

How To Win With Non-Gaussian Data: Poisson Goodness-of-Fit

A. Connors

CHASC and Eureka Scientific; (aconnors@eurekabayes.com)

D. A. van Dyk

CHASC and University of California, Irvine; (dvd@uci.edu)

Abstract.

We propose a new method to test for goodness-of-fit of a model for low-count Poisson data. Our approach does not resemble the usual methods of approximations to χ^2 , but instead explicitly uses the full Poisson distribution. First, we propose to use a simple (Poisson-specific) multiscale model to characterize the “mismatch” between a best-fit physical model and the data. Next, we embed this multiscale model into a probabilistic/likelihood framework (via hierarchical Bayes), allowing us to handle statistical uncertainties. We then use MCMC to map out the shape of the joint posterior probability of all of the unknown parameters. Finally, we note that this is a generalization of a problem with a known solution: whether an additional component of known shape is justified by the data. Hence, when the multiscale structure that we use to model the “mismatch” between the data and the physical model accounts for a significant number of counts, and/or when the scale-factor for the best-fit physics model is significantly different than one, then the original fit was not “good-enough”. That is, the fitted multiscale “mismatch” isolates the discrepancy between model and data. We use models of the Gamma-Ray sky as viewed by CGRO/EGRET as our example. We demonstrate that our method works nicely on several examples, but further work is needed to investigate the method’s power.

The authors and CHASC acknowledge support from the Chandra X-Ray Center, NSF(DMS 04-06085), and AISRP NAG5-12082. It is a particular pleasure to acknowledge the SAMSIO6 Special Program in Astrostatistics.

1. Introduction: Two Tough Problems

We begin by discussing two tough related questions that often arise in astronomical image analysis: “Is a fitted model *Good Enough*?” and “Is there additional unknown structure in the image (or spectrum)?” That is, faced with a dataset, and a physical model¹ that is to be fit to the data an astrophysicist generally confronts four main questions:

QUESTION 1: What physical model parameters best fit the data, and what are their uncertainties;

¹Throughout this paper we use “physical model” to mean a model of the photon source. It is assumed to be a parametric model based on prior physical information, and of (relatively) low dimension. The “statistical model”, on the other hand, is a model of the complete data generating mechanism. It incorporates both the physical model and the statistical distribution of the data including any calibration and measurement uncertainties.

QUESTION 2: What is the evidence for an additional component of known shape (such as a line in an energy spectrum or a point-source in an image);

QUESTION 3: Is this best-fit good enough; and

QUESTION 4: If the fit is not good enough how does one quantify the unknown additional structure in the image and the uncertainties in this structure?

Most astrophysicists know reasonably principled procedures for QUESTION 1 (exploring the probability space around the maximum likelihood estimate) and QUESTION 2 (such as computing limits on the size of the possible additional component, or calibrating the rise in log-likelihood as in Protassov et al. (2001) and the references therein). However outside the Gaussian realm, QUESTIONS 3 and 4—is the fitted physical model good enough and how does one quantify additional structure in the image—are tough. Since we will use the structure later, we next sketch four example problems, rather pedantically, from simplest to more troublesome.

1.1. One Bin Problems

PROBLEM 1: ONE BIN, GAUSSIAN STATISTICS. In the first problem, data are a rate in a single bin, with a known physical model giving the mean (and previously-measured instrumental characteristics giving the variance). It is straightforward to obtain the tail probability of the data under the Gaussian model. There is a well-known reference distribution (i.e., χ_1^2) that does not depend on details of the parameters or the model and that is easy to calculate once ahead of time using a standard table. Further, any extra or missing component can be estimated simply as the data minus the model (i.e., the residual), whether one prefers a mean, a mode, or a median, since they are the same. Unfortunately, the simplicity of this analysis is unique to this simple case.

PROBLEM 2: ONE BIN, POISSON COUNTS. In this case, the data are again in a single bin, with the known physical model giving the mean. But now the “rate” is a discrete variable, i.e., the Poisson-distributed counts in the bin (or detector) of known size (or known “effective area”). Once again, the computation of a tail probability is straightforward (if more tedious—it must be calculated rather than looked up; see Bevington 1969, Loredò 1992). One has a choice of estimator for both the physical model and the excess structure. The Bayesian posterior mean, mode, and median, for example, are not the same. Nonetheless the procedures are conceptually simple (if numerically more tedious).

1.2. Multi-Bin Problems

With more bins, we can model shape or structure in how the flux varies among the bins as well as the total flux. At the same time, the two tough questions, QUESTIONS 3 and 4, require more complex treatment.

PROBLEM 3: MULTIPLE BINS, GAUSSIAN STATISTICS. Presume one has a spectral energy distribution or image, with independent Gaussian measurements and heteroscedastic errors (i.e., the variances are known ahead of time, but differ among the energy channels or image pixels).

QUESTION 1: *Parametrized physical model—best fit and uncertainties.* As Lampton, Margon and Bowyer (1976) and Cash (1976) suggest, one finds the mode (or modes) by maximizing the log-likelihood function via one of a number of standard numerical optimization methods (as in Press et al 1992). The region around the mode can then be mapped. The correspondence between drop in log-likelihood and interval coverage is assumed (via the Central Limit Theorem) to be given by χ^2_ν regardless of the details of the model. Alternatively, a Bayesian procedure maximizes and maps out the posterior probability of the parameters given the model and data. The probability that a parameter is within a certain range (“credible region”) is then calculated directly (as in Loredo 1992 and references therein).

QUESTION 2: *Evidence for an additional component of known shape.* Since a parametric function for the possible additional component is known, one builds on the procedure described above for one bin in two ways. First, one investigates the limits on the total flux (or equivalent) of the unknown component by computing confidence or credible intervals. Second, one investigates whether the rise in log-likelihood upon the addition of this component signal it is statistically significant. (We remind astrophysicists that for many applications, often including point sources and spectral lines, the mapping between the rise in log-likelihood and the chance significance cannot be compared to a pre-computed reference distribution such as χ^2 , but must be calibrated via appropriate Monte Carlo; Protassov et al. 2002)

QUESTION 3: *Is the best fit good enough?* One can compute a χ^2 on ν degrees of freedom, and compare it to a previously tabulated “reference distribution” (i.e., χ^2_ν) by invoking the Central Limit Theorem (e.g., Eadie et al. 1971 and Bevington 1969). There is nothing technically incorrect about this simple, popular procedure. However χ^2 is almost always an approximation to the true distribution. Hence any statistic that relies on its tail (where the differences would be the most pronounced) should be viewed with skepticism. Also, Loredo (1992) points out that occasionally one should *expect* data that are not a “good fit” to one’s model, by chance.

QUESTION 4: *Quantify unknown additional structure and its uncertainties.* If the physical model is not sufficient to describe the data, the difference between the data and the physical model, i.e., the residual, is a reasonable estimate of additional unknown component. Uncertainties can be computed by appropriately scaling the pixel-wise observational variances. If something is known about the structure of the additional component (such as local smoothness) one could gain additional power by using a model that formalizes this knowledge. However there is nothing actually wrong with simply subtracting the physical model from the data, as is usually done.

PROBLEM 4: MULTIPLE BINS, POISSON COUNTS. This is the same as PROBLEM 3, but with Poisson data, i.e., discrete counts.

QUESTION 1: *Parametrized physical model—best fit and uncertainties* and QUESTION 2: *Evidence for an additional component of known shape.* The procedures for these questions are essentially the same as for the Gaussian case (Cash 1979; Protassov et al. 2002).

QUESTION 3: *Is the best fit good enough?* Here is our challenge. In the Poisson case, there is no pre-computed, single, reference distribution that does not depend on the details of the model and data.

QUESTION 4: *Quantify unknown additional structure and its uncertainties.* If one has no good prior knowledge of a simple parametric model for the “mismatch” between the data and physical model, how does one estimate shape and significance of the mismatch? Subtracting the model from the data is no longer a handy estimator even of the mean. Even if the counts are relatively high, when one is interested in low probability events—such as the counts from the unknown additional structure—it is important to take full account of the skew and long tail of the underlying Poisson distribution.

We briefly note that although we have not discussed instrumental “blurring” or “smearing”, if the blurring matrix or point spread function is assumed to be known completely, it is straightforward to convolve it with the sky and physical models before comparison with the data. Thus, we mainly ignore these issues throughout this article. (The case of uncertainty in instrumental response is being addressed in a hierarchical way in a separate effort; see Drake et al. 2006.)

2. High Energy Gamma-Ray Bridge Over the Milky Way, circa 1999

2.1. Brief Gamma-Ray Astronomy Overview: Matter & Energy in our Galaxy

Gamma-ray emission (and hence detection) is an intrinsically Poisson process (Longair 1992). At very high energies, even after almost a decade of high-resolution satellite observations are summed, there are many regions in the sky with very low counts per bin (see Figure 1a, Left). Gaussian approximations are not suitable.

Detailed physics theory — such as quantum mechanics — allows researchers to ‘decode’ the basic physics of distant objects via their remotely-detected photons. Across the entire electro-magnetic spectrum, from radio (lowest energies) through infrared, optical, UV, X-ray, and gamma-rays (highest energies), photons tell us about specific processes in the sky. Gamma-ray emission originates, for example, from high-energy cosmic processes including nuclear decay and high energy particles impinging on the magnetic fields, dense material, or lower energy photons. In particular, gamma-rays reveal the whereabouts of extremely energetic particles, in our Milky Way Galaxy and beyond.

Given this, do we understand the production cycle of such highly energetic matter and energy in our Galaxy? Or is there more “stuff” we are missing? From our vantage point two thirds of the way out from the center of the Milky Way Galaxy, how well can we use the information packed in gamma-ray photons to learn about these processes?

2.2. The Gamma-Ray Sky

Instruments. These were the questions and techniques we faced in the 1990’s, when the *Compton Gamma-Ray Observatory* (CGRO), the first of NASA’s “Great Observatories”, surveyed the entire sky for gamma-ray emission. CGRO carried two survey telescopes: COMPTEL (0.5-30 MeV, the nuclear regime); and EGRET (30 MeV-100 GeV). It also carried two non-imaging instruments: BATSE, which focused on gathering signals from the whole sky as a function of time (in order to observe gamma-ray bursts); and OSSE, which concentrated on spectra.

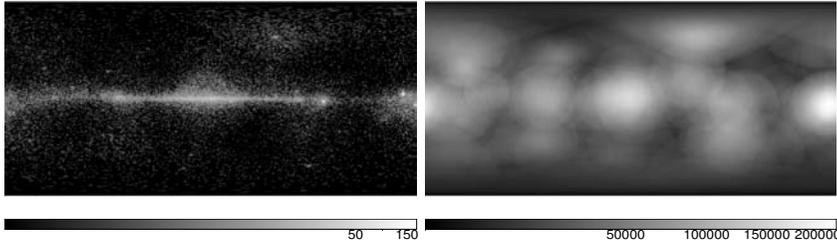


Figure 1. CGRO/EGRET data: a) Left: >1 GeV counts measured over the whole sky over the lifetime of the mission (from HEASARC), plotted in Galactic coordinates in $0.5^\circ \times 0.5^\circ$ bins. b) Right: corresponding exposure (effective area times time).

Of these, the instrumental response of CGRO-EGRET’s Spark Chamber is the simplest, and its particle induced background is the lowest. Hence for this paper, where we ignore “deconvolution”, we used > 1 GeV EGRET Spark Chamber data (real and simulated), binned into $1.5^\circ \times 2^\circ$ bins, a level of resolution where instrumental blurring is minimal.

CGRO/EGRET data. In Figure 1a (on the left), we show the >1 GeV counts (per $0.5^\circ \times 0.5^\circ$ bin) collected by the CGRO/EGRET telescope during its $\sim 9+$ years viewing the sky (Nolan et al. 1992, Thompson et al. 1993, Esposito et al. 1999, Mattox et al. 1992). On the right is the corresponding EGRET telescope effective area, in units of cm^2 . Visible in the plot of gamma-ray counts are a couple hundred “point sources” (Hartman et al. 1999). A number are thought to be spatially unresolved glowing gas from supernova remnants (SNRs) (Esposito et al. 1996). Virtually all of the others are understood to be compact objects such as neutron stars and accreting black holes. These include pulsars (neutron stars with magnetic fields roughly a dozen orders of magnitude greater than that of the Earth) and various kinds of active galaxies.

The bulk of the signal in this low-background, low count regime, however, comes from diffuse gas and nearby background photons along the plane of our galaxy (e.g., Cillis and Hartman 2005, Elwe 2003, and references therein), glowing as they are impacted by energetic particles called cosmic rays. There were physics-based models available at the time that predicted what the diffuse glow should look like (Bertsch et al. 1993; Hunter et al. 1997, Strong and Moskalenko 1998, Moskalenko, Strong, and Reimer 1998, Moskalenko and Strong 1998). These models of the Galaxy included supernovae as sources of energetic cosmic ray particles; the spatially complex gamma-ray glow from energetic particles impinging on gas clouds (Bremmstrahlung and Pion decay); and a broader, smoother component due to these energetic particles “boosting” the energy of photons from the ubiquitous low energy (microwave and infrared) background (Inverse Compton). The components of these physical models were measured separately. High resolution radio observations were used to map out the gas distribution, and local measurements of cosmic ray particles were used to constrain their energy spectra and abundances.

With the physical model and CGRO observations in place, a fundamental question was whether the models matched the observed data.

2.3. Imaging Methods for the Whole Sky in Gamma-Rays

CGRO era. In order to answer this question, Dixon et al. (1999) applied the TIPSH (translation invariant Poisson smoothing using Haar wavelets) method of Kolaczyk (1997) and Dixon and Kolaczyk (2000). That is, in QUESTION 4 under sc Problem 4 above, they modeled the “mismatch” between the physical model and CGRO/EGRET all-sky data with a multiscale (boxy-shaped Haar wavelet, with cycle-spinning) model. The TIPSH “keep or kill” thresholds were carefully tailored to match the Poisson statistics. At the time, they did find a mismatch that looked rather like a bridge above the plane of the Galaxy. They famously emphasized, however:

“The immediate question arises as to the statistical significance of this feature. Though we are able to make rigorous statements about the coefficient-wise and level-wise FDR [False Discovery Rate], similar quantification of object-wise significance (e.g., ‘this blob is significant at the n sigma level’) are difficult.”

The use of multi-scale methods was a notable improvement in modeling unknown diffuse emission, which is expected to show both the sharp and broad features of the complicated diffuse gas, the photon field, and sea of energetic particles (cosmic rays) from which it arises. Previously, among astrophysicists, other prominent “model-free” or non-parametric (i.e., a model with about as many parameters as data bins) methods include Richardson-Lucy (the EM algorithm applied to an unconstrained grid of poisson parameters; Richardson 1972, Lucy 1974, Shepp and Vardi 1982) and Maximum-Entropy-related methods (such as MaxEnt and successors; Skilling 1989; also Strong 2003). These methods lacked the multiscale representation of both diffuse and sharp features.

Progress 1: NK. In the ensuing years, several potential improvements were tried. (See also Willett 2006 and Young 2006, this volume; and the different approaches of, say, Bourdin et al. 2001; Starck, Pantin and Murtagh 2002, Starck and Murtagh 2006 and references therein.) The TIPSH wavelet structure was revamped into a more general multiscale structure on certain ratios of expected counts to exploit the statistical properties of the Poisson likelihood and to allow the inclusion of instrument response (“smearing” or blurring) in an elegant (and speedy) way (Nowak and Kolaczyk 2000; Kolaczyk and Nowak 2004, 2005; henceforth this method is referred to as NK). Like wavelets (which often use recursive dyadic partitioning), the NK structure breaks an image into four quadrants (or two segments, in 1D) and then successively breaks each of these quadrants into its four quadrants until the inherent pixelization is reached. Unlike wavelets, for which the basis function is typically some sort of differencing function, for NK the basic shape is a constant intensity block. Further, rather than working with the intensity itself, we consider the ratios of the intensities of each quadrant to the whole (see Willett 2006, this volume). These “split probabilities” are described by prior distributions tuned by smoothing parameters, in analogy to the way that shrinkage (or complexity penalties) are used to estimate the best values of coefficients in a wavelet decomposition. Further, NK is formulated so that the problem is convex and there is

only one mode, so good convergence is assured. Finally, as with most wavelet analyses, cycle-spinning is recommended to remove artifacts in the fitted image (Coifman and Donoho, 1995). Still, this used the EM algorithm to get a “best fit” image, and uncertainty estimates were not available.

Progress 2: EMC2. Later, Esch et al. (2004) embedded the NK multiscale model into a hierarchical Bayesian framework. (See Karovska et al., 2005 for applications with beautiful graphics.) This also allows the smoothing parameters used in NK to be fit to the data. The range of variation of the tuning parameters is effectively governed by a hyper-parameter that is set ahead of time. In order to fully map out the posterior distribution of the various model and tuning parameters, Esch et al. (2004) uses a Markov Chain Monte Carlo sampler. This procedure produces not only a “best fit” image (the mean, rather than the mode), but also a full exploration of the probability space, via the MCMC samples. Hence in principle one has all the information necessary for calculating any summary of the posterior distribution and to quantify uncertainty in the fitted image.

The results of an EMC2 run on a Poisson image are often not as pleasing to the eye as, say, a Gaussian-kernel-smoothed picture, or the related popular (commercial) package *Pixons* (Pina and Puetter 1993). However, since the original NK algorithm elegantly factors the log-likelihood, it is both complete and speedy. The examples later in the paper each took about 40 minutes to run on a small iBook G4. Further, in theory, one has all the information one needs to understand the uncertainties in the fitted image. Unfortunately, it was not immediately obvious how this information could be synthesized to answer the questions of direct scientific interest.

We emphasize that although we are not discussing it here, NK and EMC2 do have the capacity to “deconvolve” the image, see Esch et al. (2004) and Karovska et al. (2005). (More precisely, EMC2 employs forward-folding of the multiscale model convolved with the instrument response.)

New features. During the SAMSI06 Special Session on Astrostatistics, the second author worked on modularizing, speeding up, and adding new features to the EMC2 code. In this paper, we focus particularly on the ability to add a shaped baseline component and to fit this component’s overall intensity at the same time as inferring the shape of an additional NK-type component.

The critical question was still: How can we use these methods, in practice, to answer the question Dixon et al. (1999) posed? How can we begin to quantify the statistical significance of features in the fitted image and how can we assess goodness of fit of the statistical model?

3. How to Define a ‘Feature’

Usual thinking. An astrophysicist investigating possible unknown diffuse emission (or an unknown spectral feature) usually tries to draw a boundary around it and then tries to estimate its significance. Whether doing simple box-car smoothing or more robust Gaussian kernel smoothing (Ebeling, White, and Rangarajan 2006), or even wavelet-based detection of potential sources at different scales (Freeman et al. 2002, 2003), the usual algorithms can be thought of as defining a region of ‘significant’ emission by drawing a boundary around it and calculating its significance. That is, once the

boundary is drawn, one can ask: are there a significant number of counts in this new feature, compared to statistical fluctuations expected from my “baseline”, or known physical model? (See also wavelets in an EM context in Knödelseder et al. 1999). Still, as in Dixon et al. (1999), the researcher is essentially looking at coefficient-wise or level-wise (rather than “object-wise”) significance. Even the first uses of EMC2 followed (roughly) this usual thinking (Esch et al. 2004; Karovska et al. 2005), defining the boundary at the pixel-wise 3σ limit.

However, the larger question of how to quantify the uncertainty in the boundary itself remained unanswered.

Key breakthrough—It’s in there. After much discussion, we realized that the multiscale model itself had already “chosen” the boundaries. In real astrophysical sources, boundaries of glowing gas clouds or photon fields are not sharply defined, nor are they smooth or even continuous. No additional step was needed: we could start with the best-fit physical model and add an additional component model via the NK/EMC2 flexible multiscale model. Since we know how to fit this model, we have, in essence, reduced our original two tough problems (QUESTIONS 3 and 4 under PROBLEM 4) to one that we already know how to handle (i.e., QUESTION 2, but with a flexible semi-parametric model as our proposed additional component, rather than a low-dimensional physics-based model).

We can quantify the evidence for including the added multiscale component in a simple manner by looking at the posterior distribution of the total flux from this component using the MCMC sampler, as we have done informally (visually) in Figures 6 and 9. A more computationally intensive but possible approach is to run a simulation to calibrate the distribution of the change in log-likelihood with the addition of this component. This strategy is similar to the calibration methods for the LRT promoted by Protossov et al. (2002).

In summary, our method starts with the physical model and adds a multiscale model. We aim to investigate whether any statistically significant structure above the supposed physical model is captured by the multiscale component (detection). This strategy simultaneously gives us a model for the unknown feature’s shape (characterization). Finally, we need a strategy for determining if the original physical model is “good enough”. That is, we need a procedure that gives a “goodness of fit” measure for our best-fit physical model and describes the shape and significance of the “residuals”.

For this final objective we suggest calibrating the null distribution (for what no added feature looks like) by simulating several datasets from the physical model (our “baseline”, or null hypothesis). On each of these simulated data sets, we run the EMC2+baseline fitting procedure. The baseline corresponds to the physical model and the multiscale model is used for the added model component or feature. We then compare the joint distribution of the physical model scale factor and the total expected counts from the multiscale component for each simulation with that of the real data. If the multiscale component is more important when fit to the real data than to the simulations, we conclude the new extra component may be real.

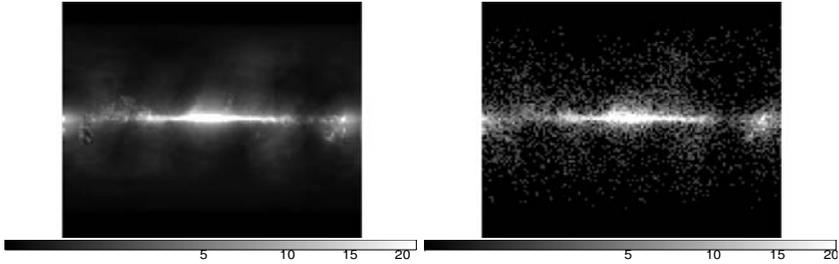


Figure 2. Quantifying Nothing—Model+Data: a) Left: Physical model of all-sky diffuse emission >1 GeV; b) Right: Poisson data simulated from the physical model with no extra component. Top and bottom are padded with zeros to make an even 128×128 bins.

4. Worked Example: Simulations of All-Sky CGRO/EGRET Data

Simulation summary. We began with simulated data based on the CGO EGRET summed phase 1-9 exposure and all-sky count-rates, courtesy of HEASARC.²

We use the output of the physics-based model GALPROP (version `fits_49_700405`) provided via A. Strong³ (Strong, Moskalenko, and Reimer 2004; Strong, et al. 2004; 2005); together with the CGRO/EGRET exposure map of Figure 1b, to simulate several physically realistic maps of the gamma-ray sky >1 GeV. We simulated Poisson datasets that matched the physical model assuming an integration time of about two weeks, as well as datasets with extra components and missing pieces. Our procedure is:

1. The physical model is used as a baseline in EMC2.
2. An additional component of unknown shape is modeled using the NK/EMC2 multiscale model.
3. The normalization of the baseline physical model is fit.
4. If the multiscale model component contains a significant number of counts, and/or if the normalization of the baseline is significantly different than one, then we conclude the physical model is not a good enough fit.
5. Significance is assessed by comparison to the posterior distributions produced by running EMC2 on simulations from the uncontaminated physical model (i.e, the baseline or null model).

Measuring nothing (no added feature). In the first simulation study, we demonstrate how this procedure works when our data has no extra component above the physical model (i.e., baseline or null model; Figure 2). Even when started far from the

²coss.gsfc.nasa.gov/docs/cgro/egret/egret.doc.html.

³www.gamma.mpe.mpg.de/~aws/aws.html

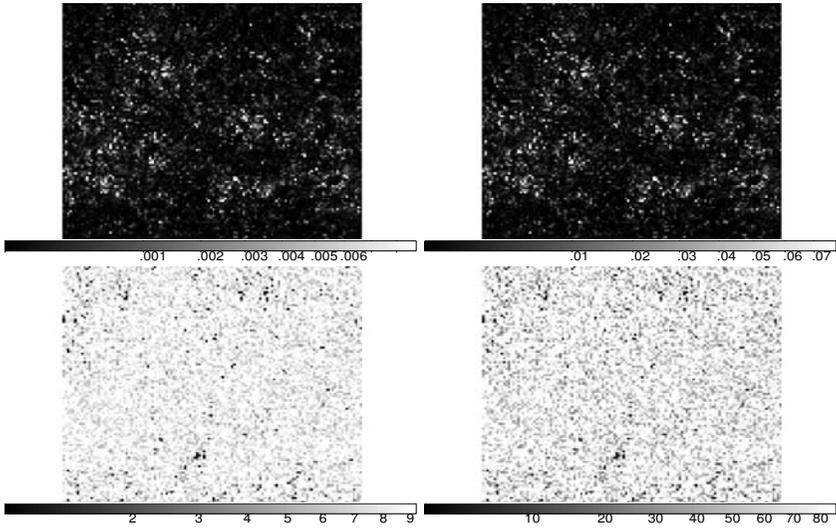


Figure 3. Quantifying Nothing—Results: a) Top Left: mean value in each pixel from 900 draws of multiscale model (i.e., EMC2 run); b) Top Right: sigma of same; c) Bottom Left: skew; d) Bottom Right: kurtosis.

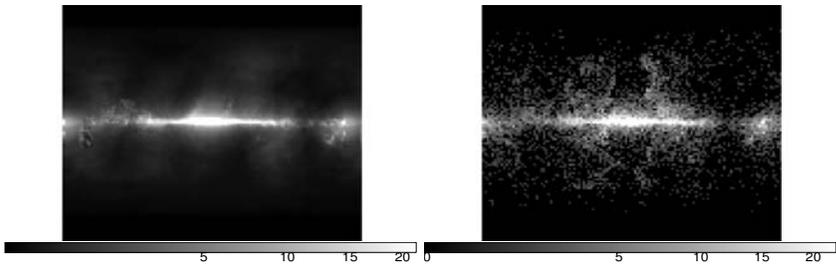


Figure 4. Bright Unknown (Discontinuous) Source—Model+Data: a) Left: Again, the baseline is the physical model of all-sky diffuse emission >1 GeV; b) Right: Poisson data simulated from the physical model with a bright added component. Top and bottom are padded with zeros to make an even 128×128 bins.

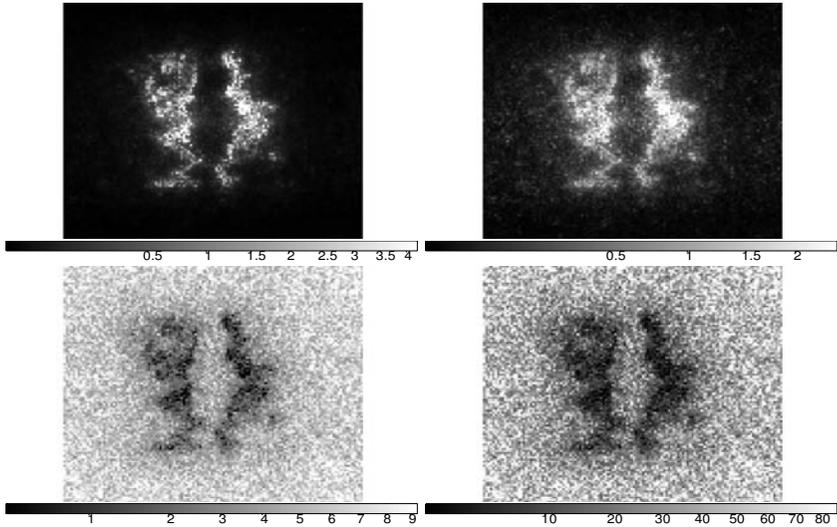


Figure 5. Bright Unknown (Discontinuous) Source—Results: a) Top Left: mean value in each pixel from 900 draws of multiscale model (i.e., EMC2 run); b) Top Right: sigma of same; c) Bottom Left: skew; d) Bottom Right: kurtosis. Recall these moments display only the shape of the unknown component. Significance is assessed as in Figure 6.

true value, the MCMC sampler converges correctly, showing no significant structure or counts in our multiscale component, see Figure 3. There is no evidence for multiple modes; the inferred mean (and mean/sigma) are very small compared to the counts (and baseline rate); and the skew (and kurtosis) are large, indicating the posterior distribution of the multiscale count is highly skewed towards zero, as it should be. In order to spread out the sharply peaked posterior probability of our two parameters of interest (the baseline scale factor and total expected count from the multiscale component) we have plotted them on a \log_{10} scale.

Bright discontinuous unknown. The results appear very different when we add a bright but discontinuous, irregular, and low surface brightness second component to the physical model, see Figure 4. The difference between the baseline physical model (Figure 4a) and simulated data (Figure 4b) is visible to the unaided eye. Yet from plots of the mean (and sigma) counts per pixel (Figure 5a and 5b), one sees the rate per pixel is not significant. But the overall multiscale component is hugely significant (Figure 6). Again, even when starting far from the true value, the MCMC sampler converges correctly and there is no evidence for multiple modes. The unknown component's shape—i.e., the difference between areas where the two irregular diffuse components are statistically significant and where they are not—is easily seen in the plots of the mean, sigma, skew, and kurtosis. Again, in order to spread out the sharply peaked posterior probability of our two parameters of interest, the baseline scale factor and the total expected counts due to the multiscale component, we use the \log_{10} scale.

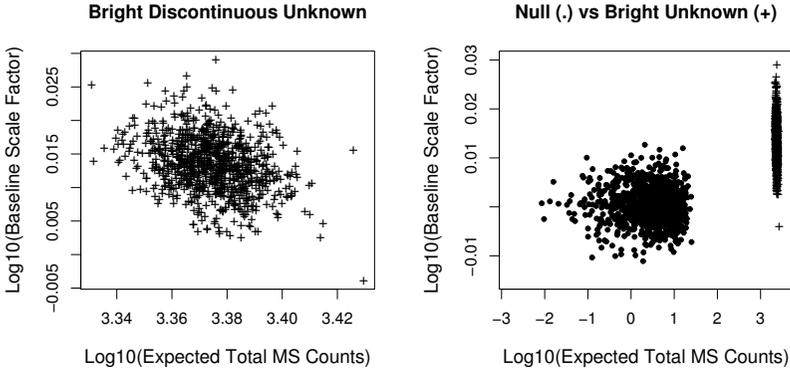


Figure 6. Calibrating Results—Null vs Bright Unknown Component: Scatter plot of \log_{10} of total number of counts inferred to be in the multiscale model for each of the draws, versus \log_{10} of the baseline scale factor. a) Left: Bright unknown component; b) Right: Same, compared to results on the baseline (null). Bullets indicate data with no signal (null hypothesis), while ‘+’ indicates results from data with a bright additional component.

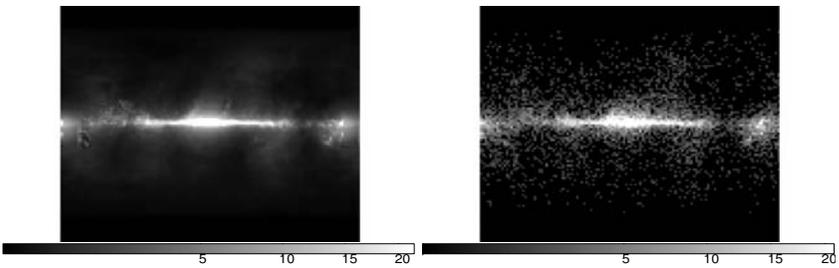


Figure 7. Faint Unknown Source, Baseline Shape Mis-Specified — Model+Data: a) Left: Baseline is physical model of all-sky diffuse emission >1 GeV, but with intensity distribution of IC component gradually suppressed above the Galactic plane; b) Right: Poisson data simulated from same spatial shape as the physical model in Figure 2.

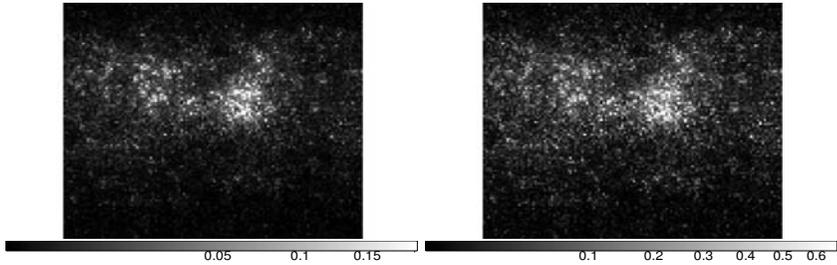


Figure 8. Faint Unknown Source, Baseline Shape Mis-Specified—Results, Scale Fixed: a) Left: mean value in each pixel from 900 draws of multiscale model (i.e., EMC2 run); b) Right: sigma of same. (The skew and kurtosis follow the same pattern as in the previous example.) Recall these moments display only the shape of the unknown component. Significance is assessed as in Figure 9.

Faint unknown source, baseline shape mis-specified. In our final simulation, the total difference between the model and simulated data is only about $\sim 10^2$ counts (rather than $\sim 10^3$ counts, as above). Further, as careful visual comparison of the baseline (Figure 7a) and the simulated data (Figure 7b) suggests, the Poisson counts were simulated from a distribution with the same spatial shape as the baseline physical model, but with overall intensity of the Inverse Compton (IC) component suppressed via a power-law ($\sim b^{-2}$) as one moves above the Galactic Center. This roughly represents a realistic mistake in the physics model.

We have processed these data two ways: with the baseline scale component fixed; and with it allowed to float. The multiscale component picks up the missing broad glow quite easily when the baseline scale factor is fixed; with the brightest parts being where the exposure is greatest (Figure 8). However, interestingly, when the scale factor is allowed to float, the procedure prefers to increase the scale factor rather than put counts into a new component (Figure 9).

5. Conclusions: General Principals, and Further Work

5.1. Brief Mention of Other Methods

There are many methods not mentioned here (e.g., Rice, 2006, these proceedings). In a general sense, they share with our proposed technique an overarching strategy of non- or semi- parametrically looking for regions of fluctuations in the data beyond those expected by chance from the underlying distribution of the data. Our method, however, is specifically tailored to Poisson images and energy spectra. We note that other Poisson multiscale (or very flexible) models could also be used in place of EMC2 (e.g., Bourdin et al. 2001, Starck and Murtagh 2006, and references therein).

5.2. Further Work

The simulations described here, though promising are just the beginning. We hope to develop methods to quickly calibrate and precisely quantify the improvement offered by the addition multiscale component. Rather than a simple visual inspection of the posterior distribution, such methods would offer a precise numerical summary. The

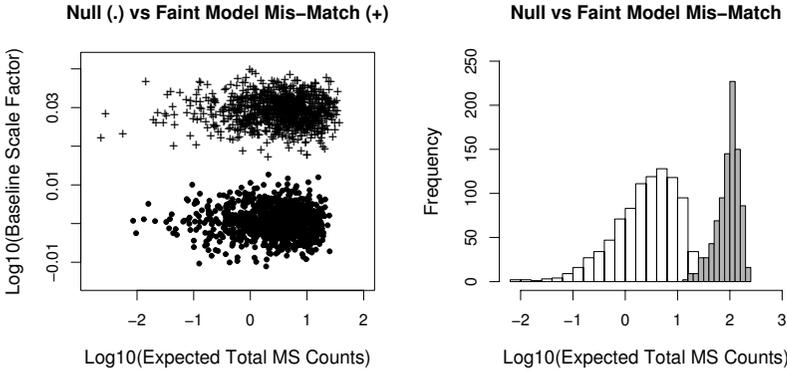


Figure 9. Calibrating results—Null vs Faint Unknown: a) Left — Baseline scale floats free: Scatter plot of \log_{10} of total number of counts inferred to be in the multiscale model for each of the draws, versus \log_{10} of the baseline scale factor. Bullets indicate no signal (null hypothesis). A ‘+’ indicates results from data with a faint additional component. b) Right — Baseline scale fixed (at unity): Histogram of \log_{10} of total number of counts inferred to be in the multiscale model for the null (white) and data with faint additional component (gray).

power of the method to detect weak added source is another area that we hope to explore. (Preliminary tests, however, show that where the total difference between our usual baseline physical model—which contains several thousand counts over the whole sky—and the simulated data is a factor of 10 lower than the faintest example shown here, we see no visible detection.) Finally, although the capability for incorporating an instrument response has been built into the code, it has not been extensively tested on these data.

There are many practical directions in which we would like to improve the functionality of our method. For example, using all-sky spherical data rather than nested square pixels; allowing for a PSF or instrument response that varies over the field of view; incorporating simple parametric fitting in the baseline physical model; using the full energy information together with the imaging information; incorporating uncertainties in the physical model and/or “instrument” model; and including more complex prior physical information (say, high resolution radio data). We expect to work on some of these in the near future.

5.3. Soapbox

Traditionally, physicists receive very little formal training in probability and statistics. Our cultural habit is to think in terms of Gauss-Normal statistics and χ^2 . (“How many sigma is that result?”, we ask of a detection where, say, the expected background counts are 0.2 and measured source plus background counts are 2.) At the same time, physicists tend to be comfortable with likelihood methods, and well-versed in Monte Carlo — and so have understanding of the basic building blocks for tackling non-Gauss-Normal problems directly, rather than by approximations to older Gauss-Normal statistics, such as χ^2 .

Here, we have tackled Poisson data. It is good as one specific example of the Non-Gaussian regime. This forces us (astronomers and physicists) to *think*. We cannot rely on the symmetry and limiting properties that make, say, “filtering” or “subtracting” or “correcting” *data* mathematically/probabilistically/statistically equivalent to a thorough likelihood analysis.

Instead we must pay careful attention to the unfolding of the *model*: astrophysical + instrumental + statistical. We must leave the *data* alone; our tampering will not improve them.

In this paper, we have used hierarchical methods of taking apart and putting together models and data. This enabled us to propose a practical solution to two tough long-standing problems for Poisson data: “good-enough” fit and querying the data for evidence of an unpredicted component of unknown shape. We have at least qualified success in testing a solution.

But we argue the larger lesson is: many supposedly intractable problems can be analyzed using this same framework of carefully breaking the difficult problem apart into smaller, more manageable pieces. Likelihood-based methods such as Hierarchical Bayes offer ideal tools to this end.

Acknowledgments. The authors and CHASC acknowledge support from Chandra X-Ray Center and NSF (DMS 04-06085). AC acknowledges AISRP NAG5-12082 “Astrostatistical Tools in Python”, and the work of her co-investigators (T. Loredo and T. Oliphant). It is a particular pleasure to acknowledge the SAMSI06 Special Program in Astrostatistics. This research has made use of data obtained from the High Energy Astrophysics Science Archive Research Center (HEASARC), provided by NASA’s Goddard Space Flight Center.

References

- Bertsch, D.L., et al., 1993, “Diffuse Gamma-Ray Emission in the Galactic Plane from Cosmic-Ray, Matter, and Photon Interactions”, *Astrophysical Journal*, 416, 587.
- Bevington, P.R., 1969, *Data Reduction and Error Analysis for Physical Scientists*, (New York: McGraw Hill).
- Bourdin, H.; Slezak, E.; Bijaoui, A.; Arnaud, M., 2001, “A multiscale regularized restoration algorithm for XMM-Newton data”, in *Clusters of galaxies and the high redshift universe observed in X-rays, Recent results of XMM-Newton and Chandra, XXXVth Rencontres de Moriond, XXIst Moriond Astrophysics Meeting, March 10-17, 2001 Savoie, France*. Edited by D.M. Neumann & J.T.T. Van. /[arXiv:astro-ph/0106138](https://arxiv.org/abs/astro-ph/0106138)
- Cash, W., 1976, “Generation of Confidence Intervals for Model Parameters in X-ray Astronomy”, *Astronomy & Astrophysics* 52, 307.
- Cash, W., 1979, “Parameter estimation in astronomy through application of the likelihood ratio”, *Astrophysical Journal*, 228, 939.
- Coifman, R.R., and Donoho, D.L., 1995, “Translation-Invariant De-Noising”, in *Wavelets and Statistics. Lecture Notes in Statistics*, (Springer Verlag), 125.
- Dixon, D., et al., 1999, “Evidence for a Galactic Gamma-Ray Halo” *New Astronomy*, 3, 539.
- Dixon, D., and Kolaczyk, E., “Nonparametric Estimation of Intensity Maps Using Haar Wavelets and Poisson Noise Characteristics” *Astrophysical Journal*, 534, 490.
- Drake, J., et al., 2006, “Monte Carlo processes for including Chandra instrument response uncertainties in parameter estimation studies”, *SPIE*, 6270E, 49.
- Eadie, W. T., Drijard, D., James, F. E., Roos, M., & Sadoulet, B. 1971, *Statistical Methods in Experimental Physics* (New York: North-Holland).

- Ebeling, H.; White, D. A.; Rangarajan, F. V. N., 2006, "ASMOOTH: a simple and efficient algorithm for adaptive kernel smoothing of two-dimensional imaging data", *Monthly Notices of the Royal Astronomical Society*, 368, 65.
- Elwe, D., 2003, "Comparison of Point-Source Subtracted EGRET Data with GALPROP", *Masters Thesis*, Royal Institute of Technology Department of Physics, Stockholm.
- Esch, D., Connors, A., Karovska, M., and van Dyk, D., 2004, "An Image Restoration Technique with Error Estimates", *Astrophysical Journal*, 610, 1213.
- Esposito, J. A.; Sreekumar, P.; Hunter, S. D.; Kanbach, G., 1996, "EGRET Observations of Radio-bright Supernova Remnants" *Astrophysical Journal*, 461 820.
- Esposito, J., et al., 1999, "In-Flight Calibration of EGRET on the Compton Gamma-Ray Observatory", *Astrophysical Journal Supplement*, 123 203.
- Freeman, P. E.; Kashyap, V.; Rosner, R.; Lamb, D. Q., 2002, "A Wavelet-Based Algorithm for the Spatial Analysis of Poisson Data" *Astrophysical Journal Supplement*, 138, 185.
- Freeman, P. E.; Kashyap, V.; Rosner, R.; Lamb, D. Q., 2003, "The statistical challenges of wavelet-based source detection", *Statistical Challenges in Modern Astronomy (SCMA III)*, (Springer Verlag, New York), 365.
- Hartman, R., et al., 1999, "The Third EGRET Catalog of High-Energy Gamma-Ray Sources", *Astrophysical Journal Supplement*, 123, 79.
- Hunter, S.D., et al., 1997, "EGRET Observations of the Diffuse Gamma-Ray Emission from the Galactic Plane", *Astrophysical Journal*, 481, 205.
- Karovska, M., et al., 2005, "A Large X-Ray Outburst in Mira A", *Astrophysical Journal*, 623L,137.
- Knoedelseder et al., 1999, "Image reconstruction of COMPTEL 1.8 MeV (26) AL line data", *Astronomy & Astrophysics*, 345, 813.
- Kolaczyk, E., 1997, "Nonparametric Estimation of Gamma-Ray Burst Intensities Using Haar Wavelets", *Astrophysical Journal*, 483 340.
- Kolaczyk, E.D. and Nowak, R.D., 2004, "Multiscale likelihood analysis and complexity penalized estimation" *Annals of Statistics*,32, 500.
- Kolaczyk, E.D. and Nowak, R.D. 2005, "Multiscale generalized linear models for nonparametric function estimation" *Biometrika*, 92:1, 119.
- Lampton, M., Margon, B., and Bowyer, S., 1976, "Parameter estimation in X-ray astronomy" *Astrophysical Journal*, 208, 177.
- Longair, Malcom S. 1992 *High Energy Astrophysics – V.1 - Particles Photons and Their Detection - ED.2* (Cambridge University Press: Cambridge, England).
- Loredo, T., 1992, "The Promise of Bayesian Inference in Astrophysics", in *Statistical Challenges in Modern Astronomy*, G.J. Babu and E.D. Feigelson (eds.), (Springer Verlag, New York), 275.
- Lucy, L. B. 1974 "An iterative technique for the rectification of observed distributions", *Astronomical Journal*, 79, 745.
- Mattox, J., et al. 1992, *Proceedings of the Compton Observatory Science Workshop, Annapolis, Sept 23-35, 1991*, edited by C. Shrader, N. Gehrels, & B. Dennis, NASA conf. pup. No.3137, 126.
- Moskalenko, I.V. and Strong, A., 1998, "Production and Propagation of Cosmic-Ray Positrons and Electrons", *Astrophysical Journal*, 493, 694.
- Moskalenko, I.V. and Strong, A., Reimer, O., 1998, "Diffuse galactic gamma rays, cosmic-ray nucleons and antiprotons", *Astronomy and Astrophysics Letters*, 338L, 75.
- Nolan et al., 1992, "Performance of the EGRET astronomical gamma ray telescope" *IEEE Transactions on Nuclear Science*, 39, 993.
- Nowak, R.D. and Kolaczyk, E.D., 2000, "A Bayesian multiscale framework for Poisson inverse problems" *IEEE Transactions on Information Theory*, 46:5, 1811.
- Pina, R. K.; Puetter, R. C., 1993, "Bayesian image reconstruction - The pixon and optimal image modeling", *Publications of the Astronomical Society of the Pacific*, 105, 630.
- Press, W., Teukolsky, S., Vetterling, W., and Flannery, B., 1992, *Numerical Recipes, 2nd Edition*, (Cambridge University Press: New York).

- Protassov, R., et al., 2002, "Statistics, Handle with Care: Detecting Multiple Model Components with the Likelihood Ratio Test" *Astrophysical Journal*, 571, 545.
- Richardson, W. H., 1972, "Bayesian-based iterative method of image restoration", *J. Opt. Soc. Am.*, 62, 55.
- Scargle, J., this volume
- Shepp, L.A, and Vardi, Y., 1982, "Maximum likelihood reconstruction for emission tomography", *IEE Trans. Med. Imaging* M1-2, 113.
- Skilling, J., 1989, "Maximum entropy and bayesian methods", in *Proceedings of the 8th Workshop on Maximum Entropy and Bayesian Methods*, J. Skilling, ed., (Dordrecht: Kluwer).
- Starck, J. L.; Pantin, E.; Murtagh, F, 2002, "Deconvolution in Astronomy: A Review", *The Publications of the Astronomical Society of the Pacific*, Volume 114, Issue 800, 1051.
- Starck, J.-L.; Murtagh, F., 2006, *Astronomical Image and Data Analysis*, (Berlin: Springer).
- Strong, A. and Moskalenko, I.V., 1998, "Propagation of Cosmic-Ray Nucleons in the Galaxy", *Astrophysical Journal*, 509, 212.
- Strong, A.W., 2003, "Maximum Entropy imaging with INTEGRAL/SPI data", *Astronomy and Astrophysics Letters*, 411, L127.
- Strong, A.W., Moskalenko, I.V., Reimer, O., "Diffuse Galactic continuum gamma rays. A model compatible with EGRET data and cosmic-ray measurements" 2004, *Astrophysical Journal*, 613, 962.
- Strong, A.W., Moskalenko, I.V. Reimer, O., Digel, S., Diehl, R., 2004, "The distribution of cosmic-ray sources in the Galaxy, gamma-rays and the gradient in the CO-to-H2 syrelation" *Astronomy & Astrophysics*, Letters, 422, L47.
- Strong, A.W., Diehl, R., Halloin, H., Schoenfelder, V., Bouchet, L., Mandrou, P., Lebrun, F., Terrier, R., 2005, "Gamma-ray continuum emission from the inner Galactic region as observed with INTEGRAL/SPI" *Astronomy & Astrophysics*, 444, 495.
- Thompson, D.J., et al. 1993, "Calibration of the Energetic Gamma-Ray Experiment Telescope (EGRET) for the Compton Gamma-Ray Observatory" *Astrophysical Journal Supplement*, 86, 629
- Willett, R., 2006, this volume.
- Young, C.A., 2006, this volume