course of this PhD. The first is the X-ray emission metamodel, which can be used to sample from the posterior of the catalog space of X-ray point sources consistent with deep field X-ray images. These samples then can be correlated with the crowded optical catalogs of the same region, yielding a posterior of the redshift and optical properties of the faint X-ray point sources. Second, `PCAT` can also assume the standard Gaussian mixture model as its metamodel, which can be used to sample from the posterior of the multiplicity, amplitude, and size distributions of *clusters* on the integrals-of-motion plane (kinetic energy and momenta) of stars, which allow inferences of their formation history. These two metamodels have been omitted from this thesis.

5.1 Proposal scale

MCMC can efficiently sample from a given probability distribution only when the proposal distribution is *similar* to the target probability distribution. For instance, when the target probability distribution is a Gaussian with a standard deviation of unity, the mixing of the chain will be slow when

the proposal scale is either much smaller or larger than unity. In the former case, the chain takes too small steps, achieves an acceptance ratio of $\sim 100\%$, increases the correlation between successive samples, and requires more thinning to obtain uncorrelated samples. In the latter case, proposals get frequently rejected and the state of the chain will evolve at a time scale longer than the number of samples taken by the sampler. The optimal scaling of MCMC proposals requires acceptance ratio to be 23.4% in the large dimension limit[183].

PCAT sets the proposal scale by determining the Hessian matrix, i.e., the curvature of the target probability distribution at the initial state. Because the chain is initialized at a draw from the prior, the initial Hessian estimate can bring the state closer to a fair draw from the posterior, but cannot efficiently sample from the posterior. Therefore, once the acceptance ratio drops significantly, PCAT restarts with a new Hessian estimate and this scale learning process continues iteratively until successive iterations start generating fair samples from the posterior. Frequently, however, one has good priors on what the proposal scale should be for the posterior of interest. Therefore, PCAT also allows one to bypass this iterative scheme by letting the user define the proposal scale for every parameter. In order to maintain the optimal acceptance rate, the scale of within-model proposals is normalized by the square root of the number of parameters whose values are proposed to change simultaneously,

$$\sigma'_P = \frac{1}{\sqrt{N_{per}}} \sigma_P. \tag{5.1}$$

where σ_P and σ'_P are the single and multiple parameter proposal scales in the unit space and N_{per} is

the number of parameters perturbed per within-model proposal.

The relative frequencies of different proposal groups can also be adjusted by the user. At low signal-to-noise, birth and death proposals become the principle proposals that align with the data uncertainties, whereas as the signal-to-noise increases, split and merge proposals become the principle transdimensional proposals that probe covariances.

The acceptance ratio of birth and death proposals is set by the minimum of the prior on the element amplitudes and the regularization prior. When the regularization prior is applied, a too-low minimum amplitude results in an acceptance ratio much less than 0.1. Conversely, when the regularization prior is not applied, a too-low minimum amplitude causes the acceptance ratio to increase to $\sim$ 1. Therefore, by default, PCAT tunes this minimum to the value that yields an acceptance ratio of $\sim$ 0.1.

5.2 TIME COMPLEXITY

Probabilistic cataloging requires sampling from the posterior. Therefore, its time complexity depends on the typical number of elements (i.e., dimensionality) of the metamodel and can be high for even moderate numbers of elements. Furthermore, the hierarchy and transdimensionality of the metamodel can further increase the time complexity. As a result, in general, the time complexity of probabilistic cataloging is higher than that of traditional cataloging roughly by a factor equal to the number of fair samples taken from the posterior. Therefore, caution must be exercised to ensure that probabilistic cataloging is scalable.

For each proposal in MCMC, the current parameter vector is used to compute the forward-model, which is then compared with the data through the likelihood. This forward-modeling is the leading term in the time budget of a single proposal. In PCAT, forward-modeling of element parameters is performed perturbatively. During each proposal, the current forward-model is modified for the change in the element that is proposed to change, be added or be killed. This decreases the time complexity of processing a single sample from $\mathcal{O}(N_D N)$ to $\mathcal{O}(N_D)$, where N is the number of elements and N_D is the number of data points. Furthermore, it is assumed that the affect of the elements on the forward-model vanishes outside some circle around the element, resulting in a *local* evaluation of the forward-model. Assuming that the closest pixel to the nth element has index j, the evaluated element kernel is

$$
K_{ijm} = \begin{cases} \mathcal{K}_{im}(\theta_{1,n} - \theta_{1,j}, \theta_{2,n} - \theta_{2,j}) \text{ if } j \text{ in } L_h \\ \\ 0 \text{ otherwise} \end{cases} \tag{5.2}
$$

where $\mathcal{K}_{im}$ is the element kernel, K_{ijm} is the element kernel evaluated on the forward-model grid, $\theta_{1,j}$ and $\theta_{2,j}$ are the horizontal and vertical positions of the jth pixel, and L_h is the list of nearby pixels to pixel j for the hth amplitude bin. L_h is precomputed for all amplitude bins before sampling and stored as a look-up table. The radius of the circle that L_h extends over, depends on the amplitude of the element and is determined such that the approximation error by the largest amplitude allowed by the prior contaminates the lowest amplitude at most by one percent.
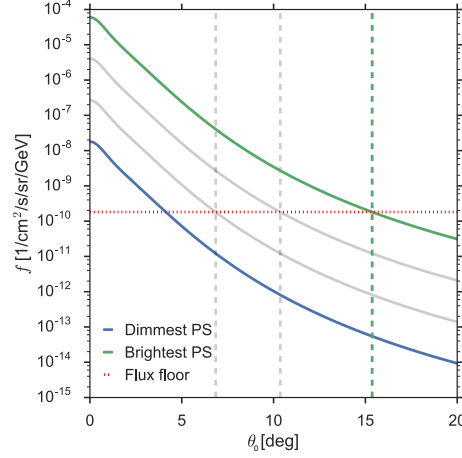
Figure 5.1: The radial profile of emission from a point source as reconstructed by the Fermi-LAT in different flux bins. The tail of the PSF is a power-law and is not exponentially suppressed. The horizontal line highlights the minimum flux allowed by the prior multiplied by 0.01. When the position, flux or the color of a point source is changed, its contribution to the total model is updated only inside a circle centered at the pixel closest to the point source. The radius of the circle is set such that this approximation causes a negligible error.

The circle radii are illustrated in Figure 5.1 for the emission metamodel, where the horizontal line indicates $0.01 f_{min}$, where f_{min} is the minimum flux allowed by the emission metamodel. Hence, point sources contribute to the forward-model up to the radius, after which the bias introduced by their neglect is below one percent of the contribution of the faintest point source. The error introduced by this approximation is monitored throughout sampling to ensure that the bias is negligible. As a result of this approximation the time complexity of a single proposal is further reduced to $\mathcal{O}(N_{Dk})$, where N_{Dk} is the number of data points where the element kernel does not vanish.

For element kernels that are approximately Gaussian, such as PSFs in X-rays and optical wavelengths, the evaluation in Equation 5.2 can be further accelerated by linearizing the element kernel and replacing interpolation with a matrix multiplication of expansion coefficients and small (i.e.,

leading order) shifts on the forward-model grid. This is then used to evaluate in parallel the forward-models of elements with more complex morphologies such as galaxies with Sérsic profiles, as the sum of linearized kernels called *phonions*. This was implemented in a different software of probabilistic cataloging optimized for optical photometry[180], but has not been included in this thesis.

Moreover, inference of large fields can be broken down into independent inferences of tiles not much larger than $\sim 10 - 20$ times the size of the element kernel. Because elements further away have vanishing covariances, their posteriors can be sampled independently. The regularization prior also has the advantage of significantly reducing the time complexity of the probabilistic cataloging by only allowing relatively significant subthreshold elements, $\gtrsim 1\sigma$, into the model. In addition, the term $\ln k_{ij}^{\mathrm{D}}$ in the log-likelihood is fixed for a given data set and neglected when calculating the differences in log-likelihood. PCAT also achieves parallelism at the process level via bypassing the Global Interpreter Lock (GIL) of Python, where the main (parent process) spawns multiple, non-interacting child processes to accelerate computation.

The resulting time budget of PCAT per sample per band per data quality class are 1 ms, 1 ms, 3 ms and 12 ms for the X-ray emission metamodel, Gaussian mixture metamodel, gamma-ray emission metamodel, and the lensed emission metamodel, respectively. Because X-ray emission is forward-modeled on a Cartesian grid and the element kernel has ~ 400 pixels, it has the smallest time budget. Gamma-ray emission is evaluated on HealPix and the PSF in the lowest energy is oversampled due to PSF tails, resulting in a larger (~ 5000) number of pixels per band in the element kernel. Lensed emission is the slowest forward-modeling because of the solution of the lens equation.

5.3 CONVERGENCE

MCMC allows sampling from multimodal and covariant posteriors with the caveat that convergence to the stationary distribution can require a long simulation time. When the posterior extends to models with thousands of parameters, the finite available MCMC time may preclude the convergence of the sampled distribution to the desired target distribution.

5.3.1 ACROSS-CHAIN VARIANCE

One method for evaluating the chain convergence is to calculate the variance of the sampled chain. However, the variance of a single MCMC chain can underestimate the true variance, since the chain may not have converged to the target distribution despite having a small variance. Therefore, multiple, ~ 4, chains are run and the mean of the chain variances are compared with the variance of the means of the chains[104], ensuring that the initial states of the chains are over-dispersed relative to the target distribution. Having N_c chains and N_s samples in each chain, the resulting test statistic,

$$\hat{R} = \sqrt{1 + \frac{B}{W} - \frac{1}{N_s}},$$

(5.3)

is known as the Potential Scale Reduction Factor (PSRF). It estimates the factor by which the statistical uncertainty of a random variable could be reduced by running the chains longer. Therefore, it can be used to assess whether the chains have converged to the same distribution. Here, W is the

within-chain variance, i.e., the mean of the chain variances,

$$W = \frac{1}{N_c} \sum_{c}^{N_c} \left(\frac{1}{N_s - 1} \sum_{s}^{N_s} (y_s - \bar{y}_c)^2 \right), \tag{5.4}$$

B is the between-chain variance, i.e., the variance of the chain means,

$$B = \frac{1}{N_c - 1} \sum_{c}^{N_c} \left(\bar{y}_c - \frac{1}{N_c} \sum_{c}^{N_c} \bar{y}_c \right)^2, \tag{5.5}$$

and the mean of the cth chain is

$$\bar{y}_c = \frac{1}{N_s} \sum_{s}^{N_s} y_s. \tag{5.6}$$

This definition of B differs from that commonly found in the literature by a factor of N_s, which is absorbed into the definition of $\hat{R}$. In a well-mixed chain the PSRF should be close to unity.

Because of the labeling degeneracy in probabilistic cataloging, there is an $N!$-fold degeneracy in the likelihood of a member model with N elements. For any reasonably large N, $N!$ is larger than the number of samples that can be drawn from the posterior. Therefore, formal convergence is not possible, but also unnecessary, since well sampling only one of the degenerate likelihood peaks reveals the unique likelihood topology of the problem.

When sampling from a target probability distribution by constructing Markov chains, one needs to ensure that all walkers reach the same stationary distribution. Because the chain is transdimensional, however, the PSRF is calculated on the forward-model. Figure 5.2 illustrates the PSRF distribution obtained in Section 4, which implies that the inference has a $\lesssim 10\%$ uncertainty that adds

in quadrature with the statistical uncertainty derived by assuming that all chains have sampled from the same target distribution, i.e., the posterior of the metamodel.

5.3.2 AUTOCORRELATION

In MCMC, proposals are made to change the current state by some random, but small amount. Therefore, nearby samples in the chain will be similar. In order to retain the Markovian property, the chain is thinned by a factor equal to the autocorrelation time of the resulting chain. Given the high dimensionality, ~ 1000, of probabilistic cataloging, this thinning factor is typically $\sim 10^4$.

The autocorrelation of the chain is computed using a similar method to that followed to calculate
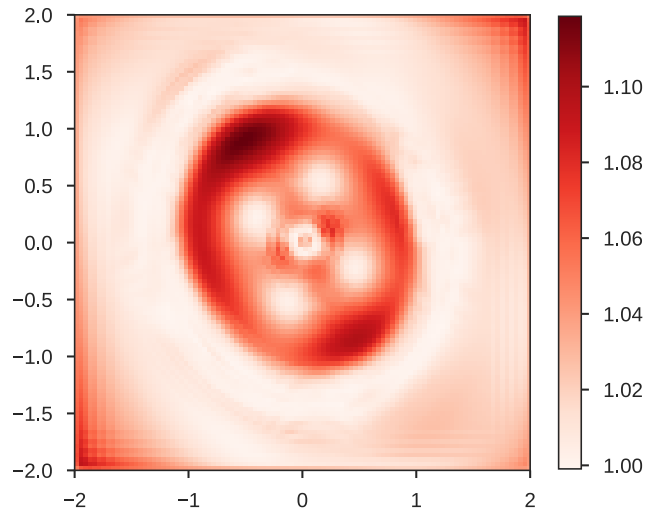
Figure 5.2: PSRF evaluated in each pixel in an inference of lensed emission metamodel.
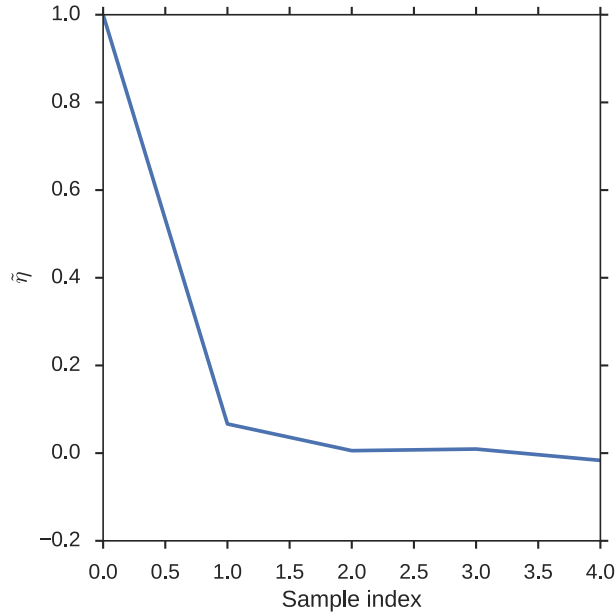
Figure 5.3: Autocorrelation of the diluted chain normalized by its value at zero lag.

the PSRF of the forward-model. After calculating the autocorrelation of each common parameter and the forward-model in each data cube, the maximum autocorrelation is determined. The auto-correlation of the chain obtained in Section 3.2 is plotted in Figure 5.3, which shows that samples in the diluted chain is memoryless.

5.4   SYSTEMATICS

5.4.1   COMPLETENESS AND FALSE DISCOVERY RATE

Using a simulated data set, the posterior of the metamodel can be assessed by comparing the posterior of the metamodel with the parameters of the true metamodel from which the simulated data

was drawn. An important part of this assessment is the comparison of the posterior of the catalog space with the elements of the true metamodel. In this context, the elements of the true metamodel are treated as a traditional catalog.

When comparing the proximity of a sample catalog with a reference catalog, there are two relevant quantities: the *completeness* and the *false discovery rate*, i.e., Type I error. The completeness, $\mathcal{C}$, is given by

$$\mathcal{C} = \frac{N_{ta}}{N_{ref}}, \tag{5.7}$$

where N_{ref} is the number of elements in the reference catalog and N_{ca} is the number of correct associations. In contrast, the false discovery rate, $\mathcal{F}$, is

$$\mathcal{F} = \frac{N_{fa}}{N_{sam}}, \tag{5.8}$$

where N_{sam} is the number of elements in the sample catalog and N_{fa} is the number of false associations.

The associations with the true catalog is performed as discussed in Section 2.2.3. Figure 5.4 shows the posterior of the completeness (left) and false discovery rate (right) when the regularization prior is applied. Posterior completeness is unity at high flux, low color, and high significance, whereas it drops to zero at low flux, high color, and low significance. The false discovery rate shows the opposite behavior as expected. Here, the true color distribution is such that soft sources are mostly

subthreshold. The transition occurs at a significance of ~ 20 despite the fact that the true catalog is highly crowded (i.e., on average one true source inside $2°$ by $2°$ region).

When the regularization prior is lifted, the false discovery rate generally increases and the completeness drops. This is shown in Figure 5.5.

Completeness and false discovery rate also depend on the association scale, which allows the discrimination threshold (i.e., level of degradation) to be controlled, placing the sampler at different points on the Receiver Operating Characteristics (ROC). Figure 5.6 shows the effect of changing the association radius.

5.4.2 Break in the amplitude distribution

An important feature of probabilistic cataloging is its ability to use subthreshold information in the data. Figure 5.7 presents the mean of the posterior histogram of point source fluxes and significances (blue), the posterior of the hyperdistribution (black), and the true catalog (green), when the true metamodel has a *broken* flux distribution with a break at 3×10^{-8} cm^{-2}s^{-1}GeV^{-1}. As a result, the posterior of the lower slope of the flux distribution is only informed by subthreshold and marginally significant point sources. Nevertheless, the inference does recover the break in the posterior, although the application of the regularization prior (left) causes the slope to be predicted higher.
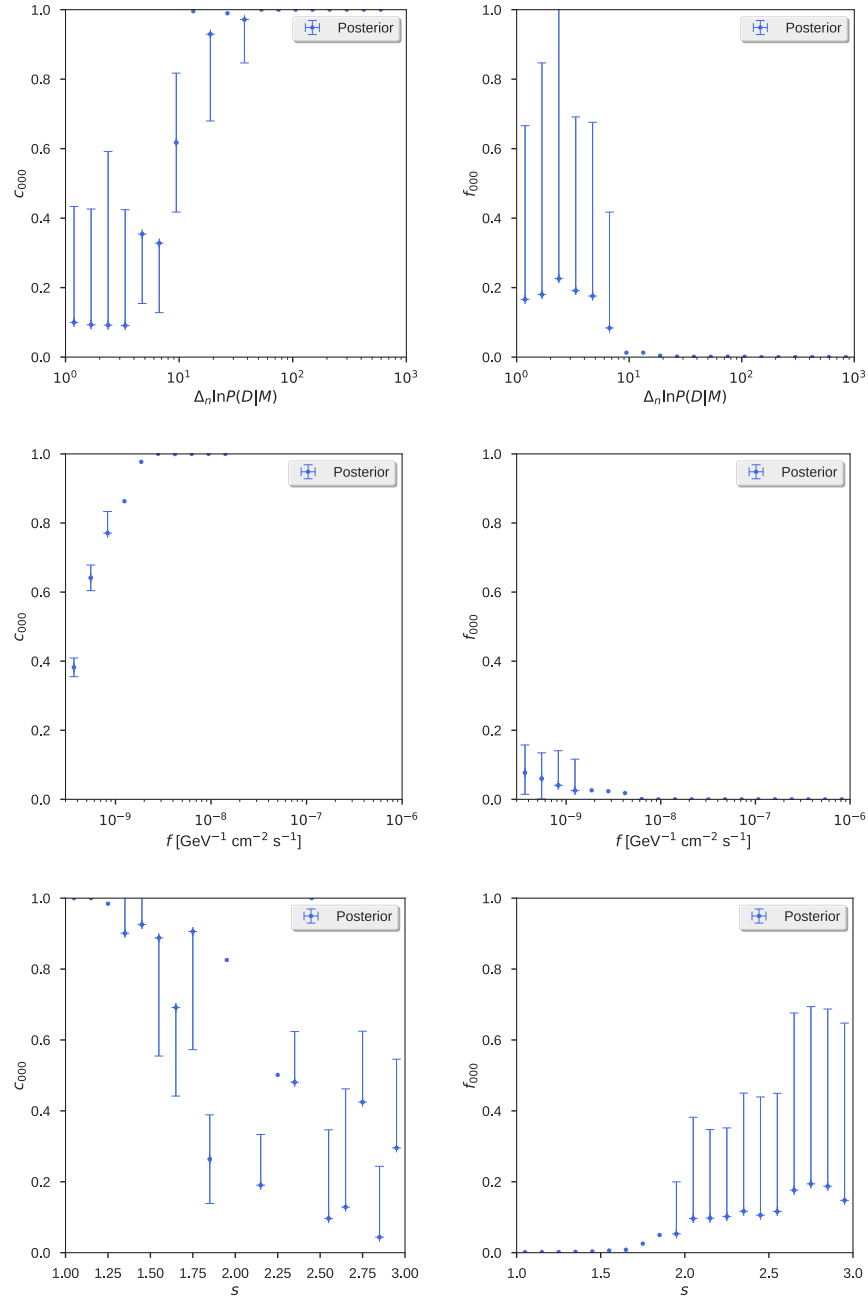
Figure 5.4: Posterior of the completeness (left) and false discovery rate (right) as a function of point source significance (top), flux (middle), and color (bottom). Point source parameters on the horizontal axes correspond to the reference (i.e., true) and the model point sources for the completeness and false discovery rate, respectively.
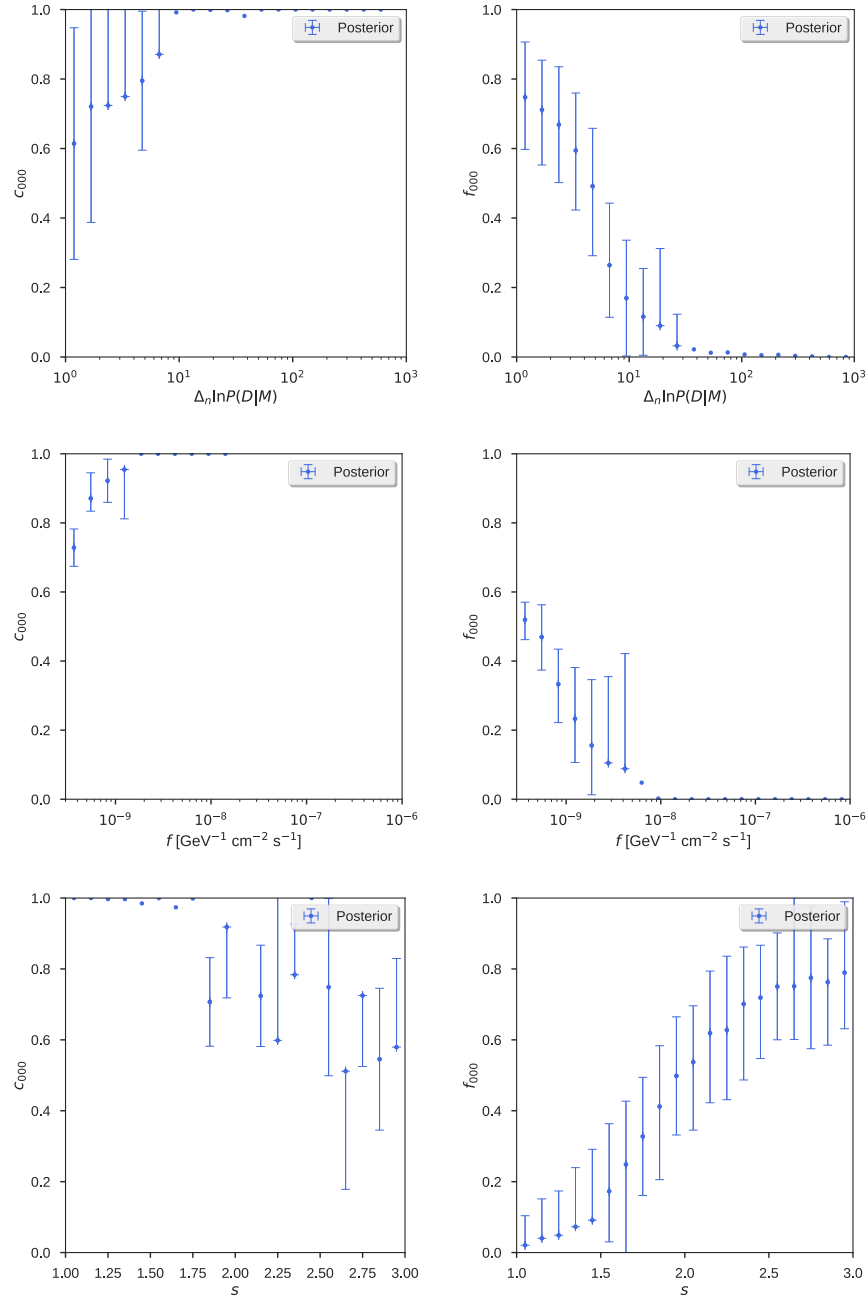
Figure 5.5: Similar to Figure 5.4, but when the regularization prior is not applied.
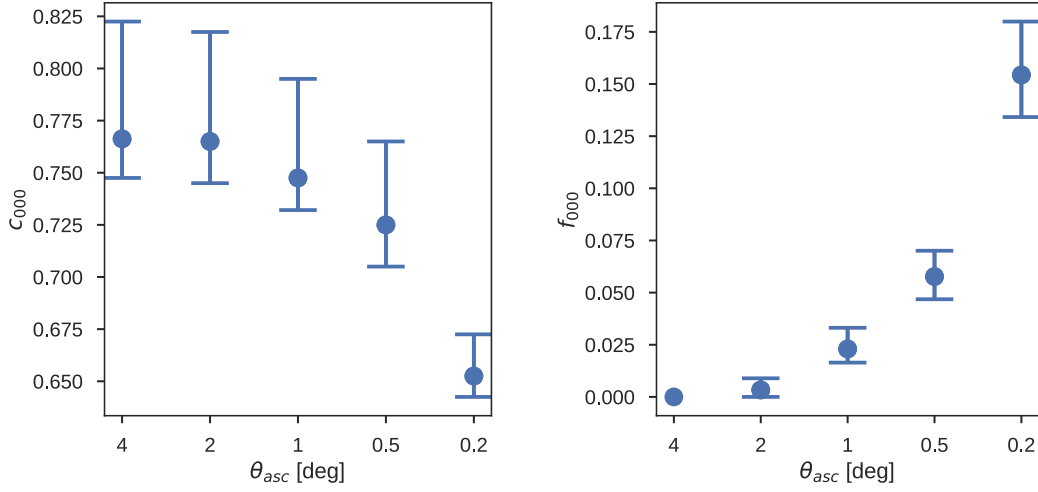
Figure 5.6: Posterior of the Completeness and false discovery rate as a function of the association scale.

5.4.3 VARYING THE REGULARIZATION PRIOR

A regularization prior of -1/2 per degree of freedom is motivated for the case when the residuals are independent, identically distributed, and Gaussian. In general, however, there can be degeneracies between the element parameters, which reduces the effective number of parameters of elements. Furthermore, residuals may deviate from the χ^2 distribution. Lastly, because this prior is intended to reduce model complexity, its exact value should not change the results of the inference, as long as the posterior is corrected for the effect of the regularization prior.

Figure 5.8 shows the posterior of the number of elements (top left), slope of the flux distribution (top right), and the background normalization (bottom), when the regularization prior is scaled by a factor α_p such that
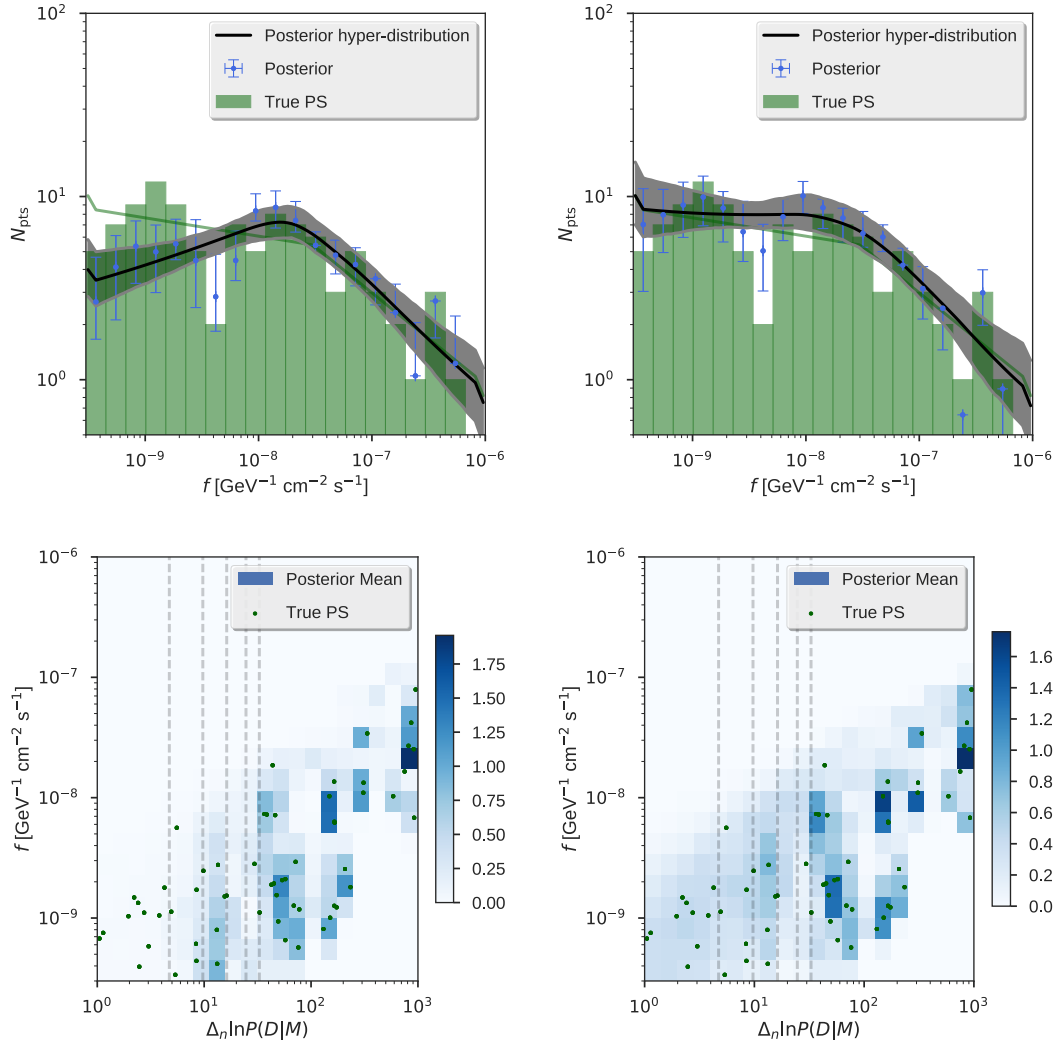
Figure 5.7: Top: Posterior (blue) of the flux histogram when the true flux distribution (green) has a break, imposing (left) and not imposing the regularization prior (right). Bottom: Mean of the posterior of the joint histogram of the flux and significance imposing (left) and not imposing (right) the regularization prior. True point sources are shown with green markers.

$$\ln P(N) = -\frac{1}{2}\alpha_p N_{ep} N. \tag{5.9}$$

The true values of the parameters are shown with the green lines. Posterior of the number of elements decreases as α_p increases. Similarly, the slope of the flux distribution decreases (i.e, flux distribution becomes shallower) and the background normalization increases as α_p increases. These posteriors agree with the true metamodel at $\alpha_p = 0.5$. This is largely due to the fact that the likelihood is Poisson instead of Gaussian, which causes larger uncertainties in the forward-model and require a smaller penalty for more complex models.

However, in general, the posterior of different parameters can agree with the true metamodel at different values of α_p, because they may have different covariances with the model indicator. Therefore, the correction that must be applied to the posterior to undo the effect of the parsimony prior depends on the parameter of interest as in the Focused Information Criterion (FIC)[56].
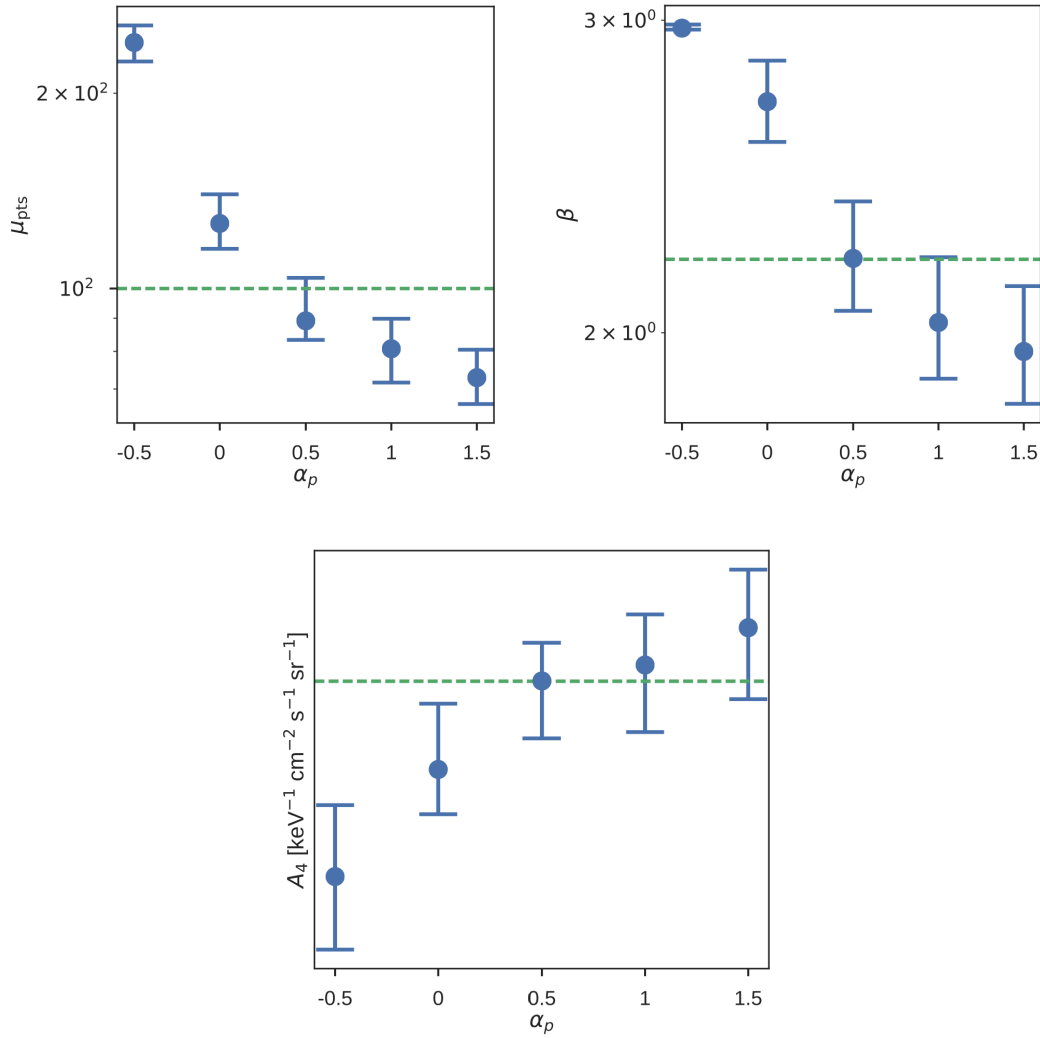
Figure 5.8: Posterior mean number of elements (top left), flux distribution slope (top right), and background normalization (bottom), when the strength of the regularization prior is scaled to different values.

6
Conclusion

This thesis implements a transdimensional, threshold-free, hierarchical, and Bayesian inference framework. It is based on the foundations of statistics, i.e., the epistemological study of how inference should be made, but also relates to contemporary problems in cosmology. Its central idea is that proper uncertainty propagation and marginalization is a prerequisite to performing robust inference.

Although there is growing evidence for the dark matter hypothesis based on its gravitational interactions, an observation that can only be explained by the non-gravitational interaction of dark matter would yield substantial evidence for its existence. Obtaining this evidence, if there will be any data to support it, requires robust error propagation and marginalization over nuisance parameters. The thesis applies probabilistic cataloging to two such data sets and explores their inference potential. The thesis focuses on the indirect detection of WIMPs not because WIMPs are theoretically more justified than other dark matter candidates. In the absence of data that yields substantial evidence for a particular model, any model choice is largely prior-driven.

The thesis also explores the notion of model complexity. The parsimony principle is the view that the best description of some data is the one that uses the fewest symbols (i.e., minimum description length) to explain it. Regularization of model complexity has implications for any inference using the scientific method. In particular, regularization of model complexity has applications from reducing overfitting in Q-learning, e.g., constructing intelligent agents to take optimal actions without exploiting only short-term benefits, to obtaining latent representations of large data sets. It is plausible to expect that the increase in the computational resources will allow a broader range of inferences to use ideas in probabilistic cataloging to take into account transdimensional modeling covariances.

References

[1] K. Abazajian, G. M. Fuller, and M. Patel. Sterile neutrino hot, warm, and cold dark matter. *Phys. Rev. D*, 64(2):023501, July 2001.

[2] K. N. Abazajian and M. Kaplinghat. Detection of a gamma-ray source in the Galactic Center consistent with extended emission from dark matter annihilation and concentrated astrophysical emission. *Phys. Rev. D*, 86(8):083511, October 2012.

[3] A. A. Abdo et al. The Fermi-LAT High-Latitude Survey: Source Count Distributions And The Origin Of The Extragalactic Diffuse Background. *The Astrophysical Journal*, 720(1):435–453, sep 2010.

[4] F. Acero et al. Fermi Large Area Telescope Third Source Catalog. *The Astrophysical Journal Supplement Series*, 218(2):23, 2015.

[5] F. Acero et al. Development Of The Model Of Galactic Interstellar Emission For Standard Point-Source Analysis Of Fermi Large Area Telescope Data. *The Astrophysical Journal Supplement Series*, 223(2):26, apr 2016.

[6] Lotty Ackerman, Matthew R. Buckley, Sean M. Carroll, and Marc Kamionkowski. Dark Matter and Dark Radiation. *Phys. Rev. D*, 79:023519, 2009.

[7] M. Ackermann et al. Constraints On The Cosmic-Ray Density Gradient Beyond The Solar Circle From Fermi γ-Ray Observations Of The Third Galactic Quadrant. *The Astrophysical Journal*, 726(2):81, 2011.

[8] M. Ackermann et al. Determination Of The Point-Spread Function For The Fermi Large Area Telescope From On-Orbit Data And Limits On Pair Halos Of Active Galactic Nuclei. *The Astrophysical Journal*, 765(1):54, mar 2013.

[9] M. Ackermann et al. Searching for Dark Matter Annihilation from Milky Way Dwarf Spheroidal Galaxies with Six Years of Fermi Large Area Telescope Data. *Physical Review Letters*, 115(23):231301, December 2015.

[10] P. A. R. Ade, N. Aghanim, M. I. R. Alves, C. Armitage-Caplan, M. Arnaud, M. Ashdown, F. Atrio-Barandela, J. Aumont, H. Aussel, and et al. Planck 2013 results. I. Overview of products and scientific results. *A&A*, 571:A1, November 2014.

[11] M. L. Ahnen et al. Limits to dark matter annihilation cross-section from a combined analysis of MAGIC and Fermi-LAT observations of dwarf satellite galaxies. *J. Cosmology Astropart. Phys.*, 2:039, February 2016.

[12] M. Ajello et al. Fermi-LAT Observations of High-Energy Gamma-Ray Emission toward the Galactic Center. *ApJ*, 819:44, March 2016.

[13] A. Albert et al. Searching for Dark Matter Annihilation in Recently Discovered Milky Way Satellites with Fermi-Lat. *ApJ*, 834:110, January 2017.

[14] Andreas Albrecht and Paul J. Steinhardt. Cosmology for grand unified theories with radiatively induced symmetry breaking. *Phys. Rev. Lett.*, 48:1220–1223, Apr 1982.

[15] C. Alcock, R. A. Allsman, D. R. Alves, T. S. Axelrod, A. C. Becker, D. P. Bennett, K. H. Cook, N. Dalal, A. J. Drake, K. C. Freeman, M. Geha, K. Griest, M. J. Lehner, S. L. Marshall, D. Minniti, C. A. Nelson, B. A. Peterson, P. Popowski, M. R. Pratt, P. J. Quinn, C. W. Stubbs, W. Sutherland, A. B. Tomaney, T. Vandehei, and D. Welch. The MACHO Project: Microlensing Results from 5.7 Years of Large Magellanic Cloud Observations. *ApJ*, 542:281–307, October 2000.

[16] R. A. Alpher, H. Bethe, and G. Gamow. The origin of chemical elements. *Phys. Rev.*, 73:803–804, Apr 1948.

[17] N. C. Amorisco, J. Zavala, and T. J. L. de Boer. Dark Matter Cores in the Fornax and Sculptor Dwarf Galaxies: Joining Halo Assembly and Detailed Star Formation Histories. *ApJL*, 782:L39, February 2014.

[18] T. Appelquist, H.-C. Cheng, and B. A. Dobrescu. Bounds on universal extra dimensions. *Phys. Rev. D*, 64(3):035002, August 2001.

[19] E. Aprile et al. First dark matter search results from the xenon1t experiment. *Phys. Rev. Lett.*, 119:181301, Oct 2017.

[20] N. Arkani-Hamed, A. G. Cohen, and H. Georgi. Electroweak symmetry breaking from dimensional deconstruction. *Physics Letters B*, 513:232–240, July 2001.

[21] W. B. Atwood, A. A. Abdo, M. Ackermann, W. Althouse, B. Anderson, M. Axelsson, L. Baldini, J. Ballet, D. L. Band, G. Barbiellini, and et al. The Large Area Telescope on the Fermi Gamma-Ray Space Telescope Mission. *ApJ*, 697:1071–1102, June 2009.

[22] Edward A Baltz, Phil Marshall, and Masamune Oguri. Analytic models of plausible gravitational lens potentials. *Journal of Cosmology and Astroparticle Physics*, 2009(01):015–015, jan 2009.

[23] Richard Bartels, Suraj Krishnamurthy, and Christoph Weniger. Strong Support for the Millisecond Pulsar Origin of the Galactic Center GeV Excess. *Physical Review Letters*, 116(5), 2016.

[24] Stefanie I. Becker. Guidance of attention by feature relationships: The end of the road for feature map theories? *Current Trends in Eye Tracking Research*, 19(6):37–49, 2014.

[25] J. D. Bekenstein. Relativistic gravitation theory for the modified Newtonian dynamics paradigm. *Phys. Rev. D*, 70(8):083509, October 2004.

[26] G. Benbassat. La Reconnaissance Automatique De La Parole. *Revue de Laryngologie Otologie Rhinologie*, 111(4):389–392, 1990.

[27] P. S. Bhupal Dev, A. Mazumdar, and S. Qutub. Constraining Non-thermal and Thermal properties of Dark Matter. *ArXiv e-prints*, November 2013.

[28] S. Birrer, A. Amara, and A. Refregier . Gravitational Lens Modeling with Basis Sets. *ApJ*, 813:102, November 2015.

[29] S. Birrer, A. Amara, and A. Refregier. Lensing substructure quantification in RXJ1131-1231: a 2 keV lower bound on dark matter thermal relic mass. *J. Cosmology Astropart. Phys.*, 5:037, May 2017.

[30] George R. Blumenthal, S.M. Faber, Joel R. Primack, and Martin J. Rees. Formation of Galaxies and Large Scale Structure with Cold Dark Matter. *Nature*, 311:517–525, 1984.

[31] George R. Blumenthal, Heinz Pagels, and Joel R. Primack. GALAXY FORMATION BY DISSIPATIONLESS PARTICLES HEAVIER THAN NEUTRINOS. *Nature*, 299:37–38, 1982.

[32] Paul Bode, Jeremiah P. Ostriker, and Neil Turok. Halo formation in warm dark matter models. *Astrophys. J.*, 556:93–107, 2001.

[33] Celine Boehm, Alain Riazuelo, Steen H. Hansen, and Richard Schaeffer. Interacting dark matter disguised as warm dark matter. *Phys. Rev. D*, 66:083505, 2002.

[34] Adam S. Bolton, Scott Burles, Léon V. E. Koopmans, Tommaso Treu, Raphaël Gavazzi, Leonidas A. Moustakas, Randall Wayth, and David J. Schlegel. The Sloan Lens ACS Survey. V. The Full ACS Strong-Lens Sample. *The Astrophysical Journal*, 682(2):964–984, may 2008.

[35] Adam S. Bolton, Scott Burles, Leon V. E. Koopmans, Tommaso Treu, and Leonidas A. Moustakas. The Sloan Lens ACS Survey. I. A Large Spectroscopically Selected Sample of Massive Early-Type Lens Galaxies. *The Astrophysical Journal*, 638:703, nov 2005.

[36] J. R. Bond and A. S. Szalay. The Collisionless Damping of Density Fluctuations in an Expanding Universe. *ApJ*, 274:443–468, 1983.

[37] D. Boyanovsky, H. J. de Vega, and N. Sanchez. The dark matter transfer function: free streaming, particle statistics and memory of gravitational clustering. *Phys. Rev. D*, 78:063546, 2008.

[38] Daniel Boyanovsky and Jun Wu. Small scale aspects of warm dark matter : power spectra and acoustic oscillations. *Phys. Rev. D*, 83:043524, 2011.

[39] M. Bradač, M. Lombardi, and P. Schneider. Mass-sheet degeneracy: Fundamental limit on the cluster mass reconstruction from statistical (weak) lensing. *Astronomy and Astrophysics*, 424(1):13–22, may 2004.

[40] T. D. Brandt and B. Kocsis. Disrupted Globular Clusters Can Explain the Galactic Center Gamma-Ray Excess. *ApJ*, 812:15, October 2015.

[41] Brendon J. Brewer, Daniel Foreman-Mackey, and David W. Hogg. Probabilistic Catalogs for Crowded Stellar Fields. *The Astronomical Journal*, 146(1):7, 2013.

[42] Brendon J. Brewer, David Huijser, and Geraint F. Lewis. Trans-Dimensional Bayesian Inference for Gravitational Lens Substructures. *Monthly Notices of the Royal Astronomical Society*, 455(2):1819–1829, aug 2015.

[43] A. M. Brooks, M. Kuhlen, A. Zolotov, and D. Hooper. A Baryonic Solution to the Missing Satellites Problem. *ApJ*, 765:22, March 2013.

[44] Matthew R. Buckley, Jesus Zavala, Francis-Yan Cyr-Racine, Kris Sigurdson, and Mark Vogels-berger. Scattering, Damping, and Acoustic Oscillations: Simulating the Structure of Dark Matter Halos with Relativistic Force Carriers. *Phys. Rev. D*, 90(4):043524, 2014.

[45] P. Bull et al. Beyond ΛCDM: Problems, solutions, and the road ahead. *Physics of the Dark Universe*, 12:56–99, dec 2016.

[46] E. Margaret Burbidge, G. R. Burbidge, William A. Fowler, and F. Hoyle. Synthesis of the elements in stars. *Rev. Mod. Phys.*, 29:547–650, Oct 1957.

[47] A. Burkert. The Structure of Dark Matter Halos in Dwarf Galaxies. *ApJL*, 447:L25–L28, July 1995.

[48] Francesca Calore, Ilias Cholis, Christopher McCabe, and Christoph Weniger. A tale of tails: Dark matter interpretations of the Fermi GeV excess in light of background model systematics. *Physical Review D - Particles, Fields, Gravitation and Cosmology*, 91(6), nov 2015.

[49] Francesca Calore, Ilias Cholis, and Christoph Weniger. Background model systematics for the Fermi GeV excess. *Journal of Cosmology and Astroparticle Physics*, 2015(03):038–038, mar 2014.

[50] E. Carlson and S. Profumo. Cosmic ray protons in the inner Galaxy and the Galactic Center gamma-ray excess. *Phys. Rev. D*, 90(2):023015, July 2014.

[51] E. D. Carlson, M. E. Machacek, and L. J. Hall. Self-interacting dark matter. *ApJ*, 398:43–52, October 1992.

[52] H.-C. Cheng, J. L. Feng, and K. T. Matchev. Kaluza-Klein Dark Matter. *Physical Review Letters*, 89(21):211301, October 2002.

[53] M. Chiba. Probing Dark Matter Substructure in Lens Galaxies. *ApJ*, 565:17–23, January 2002.

[54] Xiaoyong Chu and Basudeb Dasgupta. Dark Radiation Alleviates Problems with Dark Matter Halos. *Phys. Rev. Lett.*, 113(16):161301, 2014.

[55] L Ciotti and G Bertin. Analytical properties of the R$^{1/m}$ law. *A&A*, 352:447–451, dec 1999.

[56] Gerda Claeskens and Nils Lid Hjort. The focused information criterion. *Journal of the American Statistical Association*, 98(464):900–916, 2003.

[57] D. Clowe, M. Bradač, A. H. Gonzalez, M. Markevitch, S. W. Randall, C. Jones, and D. Zaritsky. A Direct Empirical Proof of the Existence of Dark Matter. *ApJL*, 648:L109–L113, September 2006.

[58] Sidney Coleman. Q-balls. *Nuclear Physics B*, 262(2):263 – 283, 1985.

[59] P. Colín, V. Avila-Reese, O. Valenzuela, and C. Firmani. Structure and Subhalo Population of Halos in a Self-interacting Dark Matter Cosmology. *ApJ*, 581:777–793, December 2002.

[60] M. Colless, G. Dalton, S. Maddox, W. Sutherland, P. Norberg, S. Cole, J. Bland-Hawthorn, T. Bridges, R. Cannon, C. Collins, W. Couch, N. Cross, K. Deeley, R. De Propris, S. P. Driver, G. Efstathiou, R. S. Ellis, C. S. Frenk, K. Glazebrook, C. Jackson, O. Lahav, I. Lewis, S. Lumsden, D. Madgwick, J. A. Peacock, B. A. Peterson, I. Price, M. Seaborne, and K. Taylor. The 2dF Galaxy Redshift Survey: spectra and redshifts. *MNRAS*, 328:1039–1063, December 2001.

[61] J. J. Condon, W. D. Cotton, E. B. Fomalont, K. I. Kellermann, N. Miller, R. A. Perley, D. Scott, T. Vernstrom, and J. V. Wall. Resolving the Radio Source Background: Deeper Understanding Through Confusion. *The Astrophysical Journal*, 758(1):23, 2012.

[62] P. Creasey, O. Sameie, L. V. Sales, H.-B. Yu, M. Vogelsberger, and J. Zavala. Spreading out and staying sharp - creating diverse rotation curves via baryonic and self-interaction effects. *MNRAS*, 468:2283–2295, June 2017.

[63] M. Cruz, E. Martínez-González, P. Vielva, and L. Cayón. Detection of a non-Gaussian spot in WMAP. *MNRAS*, 356:29–40, January 2005.

[64] R. H. Cyburt, B. D. Fields, K. A. Olive, and T.-H. Yeh. Big bang nucleosynthesis: Present status. *Reviews of Modern Physics*, 88(1):015004, January 2016.

[65] Francis Yan Cyr-Racine, Leonidas A. Moustakas, Charles R. Keeton, Kris Sigurdson, and Daniel A. Gilman. Dark census: Statistically detecting the satellite populations of distant galaxies. *Physical Review D*, 94(4):043505, jun 2016.

[66] Francis-Yan Cyr-Racine and Kris Sigurdson. The cosmology of atomic dark matter. *Phys. Rev. D*, 87:103515, 2013.

[67] N. Dalal and C. S. Kochanek. Direct Detection of Cold Dark Matter Substructure. *The Astrophysical Journal*, 572(1):25–33, nov 2002.

[68] Neal Dalal and Christopher S. Kochanek. Strong lensing constraints on small-scale linear power. 2002.

[69] Julianne J. Dalcanton and Craig J. Hogan. Halo cores and phase space densities: Observational constraints on dark matter physics and structure formation. *Astrophys. J.*, 561:35–45, 2001.

[70] R. Davé, D. N. Spergel, P. J. Steinhardt, and B. D. Wandelt. Halo Properties in Cosmological Simulations of Self-interacting Cold Dark Matter. *ApJ*, 547:574–589, February 2001.

[71] M. Davis, M. Lecar, C. Pryor, and E. Witten. The formation of galaxies from massive neutrinos. *ApJ*, 250:423–431, November 1981.

[72] Marc Davis, George Efstathiou, Carlos S. Frenk, and Simon D.M. White. The Evolution of Large Scale Structure in a Universe Dominated by Cold Dark Matter. *ApJ*, 292:371–394, 1985.

[73] T. Daylan, F.-Y. Cyr-Racine, A. Diaz Rivero, C. Dvorkin, and D. P. Finkbeiner. Probing the Small-scale Structure in Strongly Lensed Systems via Transdimensional Inference. *ApJ*, 854:141, February 2018.

[74] T. Daylan, S. K. N. Portillo, and D. P. Finkbeiner. Inference of Unresolved Point Sources at High Galactic Latitudes Using Probabilistic Catalogs. *ApJ*, 839:4, April 2017.

[75] Tansu Daylan, Douglas P. Finkbeiner, Dan Hooper, Tim Linden, Stephen K N Portillo, Nicholas L. Rodd, and Tracy R. Slatyer. The characterization of the gamma-ray signal from the central Milky Way: A case for annihilating dark matter. *Physics of the Dark Universe*, 12:1–23, 2016.

[76] P. de Bernardis et al. A flat Universe from high-resolution maps of the cosmic microwave background radiation. *Nature*, 404:955–959, April 2000.

[77] G. de Vaucouleurs. Recherches sur le nebuleuses extragalactiques. *Annales d'Astrophysique*, 11:247, jan 1948.

[78] Antonino Del Popolo and Morgan Le Delliou. Small scale problems of the ΛCDM model: a short review. *Galaxies*, 5(1):17, 2016.

[79] Charles Dennison Dermer and Berrie Giebels. Active galactic nuclei at gamma-ray energies. page 27, feb 2016.

[80] G. Despali and S. Vegetti. The impact of baryonic physics on the subhalo mass function and implications for gravitational lensing. *MNRAS*, 469:1997–2010, August 2017.

[81] R. H. Dicke, P. J. E. Peebles, P. G. Roll, and D. T. Wilkinson. Cosmic Black-Body Radiation. *The Astrophysical Journal*, 142(9):414, 1965.

[82] S. Dodelson and L. M. Widrow. Sterile neutrinos as dark matter. *Physical Review Letters*, 72:17–20, January 1994.

[83] Andrzej K. Drukier, Katherine Freese, and David N. Spergel. Detecting cold dark-matter candidates. *Phys. Rev. D*, 33:3495–3508, Jun 1986.

[84] John Dubinski and R.G. Carlberg. The structure of cold dark matter halos. \apj, 378:496, 1991.

[85] Albert Einstein. Explanation of the perihelical motion of mercury from the general theory of relativity. *St{ä}ndiger Beobachter der Preussischen Akademie der Wissenschaften, Part 2*, pages 831–839, 1915.

[86] D. J. Eisenstein et al. Detection of the Baryon Acoustic Peak in the Large-Scale Correlation Function of SDSS Luminous Red Galaxies. *ApJ*, 633:560–574, November 2005.

[87] D. J. Eisenstein, D. H. Weinberg, E. Agol, H. Aihara, C. Allende Prieto, S. F. Anderson, J. A. Arns, É. Aubourg, S. Bailey, E. Balbinot, and et al. SDSS-III: Massive Spectroscopic Surveys of the Distant Universe, the Milky Way, and Extra-Solar Planetary Systems. *AJ*, 142:72, September 2011.

[88] J. Ellis, J. E. Kim, and D. V. Nanopoulos. Cosmological gravitino regeneration and decay. *Physics Letters B*, 145:181–186, September 1984.

[89] F. Englert and R. Brout. Broken symmetry and the mass of gauge vector mesons. *Phys. Rev. Lett.*, 13:321–323, Aug 1964.

[90] A. C. Fabian and X. Barcons. The origin of the X-ray background. *Annual Review of Astronomy and Astrophysics*, 30:429–456, 1992.

[91] Ross Fadely and Charles R. Keeton. Substructure in the lens HE 0435-1223. sep 2011.

[92] E. E. Falco, M. V. Gorenstein, and I. I. Shapiro. On model-dependent bounds on H(0) from gravitational images Application of Q0957 + 561A,B. *The Astrophysical Journal*, 289:L1, 1985.

[93] Claude-André Faucher-Giguère and Abraham Loeb. The pulsar contribution to the gamma-ray background. *Journal of Cosmology and Astroparticle Physics*, 2010(01):005–005, jan 2010.

[94] Jonathan L. Feng, Manoj Kaplinghat, Huitzu Tu, and Hai-Bo Yu. Hidden Charged Dark Matter. *JCAP*, 0907:004, 2009.

[95] Alex Fitts et al. FIRE in the Field: Simulating the Threshold of Galaxy Formation. 2016.

[96] Ricardo A. Flores and Joel R. Primack. Observational and theoretical constraints on singular dark matter halos. *Astrophys. J.*, 427:L1–4, 1994.

[97] W. Forman, C. Jones, and W. Tucker. Hot coronae around early-type galaxies. *ApJ*, 293:102–119, June 1985.

[98] K. Freese, B. Fields, and D. Graff. Limits on stellar objects as the dark matter of our halo: nonbaryonic dark matter seems to be required. *Nuclear Physics B Proceedings Supplements*, 80:03, January 2000.

[99] R. Friedberg, T. D. Lee, and A. Sirlin. Class of scalar-field soliton solutions in three space dimensions. *Phys. Rev. D*, 13:2739–2761, May 1976.

[100] Yasmin Friedmann and François Bouchet. Fluctuation analysis of the far-infrared background - information from the confusion. *Monthly Notices of the Royal Astronomical Society*, 348(3):737–744, 2004.

[101] L. Gao, C. S. Frenk, M. Boylan-Kolchin, A. Jenkins, V. Springel, and S. D.M. White. The statistics of the subhalo abundance of dark matter haloes. *Monthly Notices of the Royal Astronomical Society*, 410(4):2309–2314, jun 2011.

[102] S. Garrison-Kimmel, A. R. Wetzel, J. S. Bullock, P. F. Hopkins, M. Boylan-Kolchin, C.-A. Faucher-Giguere, D. Keres, E. Quataert, R. E. Sanderson, A. S. Graus, and T. Kelley. Not so lumpy after all: modeling the depletion of dark matter subhalos by Milky Way-like galaxies. *ArXiv e-prints*, January 2017.

[103] M. J. Geller and J. P. Huchra. Mapping the universe. *Science*, 246:897–903, November 1989.

[104] Andrew Gelman and Donald B. Rubin. Inference from Iterative Simulation Using Multiple Sequences, 1992.

[105] Shy Genel, Mark Vogelsberger, Volker Springel, Debora Sijacki, Dylan Nelson, Greg Snyder, Vicente Rodriguez-Gomez, Paul Torrey, and Lars Hernquist. Introducing the Illustris project: The evolution of galaxy populations across cosmic time. *Monthly Notices of the Royal Astronomical Society*, 445(1):175–200, may 2014.

[106] G. Gentile, P. Salucci, U. Klein, D. Vergani, and P. Kalberla. The cored distribution of dark matter in spiral galaxies. *MNRAS*, 351:903–922, July 2004.

[107] Howard Georgi and A. Pais. Vacuum symmetry and the pseudo-goldstone phenomenon. *Phys. Rev. D*, 12:508–512, Jul 1975.

[108] Daniel Gilman, Adriano Agnello, Tommaso Treu, Charles R. Keeton, and Anna M. Nierenberg. Strong lensing signatures of luminous structure and substructure in early-type galaxies. oct 2016.

[109] Daniel Gilman, Adriano Agnello, Tommaso Treu, Charles R. Keeton, and Anna M. Nierenberg. Strong lensing signatures of luminous structure and substructure in early-type galaxies. *Monthly Notices of the Royal Astronomical Society*, page stx158, January 2017. arXiv: 1610.08525.

[110] G. Golse and J.-P. Kneib. Pseudo elliptical lensing mass model: Application to the NFW mass distribution. *Astronomy and Astrophysics*, 390(3):821–827, 2002.

[111] L. Goodenough and D. Hooper. Possible Evidence For Dark Matter Annihilation In The Inner Milky Way From The Fermi Gamma Ray Space Telescope. *ArXiv e-prints*, October 2009.

[112] C. Gordon and O. Macías. Dark matter and pulsar model constraints from Galactic Center Fermi-LAT gamma-ray observations. *Phys. Rev. D*, 88(8):083521, October 2013.

[113] B. R. Granett, M. C. Neyrinck, and I. Szapudi. An Imprint of Superstructures on the Microwave Background due to the Integrated Sachs-Wolfe Effect. *ApJL*, 683:L99, August 2008.

[114] Peter J. Green. Reversible jump Markov chain monte carlo computation and Bayesian model determination. *Biometrika*, 82(4):711–732, 1995.

[115] Kim Griest and Marc Kamionkowski. Unitarity limits on the mass and radius of dark-matter particles. *Phys. Rev. Lett.*, 64:615–618, Feb 1990.

[116] Alan H. Guth. Inflationary universe: A possible solution to the horizon and flatness problems. *Phys. Rev. D*, 23:347–356, Jan 1981.

[117] S. Hanany, P. Ade, A. Balbi, J. Bock, J. Borrill, A. Boscaleri, P. de Bernardis, P. G. Ferreira, V. V. Hristov, A. H. Jaffe, A. E. Lange, A. T. Lee, P. D. Mauskopf, C. B. Netterfield, S. Oh, E. Pascale, B. Rabii, P. L. Richards, G. F. Smoot, R. Stompor, C. D. Winant, and J. H. P. Wu. *ApJL*, 545:L5–L9, December 2000.

[118] David I. Hastie and Peter J. Green. Model choice using reversible jump Markov chain Monte Carlo. *Statistica Neerlandica*, 66(3):309–338, 2012.

[119] S. Hawking. Gravitationally collapsed objects of very low mass. *MNRAS*, 152:75, 1971.

[120] Y. Hezaveh, N. Dalal, G. Holder, T. Kisner, M. Kuhlen, and L. Perreault Levasseur. Measuring the power spectrum of dark matter substructure using strong gravitational lensing. *J. Cosmology Astropart. Phys.*, 11:048, November 2016.

[121] Yashar Hezaveh, Neal Dalal, Gilbert Holder, Michael Kuhlen, Daniel Marrone, et al. Dark Matter Substructure Detection Using Spatially Resolved Spectroscopy of Lensed Dusty Galaxies. *Astrophys. J.*, 767:9, 2013.

[122] Yashar D. Hezaveh, Neal Dalal, Daniel P. Marrone, Yao-Yuan Mao, Warren Morningstar, Di Wen, Roger D. Blandford, John E. Carlstrom, Christopher D. Fassnacht, Gilbert P. Holder, Athol Kemball, Philip J. Marshall, Norman Murray, Laurence Perreault Levasseur, Joaquin D. Vieira, and Risa H. Wechsler. Detection of lensing substructure using ALMA observations of the dusty galaxy SDP.81. *The Astrophysical Journal*, 823(1):37, jan 2016.

[123] Peter W. Higgs. Broken symmetries and the masses of gauge bosons. *Phys. Rev. Lett.*, 13:508–509, Oct 1964.

[124] D.W. Hogg and D. Lang. Telescopes don't make catalogues! *EAS Publications Series*, 45:351–358, aug 2010.

[125] D. Hooper and L. Goodenough. Dark matter annihilation in the Galactic Center as seen by the Fermi Gamma Ray Space Telescope. *Physics Letters B*, 697:412–428, March 2011.

[126] D. Hooper and T. Linden. Origin of the gamma rays from the Galactic Center. *Phys. Rev. D*, 84(12):123005, December 2011.

[127] D. Hooper and T. R. Slatyer. Two emission mechanisms in the Fermi Bubbles: A possible signal of annihilating dark matter. *Physics of the Dark Universe*, 2:118–138, September 2013.

[128] J. W. Hsueh, L. Oldham, C. Spingola, S. Vegetti, C. D. Fassnacht, M. W. Auger, L. V. E. Koopmans, J. P. McKean, and D. J. Lagattuta. SHARP - IV. An apparent flux ratio anomaly resolved by the edge-on disc in B0712+472. jan 2017.

[129] Wayne Hu, Rennan Barkana, and Andrei Gruzinov. Fuzzy cold dark matter: the wave properties of ultralight particles. *Physical Review Letters*, 85(6):1158–1161, mar 2000.

[130] L. Hui, J. P. Ostriker, S. Tremaine, and E. Witten. Ultralight scalars as cosmological dark matter. *Phys. Rev. D*, 95(4):043541, February 2017.

[131] P. Hut. Limits on masses and number of neutral weakly interacting particles. *Physics Letters B*, 69:85–88, July 1977.

[132] M. Kamionkowski and A. R. Liddle. The Dearth of Halo Dwarf Galaxies: Is There Power on Short Scales? *Physical Review Letters*, 84:4525–4528, May 2000.

[133] David E. Kaplan, Gordan Z. Krnjaic, Keith R. Rehermann, and Christopher M. Wells. Atomic Dark Matter. *J. Cosmology Astropart. Phys.*, 1005:021, 2010.

[134] A Kassiola and I Kovner. Elliptic Mass Distributions versus Elliptic Potentials in Gravitational Lenses. *apj*, 417:450, nov 1993.

[135] C. R. Keeton, C. S. Kochanek, and U. Seljak. Shear and Ellipticity in Gravitational Lenses. *The Astrophysical Journal*, 482:604, 1996.

[136] Charles R. Keeton. A Catalog of Mass Models for Gravitational Lensing. *Journal of Astronomy and Astrophysics*, (June):17, 2001.

[137] Charles R. Keeton, B. Scott Gaudi, and A. O. Petters. Identifying lenses with small-scale structure. i. cusp lenses. *Astrophys. J.*, 598:138–161, 2003.

[138] Charles R. Keeton and Leonidas A. Moustakas. A new channel for detecting dark matter substructure in galaxies: Gravitational lens time delays. *Astrophys. J.*, 699:1720–1731, 2009.

[139] Anatoly Klypin, Andrey V. Kravtsov, Octavio Valenzuela, and Francisco Prada. Where Are the Missing Galactic Satellites? *The Astrophysical Journal*, 522(1):82–92, jan 1999.

[140] L. V. E. Koopmans. Gravitational imaging of cold dark matter substructures. *MNRAS*, 363:1136–1144, November 2005.

[141] S. Kullback and R. A. Leibler. On Information and Sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86, 1951.

[142] B. W. Lee and S. Weinberg. Cosmological lower bound on heavy-neutrino masses. *Physical Review Letters*, 39:165–168, July 1977.

[143] Samuel K. Lee, Mariangela Lisanti, and Benjamin R. Safdi. Distinguishing dark matter from unresolved point sources in the Inner Galaxy with photon statistics. *Journal of Cosmology and Astroparticle Physics*, 2015(05):056–056, dec 2015.

[144] Samuel K. Lee, Mariangela Lisanti, Benjamin R. Safdi, Tracy R. Slatyer, and Wei Xue. Evidence for Unresolved γ-Ray Point Sources in the Inner Galaxy. *Physical Review Letters*, 116(5), 2016.

[145] Ran Li, Carlos S. Frenk, Shaun Cole, Liang Gao, Sownak Bose, and Wojciech A. Hellwing. Constraints on the identity of the dark matter from strong gravitational lenses. *Mon. Not. Roy. Astron. Soc.*, 460(1):363–372, 2016.

[146] Ran Li, Carlos S. Frenk, Shaun Cole, Qiao Wang, and Liang Gao. Projection effects in the strong lensing study of subhaloes. *Mon. Not. Roy. Astron. Soc.*, 468:1426, 2017.

[147] A.D. Linde. A new inflationary universe scenario: A possible solution of the horizon, flatness, homogeneity, isotropy and primordial monopole problems. *Physics Letters B*, 108(6):389 – 393, 1982.

[148] S. López, V. D'Odorico, S. L. Ellison, G. D. Becker, L. Christensen, G. Cupani, K. D. Denney, I. Pâris, G. Worseck, T. A. M. Berg, S. Cristiani, M. Dessauges-Zavadsky, M. Haehnelt, F. Hamann, J. Hennawi, V. Iršič, T.-S. Kim, P. López, R. Lund Saust, B. Ménard, S. Perrotta, J. X. Prochaska, R. Sánchez-Ramírez, M. Vestergaard, M. Viel, and L. Wisotzki. *A&A*, 594:A91, October 2016.

[149] Shude Mao and Peter Schneider. Evidence for substructure in lens galaxies? *Mon. Not. R. Astron. Soc.*, 295:587–594, 1998.

[150] M. Markevitch. Chandra Observation of the Most Interesting Cluster in the Universe. In A. Wilson, editor, *The X-ray Universe 2005*, volume 604 of *ESA Special Publication*, page 723, January 2006.

[151] Charlotte A. Mason, Tommaso Treu, Adriano Fontana, Tucker Jones, Takahiro Morishita, Ricardo Amorin, Maruša Bradač, Emily Quinn Finney, Claudio Grillo, Alaina Henry, Austin Hoag, Kuang-Han Huang, Kasper B. Schmidt, Michele Trenti, and Benedetta Vulcani. First Results from the KMOS Lens-Amplified Spectroscopic Survey (KLASS): Kinematics of Lensed Galaxies at Cosmic Noon. *The Astrophysical Journal*, 838(1):14, mar 2017.

[152] J. C. Mather, E. S. Cheng, D. A. Cottingham, R. E. Eplee, Jr., D. J. Fixsen, T. Hewagama, R. B. Isaacman, K. A. Jensen, S. S. Meyer, P. D. Noerdlinger, S. M. Read, L. P. Rosen, R. A. Shafer, E. L. Wright, C. L. Bennett, N. W. Boggess, M. G. Hauser, T. Kelsall, S. H. Moseley, Jr., R. F. Silverberg, G. F. Smoot, R. Weiss, and D. T. Wilkinson. Measurement of the cosmic microwave background spectrum by the COBE FIRAS instrument. *ApJ*, 420:439–444, January 1994.

[153] J. C. Mather, D. J. Fixsen, R. A. Shafer, C. Mosier, and D. T. Wilkinson. Calibrator Design for the COBE Far-Infrared Absolute Spectrophotometer (FIRAS). *ApJ*, 512:511–520, February 1999.

[154] R. B. Metcalf and A. Amara. Small-scale structures of dark matter and flux anomalies in quasar gravitational lenses. *MNRAS*, 419:3414–3425, February 2012.

[155] R. B. Metcalf and H. Zhao. Flux Ratios as a Probe of Dark Substructures in Quadruple-Image Gravitational Lenses. *ApJL*, 567:L5–L8, March 2002.

[156] R. Benton Metcalf and Piero Madau. Compound Gravitational Lensing as a Probe of Dark Matter Substructure within Galaxy Halos. *The Astrophysical Journal*, 563(1):9–20, dec 2001.

[157] M. Milgrom. A modification of the Newtonian dynamics as a possible alternative to the hidden mass hypothesis. *ApJ*, 270:365–370, July 1983.

[158] Q. E. Minor and M. Kaplinghat. Inflation that runs naturally: Gravitational waves and suppression of power at large and small scales. *Phys. Rev. D*, 91(6):063504, March 2015.

[159] Q. E. Minor, M. Kaplinghat, and N. Li. A robust mass estimator for dark matter subhalo perturbations in strong gravitational lenses. *ArXiv e-prints*, December 2016.

[160] J. W. Moffat. Scalar tensor vector gravity theory. *J. Cosmology Astropart. Phys.*, 3:004, March 2006.

[161] B. Moore. Evidence against dissipationless dark matter from observations of galaxy haloes. *Nature*, 370:629, 1994.

[162] Ben Moore, Sebastiano Ghigna, Fabio Governato, George Lake, Tom Quinn, Joachim Stadel, and Paolo Tozzi. Dark matter substructure in galactic halos. *ApJ*, 524:L19–L22, 1999.

[163] K Nakamura and Particle Data Group. Review of particle physics. *Journal of Physics G: Nuclear and Particle Physics*, 37(7A):075021, 2010.

[164] Julio F. Navarro, Carlos S. Frenk, and Simon D. M. White. The Structure of Cold Dark Matter Halos. *The Astrophysical Journal*, 462:563, may 1996.

[165] A. M. Nierenberg, T. Treu, G. Brammer, A. H. G. Peter, C. D. Fassnacht, C. R. Keeton, C. S. Kochanek, K. B. Schmidt, D. Sluse, and S. A. Wright. Probing dark matter substructure in the gravitational lens HE0435-1223 with the WFC3 grism. *ArXiv e-prints*, January 2017.

[166] A. M. Nierenberg, T. Treu, G. Brammer, A. H. G. Peter, C. D. Fassnacht, C. R. Keeton, C. S. Kochanek, K. B. Schmidt, D. Sluse, and S. A. Wright. Probing dark matter substructure in the gravitational lens HE0435-1223 with the WFC3 grism. jan 2017.

[167] A. M. Nierenberg, T. Treu, N. Menci, Y. Lu, Paul Torrey, and M. Vogelsberger. The Missing Satellite Problem in 3D. mar 2016.

[168] A. M. Nierenberg, T. Treu, S. A. Wright, C. D. Fassnacht, and M. W. Auger. Detection of substructure with adaptive optics integral field spectroscopy of the gravitational lens B1422+231. *MNRAS*, 442:2434–2445, August 2014.

[169] S. Nussinov. Technocosmology — could a technibaryon excess provide a "natural" missing mass candidate? *Physics Letters B*, 165(1):55 – 58, 1985.

[170] Masamune Oguri and Philip J. Marshall. Gravitationally lensed quasars and supernovae in future wide-field optical imaging surveys. *Mon. Not. Roy. Astron. Soc.*, 405:2579–2593, 2010.

[171] S.-H. Oh, D. A. Hunter, E. Brinks, B. G. Elmegreen, A. Schruba, F. Walter, M. P. Rupen, L. M. Young, C. E. Simpson, M. C. Johnson, K. A. Herrmann, D. Ficut-Vicas, P. Cigan, V. Heesen, T. Ashley, and H.-X. Zhang. High-resolution Mass Models of Dwarf Galaxies from LITTLE THINGS. *AJ*, 149:180, June 2015.

[172] N. Palanque-Delabrouille, C. Yèche, J. Baur, C. Magneville, G. Rossi, J. Lesgourgues, A. Borde, E. Burtin, J.-M. LeGoff, J. Rich, M. Viel, and D. Weinberg. Neutrino masses and cosmology with Lyman-alpha forest power spectrum. *J. Cosmology Astropart. Phys.*, 11:011, November 2015.

[173] N. Palanque-Delabrouille, C. Yèche, A. Borde, J.-M. Le Goff, G. Rossi, M. Viel, É. Aubourg, S. Bailey, J. Bautista, M. Blomqvist, A. Bolton, J. S. Bolton, N. G. Busca, B. Carithers, R. A. C. Croft, K. S. Dawson, T. Delubac, A. Font-Ribera, S. Ho, D. Kirkby, K.-G. Lee, D. Margala, J. Miralda-Escudé, D. Muna, A. D. Myers, P. Noterdaeme, I. Pâris, P. Petitjean, M. M. Pieri, J. Rich, E. Rollinde, N. P. Ross, D. J. Schlegel, D. P. Schneider, A. Slosar, and D. H. Weinberg. The one-dimensional Lyα forest power spectrum from BOSS. *A&A*, 559:A85, November 2013.

[174] R. D. Peccei and Helen R. Quinn. CP conservation in the presence of pseudoparticles. *Phys. Rev. Lett.*, 38:1440–1443, Jun 1977.

[175] A. A. Penzias and R. W. Wilson. A Measurement of Excess Antenna Temperature at 4080 Mc/s. *ApJ*, 142:419–421, July 1965.

[176] S. Perlmutter, G. Aldering, G. Goldhaber, R. A. Knop, P. Nugent, P. G. Castro, S. Deustua, S. Fabbro, A. Goobar, D. E. Groom, I. M. Hook, A. G. Kim, M. Y. Kim, J. C. Lee, N. J. Nunes, R. Pain, C. R. Pennypacker, R. Quimby, C. Lidman, R. S. Ellis, M. Irwin, R. G. McMahon, P. Ruiz-Lapuente, N. Walton, B. Schaefer, B. J. Boyle, A. V. Filippenko, T. Matheson, A. S. Fruchter, N. Panagia, H. J. M. Newberg, W. J. Couch, and T. S. C. Project. Measurements of Ω and Λ from 42 High-Redshift Supernovae. *ApJ*, 517:565–586, June 1999.

[177] J. Petrović, P. Dario Serpico, and G. Zaharijaš. Galactic Center gamma-ray "excess" from an active past of the Galactic Centre? *J. Cosmology Astropart. Phys.*, 10:052, October 2014.

[178] J. Petrović, P. D. Serpico, and G. Zaharijas. Millisecond pulsars and the Galactic Center gamma-ray excess: the importance of luminosity function and secondary emission. *J. Cosmology Astropart. Phys.*, 2:023, February 2015.

[179] Planck Collaboration, R. Adam, P. A. R. Ade, N. Aghanim, Y. Akrami, M. I. R. Alves, F. Argüeso, M. Arnaud, F. Arroja, M. Ashdown, and et al. Planck 2015 results. I. Overview of products and scientific results. *A&A*, 594:A1, September 2016.

[180] Stephen K. N. Portillo, Benjamin C. G. Lee, Tansu Daylan, and Douglas P. Finkbeiner. Improved Point Source Detection in Crowded Fields using Probabilistic Cataloging. mar 2017.

[181] J. I. Read, G. Iorio, O. Agertz, and F. Fraternali. The stellar mass-halo mass relation of isolated field dwarfs: a critical test of ΛCDM at the edge of galaxy formation. *MNRAS*, 467:2019–2038, May 2017.

[182] A. G. Riess, A. V. Filippenko, P. Challis, A. Clocchiatti, A. Diercks, P. M. Garnavich, R. L. Gilliland, C. J. Hogan, S. Jha, R. P. Kirshner, B. Leibundgut, M. M. Phillips, D. Reiss, B. P. Schmidt, R. A. Schommer, R. C. Smith, J. Spyromilio, C. Stubbs, N. B. Suntzeff, and J. Tonry. Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant. *AJ*, 116:1009–1038, September 1998.

[183] G. O. Roberts, A. Gelman, and W. R. Gilks. Weak convergence and optimal scaling of random walk metropolis algorithms. *Ann. Appl. Probab.*, 7(1):110–120, 02 1997.

[184] M. S. Roberts. M 31 and a Brief History of Dark Matter. In A. H. Bridle, J. J. Condon, and G. C. Hunt, editors, *Frontiers of Astrophysics: A Celebration of NRAO's 50th Anniversary*, volume 395 of *Astronomical Society of the Pacific Conference Series*, page 283, August 2008.

[185] M. S. Roberts and R. N. Whitehurst. The rotation curve and geometry of M31 at large galactocentric distances. *ApJ*, 201:327–346, October 1975.

[186] Gerald Rosen. Particlelike solutions to nonlinear complex scalar field theories with positive-definite energy densities. *Journal of Mathematical Physics*, 9(7):996–998, 1968.

[187] V. C. Rubin and W. K. Ford, Jr. Rotation of the Andromeda Nebula from a Spectroscopic Survey of Emission Regions. *ApJ*, 159:379, February 1970.

[188] R. E. Ryan, Jr., S. Deustua, M. Sosey, J. Anderson, S. M. Baggett, V. Bajaj, M. Bourque, A. Bowers, T. Dahlen, M. Durbin, C. Gosmeyer, H. Gunning, H. Khandrika, J. Mack, J. MacKenty, C. Martlin, V. Kozhurina-Platais, and E. Sabbi. The Updated Calibration Pipeline for WFC3/UVIS: a Reference Guide to calwf3 (version 3.3). Technical report, February 2016.

[189] R. K. Sachs and A. M. Wolfe. Perturbations of a Cosmological Model and Angular Variations of the Microwave Background. *ApJ*, 147:73, January 1967.

[190] A. D. Sakharov. Violation of CP Invariance, C asymmetry, and baryon asymmetry of the universe. *Pisma Zh. Eksp. Teor. Fiz.*, 5:32–35, 1967. [Usp. Fiz. Nauki161,no.5,61(1991)].

[191] M. Sambridge, K. Gallagher, A. Jackson, and P. Rickwood. Trans-dimensional inverse problems, model comparison and the evidence. *Geophysical Journal International*, 167:528–542, November 2006.

[192] T. Sawala, C. S. Frenk, A. Fattahi, J. F. Navarro, R. G. Bower, R. A. Crain, C. Dalla Vecchia, M. Furlong, J. C. Helly, A. Jenkins, K. A. Oman, M. Schaller, J. Schaye, T. Theuns, J. Trayford, and S. D. M. White. The APOSTLE simulations: solutions to the Local Group's cosmic puzzles. *MNRAS*, 457:1931–1943, April 2016.

[193] T. Sawala, P. Pihajoki, P. H. Johansson, C. S. Frenk, J. F. Navarro, K. A. Oman, and S. D. M. White. Shaken and stirred: the Milky Way's dark substructures. *MNRAS*, 467:4383–4400, June 2017.

[194] P. A. G. Scheuer and M. Ryle. A statistical method for analysing observations of faint radio stars. *Mathematical Proceedings of the Cambridge Philosophical Society*, 53(03):764, 1957.

[195] J. A. Schewtschenko, R. J. Wilkinson, C. M. Baugh, C. Boehm, and S. Pascoli. Dark matter-radiation interactions: the impact on dark matter haloes. *MNRAS*, 449(4):3587–3596, 2015.

[196] David J. Schlegel, Douglas P. Finkbeiner, and Marc Davis. Maps of Dust IR Emission for Use in Estimation of Reddening and CMBR Foregrounds. *The Astrophysical Journal*, 500(2):525, 1997.

[197] P. Schneider and D. Sluse. Mass-sheet degeneracy, power-law models and external convergence: Impact on the determination of the Hubble constant from gravitational lensing. *A&A*, 559:A37, November 2013.

[198] G Schwarz. JSTOR: The Annals of Statistics, Vol. 6, No. 2 (Mar., 1978), pp. 461-464. *The annals of statistics*, 1978.

[199] J. L. Sérsic. Influence of the atmospheric and instrumental dispersion on the brightness distribution in a galaxy. *Boletin de la Asociacion Argentina de Astronomia La Plata Argentina*, 6:41, 1963.

[200] G. Servant and T. M. P. Tait. Is the lightest Kaluza-Klein particle a viable dark matter candidate? *Nuclear Physics B*, 650:391–419, February 2003.

[201] B. D. Sherwin, J. Dunkley, S. Das, J. W. Appel, J. R. Bond, C. S. Carvalho, M. J. Devlin, R. Dünner, T. Essinger-Hileman, J. W. Fowler, A. Hajian, M. Halpern, M. Hasselfield, A. D. Hincks, R. Hlozek, J. P. Hughes, K. D. Irwin, J. Klein, A. Kosowsky, T. A. Marriage, D. Marsden, K. Moodley, F. Menanteau, M. D. Niemack, M. R. Nolta, L. A. Page, L. Parker, E. D. Reese, B. L. Schmitt, N. Sehgal, J. Sievers, D. N. Spergel, S. T. Staggs, D. S. Swetz, E. R. Switzer, R. Thornton, K. Visnjic, and E. Wollack. Evidence for Dark Energy from the Cosmic Microwave Background Alone Using the Atacama Cosmology Telescope Lensing Measurements. *Physical Review Letters*, 107(2):021302, July 2011.

[202] X. Shi and G. M. Fuller. New Dark Matter Candidate: Nonthermal Sterile Neutrinos. *Physical Review Letters*, 82:2832–2835, April 1999.

[203] Yiping Shu, Adam S. Bolton, Leonidas A. Moustakas, Daniel Stern, Arjun Dey, Joel R. Brownstein, Scott Burles, and Hyron Spinrad. Kiloparsec Mass/Light Offsets in the Galaxy Pair-Lyα Emitter Lens System SDSSJ1011+0143. *The Astrophysical Journal*, 820(1):43, mar 2016.

[204] John Skilling. Nested sampling for general bayesian computation. *Bayesian Anal.*, 1(4):833–859, 12 2006.

[205] G. F. Smoot, C. L. Bennett, A. Kogut, E. L. Wright, J. Aymon, N. W. Boggess, E. S. Cheng, G. de Amici, S. Gulkis, M. G. Hauser, G. Hinshaw, P. D. Jackson, M. Janssen, E. Kaita, T. Kelsall, P. Keegstra, C. Lineweaver, K. Loewenstein, P. Lubin, J. Mather, S. S. Meyer, S. H. Moseley, T. Murdock, L. Rokke, R. F. Silverberg, L. Tenorio, R. Weiss, and D. T. Wilkinson. Structure in the COBE differential microwave radiometer first-year maps. *ApJL*, 396:L1–L5, September 1992.

[206] D. N. Spergel, L. Verde, H. V. Peiris, E. Komatsu, M. R. Nolta, C. L. Bennett, M. Halpern, G. Hinshaw, N. Jarosik, A. Kogut, M. Limon, S. S. Meyer, L. Page, G. S. Tucker, J. L. Weiland, E. Wollack, and E. L. Wright. First-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Determination of Cosmological Parameters. *ApJS*, 148:175–194, September 2003.

[207] David N. Spergel and Paul J. Steinhardt. Observational Evidence for Self-Interacting Cold Dark Matter. *Physical Review Letters*, 84(17):3760–3763, sep 2000.

[208] V. Springel, C. S. Frenk, and S. D. M. White. The large-scale structure of the Universe. *Nature*, 440:1137–1144, April 2006.

[209] V. Springel, J. Wang, M. Vogelsberger, A. Ludlow, A. Jenkins, A. Helmi, J. F. Navarro, C. S. Frenk, and S. D M White. The Aquarius Project: The subhaloes of galactic haloes. *Monthly Notices of the Royal Astronomical Society*, 391(4):1685–1711, dec 2008.

[210] V. Springel, S. D. M. White, A. Jenkins, C. S. Frenk, N. Yoshida, L. Gao, J. Navarro, R. Thacker, D. Croton, J. Helly, J. A. Peacock, S. Cole, P. Thomas, H. Couchman, A. Evrard, J. Colberg, and F. Pearce. Simulations of the formation, evolution and clustering of galaxies and quasars. *Nature*, 435:629–636, June 2005.

[211] M Stone. Cross-Validatory Choice and Assessment of Statistical Predictions. *Journal of the Royal Statistical Society*, 36(2):111–147, 1974.

[212] L. E. Strigari, J. S. Bullock, M. Kaplinghat, J. Diemand, M. Kuhlen, and P. Madau. Redefining the Missing Satellites Problem. *ApJ*, 669:676–683, November 2007.

[213] M. Su, T. R. Slatyer, and D. P. Finkbeiner. Giant Gamma-ray Bubbles from Fermi-LAT: Active Galactic Nucleus Activity or Bipolar Galactic Wind? *ApJ*, 724:1044–1082, December 2010.

[214] S. H. Suyu, V. Bonvin, F. Courbin, C. D. Fassnacht, C. E. Rusu, D. Sluse, T. Treu, K. C. Wong, M. W. Auger, X. Ding, S. Hilbert, P. J. Marshall, N. Rumbaugh, A. Sonnenfeld, M. Tewes, O. Tihhonova, A. Agnello, R. D. Blandford, G. C.-F. Chen, T. Collett, L. V. E. Koopmans, K. Liao, G. Meylan, and C. Spiniello. HoLiCOW - I. H_0 Lenses in COSMOGRAIL's Wellspring: program overview. *MNRAS*, 468:2590–2604, July 2017.

[215] N. Suzuki et al. *ApJ*, 746:85, February 2012.

[216] Amitpal Tagore and Charles Keeton. Statistical and systematic uncertainties in pixel-based source reconstruction algorithms for gravitational lensing. *Mon. Not. Roy. Astron. Soc.*, 445(1):694–710, 2014.

[217] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1):267–288, 1996.

[218] S. S. Tie and C. S. Kochanek. Microlensing Makes Lensed Quasar Time Delays Significantly Time Variable. jul 2017.

[219] P. Tisserand, L. Le Guillou, C. Afonso, J. N. Albert, J. Andersen, R. Ansari, É. Aubourg, P. Bareyre, J. P. Beaulieu, X. Charlot, C. Coutures, R. Ferlet, P. Fouqué, J. F. Glicenstein, B. Goldman, A. Gould, D. Graff, M. Gros, J. Haissinski, C. Hamadache, J. de Kat, T. Lasserre, É. Lesquoy, C. Loup, C. Magneville, J. B. Marquette, É. Maurice, A. Maury, A. Milsztajn, M. Moniez, N. Palanque-Delabrouille, O. Perdereau, Y. R. Rahal, J. Rich, M. Spiro, A. Vidal-Madjar, L. Vigroux, S. Zylberajch, and EROS-2 Collaboration. Limits on the Macho content of the Galactic Halo from the EROS-2 Survey of the Magellanic Clouds. *A&A*, 469:387–404, July 2007.

[220] Tommaso Treu. Strong Lensing by Galaxies. *Annual Review of Astronomy and Astrophysics*, 48(1):87–125, mar 2010.

[221] Laura G. van den Aarssen, Torsten Bringmann, and Christoph Pfrommer. Is dark matter with long-range interactions a solution to all small-scale problems of Λ CDM cosmology? *Phys. Rev. Lett.*, 109:231301, 2012.

[222] Frank C. van den Bosch. Dissecting the evolution of dark matter subhaloes in the bolshoi simulation. *Monthly Notices of the Royal Astronomical Society*, 468(1):885–909, nov 2017.

[223] S. Vegetti and L. V. E. Koopmans. Bayesian strong gravitational-lens modelling on adaptive grids: objective detection of mass substructure in Galaxies. *MNRAS*, 392:945–963, January 2009.

[224] S. Vegetti and L. V. E. Koopmans. Statistics of mass substructure from strong gravitational lensing: quantifying the mass fraction and mass function. *MNRAS*, 400:1583–1592, December 2009.

[225] S. Vegetti, L. V E Koopmans, M. W. Auger, T. Treu, and A. S. Bolton. Inference of the cold dark matter substructure mass function at z = 0.2 using strong gravitational lenses. *Monthly Notices of the Royal Astronomical Society*, 442(3):2017–2035, may 2014.

[226] S. Vegetti, L. V E Koopmans, A. Bolton, T. Treu, and R. Gavazzi. Detection of a dark substructure through gravitational imaging. *Monthly Notices of the Royal Astronomical Society*, 408(4):1969–1981, oct 2010.

[227] S. Vegetti, D. J. Lagattuta, J. P. McKean, M. W. Auger, C. D. Fassnacht, and L. V. E. Koopmans. Gravitational detection of a low-mass dark satellite at cosmological distance. *Nature*, 481:341, jan 2012.

[228] T. Vernstrom, Douglas Scott, J. V. Wall, J. J. Condon, W. D. Cotton, E. B. Fomalont, K. I. Kellermann, N. Miller, and R. A. Perley. Deep 3 GHz number counts from a P(D) fluctuation analysis. *Monthly Notices of the Royal Astronomical Society*, 440(3):2791–2809, 2014.

[229] C. Vlahakis et al. The 2014 ALMA Long Baseline Campaign: Observations of the Strongly Lensed Submillimeter Galaxy HATLAS J090311.6+003906 at z = 3.042. *ApJL*, 808:L4, July 2015.

[230] M. Vogelsberger, J. Zavala, F.-Y. Cyr-Racine, C. Pfrommer, T. Bringmann, and K. Sigurdson. ETHOS - an effective theory of structure formation: dark matter physics as a possible explanation of the small-scale CDM problems. *MNRAS*, 460:1399–1416, August 2016.

[231] Matthew G. Walker and Jorge Penarrubia. A Method for Measuring (Slopes of) the Mass Profiles of Dwarf Spheroidal Galaxies. *Astrophys. J.*, 742:20, 2011.

[232] Yuan Wang, Xiaobo Zhou, Honghui Wang, King Li, Lixiu Yao, and Stephen T C Wong. Reversible jump MCMC approach for peak identification for stroke SELDI mass spectrometry using mixture model. *Bioinformatics*, 24(13):i407–13, 2008.

[233] Steven Weinberg. A model of leptons. *Phys. Rev. Lett.*, 19:1264–1266, Nov 1967.

[234] Steven Weinberg. Approximate symmetries and pseudo-goldstone bosons. *Phys. Rev. Lett.*, 29:1698–1701, Dec 1972.

[235] A. R. Wetzel, P. F. Hopkins, J.-h. Kim, C.-A. Faucher-Giguère, D. Kereš, and E. Quataert. Reconciling Dwarf Galaxies with ΛCDM Cosmology: Simulating a Realistic Population of Satellites around a Milky Way-mass Galaxy. *ApJL*, 827:L23, August 2016.

[236] M. A. Worsley, A. C. Fabian, X. Barcons, S. Mateos, G. Hasinger, and H. Brunner. The (un)resolved X-ray background in the Lockman Hole. *Monthly Notices of the Royal Astronomical Society*, 352(3), 2004.

[237] C. Yèche, N. Palanque-Delabrouille, J. Baur, and H. du Mas des Bourboux. Constraints on neutrino masses from Lyman-alpha forest power spectrum with BOSS and XQ-100. *J. Cosmology Astropart. Phys.*, 6:047, June 2017.

[238] D. G. York et al. The Sloan Digital Sky Survey: Technical Summary. *AJ*, 120:1579–1587, September 2000.

[239] N. Yoshida, V. Springel, S. D. M. White, and G. Tormen. Weakly Self-interacting Dark Matter and the Structure of Dark Halos. *ApJL*, 544:L87–L90, December 2000.

[240] Q. Yuan and B. Zhang. Millisecond pulsar interpretation of the Galactic center gamma-ray excess. *Journal of High Energy Astrophysics*, 3:1–8, September 2014.

[241] Hannes-S. Zechlin, Alessandro Cuoco, Fiorenza Donato, Nicolao Fornengo, and Andrea Vittino. Unveiling the Gamma-ray Source Count Distribution Below the Fermi Detection Limit with Photon Statistics. *The Astrophysical Journal Supplement Series*, 225(2):18, dec 2015.

[242] Y. B. Zel'dovich. Gravitational instability: An approximate theory for large density perturbations. *A&A*, 5:84–89, March 1970.

[243] Wanchuang Zhu and Yanan Fan. Relabelling Algorithms for Large Dataset Mixture Models. *arXiv preprint*, mar 2014.

[244] F. Zwicky. Die Rotverschiebung von extragalaktischen Nebeln. *Helvetica Physica Acta*, 6:110–127, 1933.