

# Bayesian Poisson Image Smoothing using a Chinese Restaurant Process with Applications in Astronomy

John T. White and Subhashis Ghosal

North Carolina State University

June 10, 2009

## Abstract

Many multi-scale approaches to smoothing images with Poisson noise are currently available, including various wavelet smoothing methods, wedgelets, and platelets. Another class of smoothing algorithms utilizes Bayesian multi-scale models. In this paper, a novel method for smoothing images with Poisson data is introduced. Our method is based on a Bayesian multi-scale modeling with the Chinese Restaurant Process, which is a sampling scheme closely related to the Dirichlet Process. More specifically, this new method combines existing ideas of Bayesian multi-scale modeling for smoothing images, but uses a combination of the Chinese Restaurant Process and Dirichlet distributions to control intensity of neighboring pixels. We provide independent samples from the posterior distribution of the underlying intensity, or true object representation. This new method still permits different degrees of smoothing at the various scales of data.

In order to compare this new methodology to the current methods, simulation examples using different photon-based images were conducted. Photon-based images are common in astronomical imaging due to the collection of different types of energy such as X-Rays. Applications to real astronomical images are given, and these consist of X-ray images from the Chandra X-ray Observatory satellite. X-ray images take a count of photons in each pixel and hence can be naturally modeled as Poisson data. Not only are these methods useful in the astronomical setting, but also in imaging systems used for diagnostic medicine. This paper shows this method outperforms many existing methods in its ability to estimate the true intensity image, as well as having other favorable qualities that some methods lack, such as obtaining smoothing parameters from the data and a relatively fast computing time.

# 1 Introduction

One type of image restoration problem addresses photon-based images, which have been modeled in statistics as a Poisson process due to the collection of photons from a detection instrument. An image consists of pixels and, more specifically, a count of photons is obtained in each pixel. For some types of images, these counts can be relatively small. An example of a photon-based image would be an astronomical X-Ray image that has a count of X-ray photons in each pixel. These photons are collected by a detection device and can be distorted due to two common factors: error or noise in collecting the photons and, in some cases, the point spread function, which convolutes the observed data and is specific to that device. A common way to resolve the first issue is to perform some type of image restoration or reconstruction technique to denoise or smooth the image. This paper focuses on smoothing, or accounting for the error in the collection of photons instead of the point spread function, or deconvolution problem.

From a modeling perspective these pixels, or counts of photons, have an underlying intensity or true form that represents the object being examined. Typically, when an instrument is used to capture this underlying intensity, the resulting observed image is not free of noise. Consider the observed image as a Poisson sample with an intensity value associated with every pixel in the image. This observed image will have noise depending on the magnitude of the underlying intensity. Other modeling efforts have used Gaussian noise, but with these types of images, especially when there is a limited number of photons collected by the detection device, a Poisson model is a much better representation of the data. In an effort for the researcher to achieve a better idea of what the true representation of the object is, an estimate is obtained of the underlying intensity.

Many methods exist to perform this type of image reconstruction with Poisson noise. Of these methods, a very popular class uses multi-scale representa-

tions of the images, such as wavelet transforms. These multi-scale representations of images provide a way to show structure in very detailed or fine objects and other larger or coarse objects in the image simultaneously. In this paper, a new Bayesian multi-scale model is proposed that uses Dirichlet priors with a Chinese Restaurant Process to produce an estimate of the true intensity image. This method has many characteristics of the methods described by Kolaczyk (1999), Nowak and Kolaczyk (2000), and Esch et al. (2004). There are some key differences in the structure of the model that will be discussed, giving this method advantages over other Bayesian models for this problem.

Other methods that will be used for comparison are platelets (Willett and Nowak 2003), another multi-scale image reconstruction technique that furthers the ideas of wedgelets (Donoho 1999), a multi-scale penalized likelihood based method introduced in Kolaczyk and Nowak (2004), and a translation-invariant Haar wavelet method using Poisson-corrected thresholds described in Kolaczyk (1997) and Kolaczyk (1998).

In §2, the new Bayesian multi-scale statistical model used in this image reconstruction process is described in detail. Further theoretical results are provided in §3. The new method is compared to other methods used for smoothing Poisson images using test images where the maximum intensity can be decided and the true relative intensity is known. Images and results from these simulations are given in §4. Examples in astronomical imaging using the Chandra X-ray observatory are shown in §5. Finally, conclusions and remarks are given in §6. A basic introduction to the Chinese Restaurant Process, which is useful in understanding the model, is given in an appendix.

## 2 Model and Methodology

### 2.1 Bayesian Multi-Scale Statistical Model

Multi-scale representations of images decompose the image into different scales of data from coarse to fine with the extremes in this case being the total count of photons or photon flux in the image and the number of photons present in each pixel of the image. The reason for this is that both very fine detailed aspects of the image can be observed and utilized in image restoration, as well as larger general objects of the image. Models that do not use multi-scale representations of image might over-smooth the very small detailed aspects of an image while preserving the overall general object or vice versa. This representation will allow for different levels of smoothing to be used at the different scales of the image. Multi-scale representations have been used in many image reconstruction methods. Some of these include platelets, any wavelet method, wedgelets, and other Bayesian multi-scale methods. Much of the information provided in §2.1 is also given in Kolaczyk (1999).

One common way to decompose an image in this form is to start with the image as one total count of photons. This block is then sectioned into four smaller equal-sized blocks. Continue to partition these blocks into four equal cells until the pixel scale data is obtained. To do this decomposition, the image must have pixels on both sides of the form  $n = 2^L$  where  $L$  is an integer that denotes the highest or finest scale in the decomposition and  $N = n^2 = 4^L$  is the total number of pixels in the image.

Let  $x_{(j,k)} = x_{L,(j,k)}$  denote the photon count in pixel  $(j, k)$  and  $x_{0,(1,1)}$  denote the total count of photons or photon flux of the image. At intermediate scales  $l = 1, \dots, L - 1$ , there are  $4^l$  block-pixels that are denoted by  $(j, k)$  where  $j, k = 1, \dots, 2^l$ . To obtain the intermediate scale data, the following formula is

used.

$$x_{l,(j,k)} = x_{l+1,(2j-1,2k-1)} + x_{l+1,(2j-1,2k)} + x_{l+1,(2j,2k-1)} + x_{l+1,(2j,2k)}, \quad (1)$$

where  $j, k = 1 \dots 2^l$ ;  $l = 0, \dots, L - 1$ .

Typically, the left side of the equation is referred to as the parent while the right side consists of the children. Collectively this is called a parent-child group. This decomposition can also be thought of as a quad-tree where every node is split into four parts as you progress down the tree. The top would be the photon flux, and the bottom would be the pixel-scale data. Each row of nodes would be a different scale of data. A statistical model is developed modeling all scales of the data.

The observed image data is comprised of pixels in which there are counts of photons. These counts are modeled with a poisson distribution such that  $x_{(j,k)} \sim \text{Poi}(\lambda_{(j,k)})$ , where  $(j, k)$  denotes the pixel with  $j$  being the horizontal index and  $k$  being the vertical index. The underlying intensity  $\lambda$  represents the true object that is being examined. While these images typically have background noise, that will be ignored since this background is usually of much smaller magnitude than the object that is being observed. Methodology that allows for background noise in the model is discussed in the conclusions as future work. Images can be of many different sizes. Thus, the image must be truncated or blank pixels must be added to make each side a length of  $2^L$ .

The entire image can be thought of by  $X \sim \text{Poisson}(\Lambda)$ . A multi-scale statistical model, which is also used in Nowak and Kolaczyk (2000), Kolaczyk (1999) and Esch et al. (2004), is given by factoring this simple statistical model

into the following:

$$P(X|\Lambda) = \mathcal{P}(x_{0,(1,1)}|\lambda_{0,(1,1)}) \times \prod_{l=0}^{L-1} \prod_{j=1}^{2^l} \prod_{k=1}^{2^l} \mathcal{M} \left( \begin{matrix} x_{l+1,(2j-1,2k-1)} \\ x_{l+1,(2j-1,2k)} \\ x_{l+1,(2j,2k-1)} \\ x_{l+1,(2j,2k)} \end{matrix} \middle| x_{l,(j,k)}, \begin{matrix} \rho_{l,(j,k)}^1 \\ \rho_{l,(j,k)}^2 \\ \rho_{l,(j,k)}^3 \\ \rho_{l,(j,k)}^4 \end{matrix} \right), \quad (2)$$

where  $\mathcal{P}$  represents the Poisson distribution,  $\mathcal{M}$  the Multinomial distribution, and  $\rho$  is a  $4 \times 1$  vector of probabilities of splitting the parent into its children. The  $\rho$ 's and  $\lambda$ 's have a one-to-one relationship, therefore, knowing the  $\rho$ 's will give  $\lambda$  by the following:

$$\begin{aligned} \rho_{l,(j,k)}^1 \lambda_{l,(j,k)} &= \lambda_{l+1,(2j-1,2k-1)}, \\ \rho_{l,(j,k)}^2 \lambda_{l,(j,k)} &= \lambda_{l+1,(2j-1,2k)}, \\ \rho_{l,(j,k)}^3 \lambda_{l,(j,k)} &= \lambda_{l+1,(2j,2k-1)}, \\ \rho_{l,(j,k)}^4 \lambda_{l,(j,k)} &= \lambda_{l+1,(2j,2k)}. \end{aligned} \quad (3)$$

To find an estimate of  $\Lambda$ , the values of  $\rho$  must be estimated. Instead of directly obtaining estimates for the intensities, an estimate is obtained of the split probabilities,  $\rho$ , and this relationship is exploited to obtain the intensities. A pixel scale true intensity estimator can be represented as  $\hat{\lambda}_{j,k} = E(\lambda_{0,(1,1)}\rho_0\rho_1\dots\rho_{L-1}|\mathbf{X})$ , which is the posterior mean of the pixel scale intensity since  $\lambda_{0,(1,1)}\rho_0\rho_1\dots\rho_{L-1} = \lambda_{(j,k)}$ .

## 2.2 Prior Distributions

The method of finding the estimates for the split probabilities sets this method apart from other Bayesian multi-scale models. More specifically, a key difference is the construction of the priors. The Chinese Restaurant Process (CRP) is incorporated into the prior for the  $\rho$ 's. The CRP, as discussed in the appendix,

allows with positive probability for groups of the  $\rho$ 's to equal each other within a parent-child group. In other words, equal splits are possible to different combinations of the children for certain block-pixels. The motivation behind this choice was that it can be assumed the  $\rho$ 's are coming from a continuous distribution with positive probability that some of them will have equal values.

The configuration of  $\rho$ 's that are grouped together is denoted by  $\mathcal{C} \in \mathcal{C}$ , where  $\mathcal{C}$  is one of the configurations of ties or lack thereof and  $\mathcal{C}$  is the space of all 15 possible configurations of groups of the  $\rho$ 's. The most simple is an equal split into each of the four cells so that all  $\rho$ 's are equal in a parent-child group or  $\rho^{(1)} = \rho^{(2)} = \rho^{(3)} = \rho^{(4)} = \frac{1}{4}$ .

This allows for optimal smoothing in very fine scales when there is no significant difference between neighboring pixels. Having the same  $\rho$ 's will ensure the same  $\lambda$ 's for children of the same parent-child group. The other easy configuration is when all  $\rho$ 's are unique. This occurs when the children are very different from one another in a group. All other configurations are also used. There probabilities are given as  $P(\mathcal{C}|M) \sim \text{CRP}(M)$ , where  $M$  is the parameter of the CRP described in the appendix. There are 15 possible configurations of the split probabilities. Derived from information in the appendix, some examples of configurations and their probabilities are as follows:

$$P(\mathcal{C} = \{(1, 2), (3, 4)\}|M) = \frac{M}{M} \times \frac{1}{M+1} \times \frac{M}{M+2} \times \frac{1}{M+3},$$

$$P(\mathcal{C} = \{(1, 2, 3, 4)\}|M) = \frac{M}{M} \times \frac{1}{M+1} \times \frac{2}{M+2} \times \frac{3}{M+3},$$

$$P(\mathcal{C} = \{(1), (2), (3), (4)\}|M) = \frac{M}{M} \times \frac{M}{M+1} \times \frac{M}{M+2} \times \frac{M}{M+3}.$$

A particular configuration can be written as  $\mathcal{C} = \{C_1, \dots, C_s\}$ , where  $s$  is the number of groups in the configuration. Once a configuration is found for a parent-child group, a prior can be specified for the split probabilities of that

parent-child group. This is represented as

$$(\rho_{l,(j,k)}^{(1)}, \rho_{l,(j,k)}^{(2)}, \rho_{l,(j,k)}^{(3)}, \rho_{l,(j,k)}^{(4)}) = H_{\mathcal{C}}(q_{l,(j,k)}^{(1)}, \dots, q_{l,(j,k)}^{(s)}) \quad (4)$$

To illustrate how this function  $H_{\mathcal{C}}$  works, some examples are given:

If  $\mathcal{C} = \{(1), (2), (3, 4)\}$ , then  $\rho_{l,(j,k)}^{(1)} = q_{l,(j,k)}^{(1)}$ ,  $\rho_{l,(j,k)}^{(2)} = q_{l,(j,k)}^{(2)}$ ,  $\rho_{l,(j,k)}^{(3)} = \rho_{l,(j,k)}^{(4)} = \frac{q_{l,(j,k)}^{(3)}}{2}$ .

If  $\mathcal{C} = \{(1, 2), (3, 4)\}$ , then  $\rho_{l,(j,k)}^{(1)} = \rho_{l,(j,k)}^{(2)} = \frac{q_{l,(j,k)}^{(1)}}{2}$ ,  $\rho_{l,(j,k)}^{(3)} = \rho_{l,(j,k)}^{(4)} = \frac{q_{l,(j,k)}^{(2)}}{2}$ .

If  $\mathcal{C} = \{(1, 2, 3), (4)\}$ , then  $\rho_{l,(j,k)}^{(1)} = \rho_{l,(j,k)}^{(2)} = \rho_{l,(j,k)}^{(3)} = \frac{q_{l,(j,k)}^{(1)}}{3}$ ,  $\rho_{l,(j,k)}^{(4)} = q_{l,(j,k)}^{(2)}$ .

If  $\mathcal{C} = \{(1, 2, 3, 4)\}$ , then  $\rho_{l,(j,k)}^{(1)} = \rho_{l,(j,k)}^{(2)} = \rho_{l,(j,k)}^{(3)} = \rho_{l,(j,k)}^{(4)} = \frac{1}{4}$ .

The prior distribution of  $q$  is given as the following Dirichlet distribution:

$$P(q_{l,(j,k)}^{(1)}, \dots, q_{l,(j,k)}^{(s)} | \mathcal{C} \in \mathcal{C}) \sim \text{Dirichlet}(s; \aleph_{C_1}, \dots, \aleph_{C_s}), \quad (5)$$

where  $\aleph$  represents the cardinality of a set and  $s$  is the number of distinct groups.

This prior distribution was chosen to allow  $M$  from the CRP to control all of the amount of equal splitting among the four split probabilities and to give equal weight a priori to all of the block-pixels in the parent-child group. Again, some examples are given to illustrate this construction.

If  $\mathcal{C} = \{(1), (2), (3, 4)\}$ , then  $P(q_{l,(j,k)}^{(1)}, q_{l,(j,k)}^{(2)}, q_{l,(j,k)}^{(3)} | \mathcal{C}) \sim \text{Dirichlet}(3; 1, 1, 2)$ .

If  $\mathcal{C} = \{(1, 2), (3, 4)\}$ , then  $P(q_{l,(j,k)}^{(1)}, q_{l,(j,k)}^{(2)} | \mathcal{C}) \sim \text{Dirichlet}(2; 2, 2)$ .

If  $\mathcal{C} = \{(1, 2, 3), (4)\}$ , then  $P(q_{l,(j,k)}^{(1)}, q_{l,(j,k)}^{(2)} | \mathcal{C}) \sim \text{Dirichlet}(2; 3, 1)$ .

If  $\mathcal{C} = \{(1, 2, 3, 4)\}$ , then  $\rho = \{.25, .25, .25, .25\}$ .

The parameter  $M$  drives the smoothing in this image reconstruction technique. Thus, determining  $M$  is of great importance. In general, decreasing the value of  $M$  will create more ties among the  $\rho$ 's while increasing its value will create more distinct groups.

An important property of an image reconstruction technique is to keep the total photon flux of the original image. Therefore, instead of placing a prior

on the overall intensity, it is set equal to the total photon flux of the observed image. Thus we set  $\lambda_{0,(1,1)}$  to  $x_{0,(1,1)}$  rather than putting a prior on  $\lambda_{0,(1,1)}$ . Once all of the priors are specified, the posteriors can be obtained to perform this image reconstruction technique.

## 2.3 Posterior Distributions

Assume, for now, that  $M$  is known. The overall goal is to find an estimate of  $E(\lambda_{(j,k)}|X)$ , where  $\lambda_{(j,k)} = \lambda_{0,(1,1)}\rho_0\rho_1\rho_2\dots\rho_{L-1}$  in corresponding block-pixels down to the pixel-scale data. We denote  $X$  as the data in general.

To obtain a sample from the posterior  $\lambda_{(j,k)}|X$ , each scale's  $\rho$  of that specific pixel scale intensity parameter must be sampled as well as each configuration corresponding to its parent-child groups since that is used in determining the prior and thus posterior distributions of  $\rho$ . To do this, it is necessary to find two general posteriors: the discrete distribution  $P(\mathcal{C}|M, X)$  and the continuous distribution  $P(q_{l,(j,k)}|\mathcal{C} \in \mathcal{C}, X)$ , both of which have the same general form for each parent-child group.

The configuration must be found first. Therefore, start with the discrete distribution in a given parent-child group. Let  $(x_1, x_2, x_3, x_4)$  denote the children of that group and  $x$  denote the parent or the sum of the children. Then we can express this distribution as follows:

$$P(\mathcal{C}|M, X) \propto P(\mathcal{C}|M)P(x_1, x_2, x_3, x_4|\mathcal{C}, x). \quad (6)$$

Since this is a discrete probability distribution, the values must sum to 1 giving the proportionality constant. The first term,  $P(\mathcal{C}|M)$ , is given by the CRP. The second term,  $P(x_1, x_2, x_3, x_4|\mathcal{C}, x)$ , is found using the following integration:

$$P(x_1, x_2, x_3, x_4|\mathcal{C}, x) \propto \int_{\Delta_s} P(x_1, x_2, x_3, x_4|\mathcal{C}, x, \rho)P(q|\mathcal{C})dq, \quad (7)$$

where  $\Delta_s$  stands for the unit simplex in  $\mathbb{R}^s$ . This integration is performed for

each possible configuration to get a closed form solution. As an example, take  $\mathcal{C} = \{(1, 2)(3, 4)\}$ , to have:

$$\begin{aligned} & P(x_1, x_2, x_3, x_4 | x, \mathcal{C}, \rho^{(1)} = \rho^{(2)} = \frac{q^{(1)}}{2}, \rho^{(3)} = \rho^{(4)} = \frac{q^{(2)}}{2}) P(q^{(1)}, q^{(2)} | \mathcal{C}) \\ &= \frac{x!}{x_1! x_2! x_3! x_4!} \left(\frac{q^{(1)}}{2}\right)^{x_1+x_2} \left(\frac{q^{(2)}}{2}\right)^{x_3+x_4} \frac{\Gamma(4)}{\Gamma(2)\Gamma(2)} q^{(1)2-1} q^{(2)2-1} \end{aligned}$$

Integrate out  $q$  to get:

$$\begin{aligned} & P(x_1, x_2, x_3, x_4 | \mathcal{C}, x) \\ &= 6 \frac{\Gamma(x_1 + x_2 + 2)}{\Gamma(x_1 + 1)\Gamma(x_2 + 1)} \times \frac{\Gamma(x_3 + x_4 + 2)}{\Gamma(x_3 + 1)\Gamma(x_4 + 1)} \times \frac{\frac{1}{2}^x}{(x + 3)(x + 2)(x + 1)} \\ &= 6 \frac{1}{\text{Beta}(x_1, x_2)} \times \frac{1}{\text{Beta}(x_3, x_4)} \times \frac{\frac{1}{2}^x}{(x + 3)(x + 2)(x + 1)} \\ &= 6 \frac{\exp[-\text{logbeta}(x_1, x_2) - \text{logbeta}(x_3, x_4) - x \ln 2]}{(x + 3)(x + 2)(x + 1)}. \end{aligned} \tag{8}$$

The logbeta function (logarithm of the beta function) is used to help with computation since the values of  $x$  can become very small or large. By repeating this process, all of the probabilities of configuration given the data and our smoothing parameter  $M$ , are obtained.

Given an  $M$ , we can generate independent samples from the discrete distribution  $P(\mathcal{C} | M, X)$  for a parent-child group. Now  $P(q_{l,(j,k)} | \mathcal{C}, X)$  is simply a Dirichlet distribution by conjugacy since the likelihood of the data is a multinomial distribution. For instance, if  $\mathcal{C} = \{(1, 2)(3, 4)\}$ , then  $(q^{(1)}, q^{(2)} | x_1, x_2, x_3, x_4, \mathcal{C}) \sim \text{Dirichlet}(2; 2+x_1+x_2, 2+x_3+x_4)$ . Thus, independent samples can be generated from this known distribution given a configuration  $\mathcal{C}$ .

Having the configuration allows the generation of the  $q$  parameters from the Dirichlet distribution, which will then give the  $\rho$  parameters. Finally, from the one-to-one relationship described above, the  $\lambda$  parameters are obtained for the next finer scale. This continues block-pixel by block-pixel, scale by scale until the pixel-scale  $\lambda$  parameters are obtained. As described above, each sample from this overall posterior is independent from one another since we

are generating independent samples from the two posteriors. The posterior mean is used to give the final smoothed image. Since each iteration produces an independent sample from the posterior, only relatively few iterations need to be performed. Even though fewer iterations are needed, we shall need cycle spinning of the original image due to the multi-scale transform so that staircase-like artifacts are not created. This is discussed in §2.5.

The original idea of this type of Bayesian multi-scale model is given in Kocolczyk (1999), where the split probabilities had a non-singular Dirichlet prior and the smoothed image was given as the MAP estimate of the posterior. By incorporating ties using the CRP in our new prior, a better way of smoothing the neighboring pixels is obtained. The smoothing parameters in this earlier method were the hyperparameters of the Dirichlet prior. In Esch et al. (2004), they take the idea further and use the mean of the posterior for an estimate of the intensity image and also place a hyper-prior distribution on the Dirichlet parameters. This leads to MCMC methods, which increases computing time dramatically and introduces other problems inherent with using MCMC algorithms, such as convergence diagnostics. They still do not allow for ties among the split probabilities. We agree with Esch et al. (2004) that the mean of the posterior should be used as the smoothed image. However, our methodology achieves smoothing without the use of MCMC. Since our samples are independently drawn from the posterior and MCMC methods must wait for the samples to converge to the posterior, our methodology requires much fewer iterations.

Esch et al. (2004) also introduce the use of significance maps that use the standard deviation of each pixel-estimate. Due to the generation of independent samples from the posterior, our methodology is capable of providing an error estimate for the pixel-scale intensity estimate allowing for other types of inference.

In this image smoothing procedure, the tuning parameter  $M$  decides how much smoothing will be done. The next section describes how these smoothing parameters are obtained using the data.

## 2.4 Selection of Smoothing Parameters

In developing this algorithm, two different methods of choosing  $M$  were used and compared. In both methods,  $M$  is different for the multiple scales of data, which allows for adjusting the levels of smoothing in each of the scales. It is intuitively obvious that the value for  $M$  should get smaller as the scale gets finer or  $l$  increases. An empirical estimate of  $M$  based on the likelihood is used as well as a full Bayesian method that puts a gamma prior on  $M$ , where  $\alpha$  and  $\beta$  are hyper-parameters that will not affect the level of smoothing as much as directly changing the values of  $M$ . In this second method, to find the configuration given  $M$ , integrate  $M$  out and vary  $\alpha$  and  $\beta$ .

The goal, after finding the tuning parameters in each case, is to obtain a configuration given the data. This distribution of the configuration is given by

$$P(\mathcal{C}|x_1, x_2, x_3, x_4, M) \propto P(\mathcal{C}|M)P(x_1, x_2, x_3, x_4|\mathcal{C}, x), \quad (9)$$

or

$$P(\mathcal{C}|x_1, x_2, x_3, x_4, \alpha, \beta) \propto P(x_1, x_2, x_3, x_4|\mathcal{C}, x) \int_0^\infty P(\mathcal{C}|M)P(M|\alpha, \beta)dM. \quad (10)$$

In either case,  $P(x_1, x_2, x_3, x_4|\mathcal{C}, x)$  has already been obtained in the last section. The methods for obtaining the smoothing parameters either optimize for a value of  $M$  directly, or place a hyper-prior on  $M$  and optimize for those parameters.

To find  $M$  directly, the marginal probability of obtaining the given sample is maximized using a Newton-Raphson algorithm:

$$\sum_{\mathcal{C} \in \mathcal{C}} P(\mathcal{C}|M)P(x_1, x_2, x_3, x_4|\mathcal{C}, x). \quad (11)$$

When using the hyper-prior, the obstacle is to find the optimal choices for

$\alpha$  and  $\beta$ , or maximize the following function for these hyper-parameters.

$$\sum_{\mathcal{C} \in \mathcal{C}} \mathbb{P}(x_1, x_2, x_3, x_4 | \mathcal{C}, x) \int_0^\infty \mathbb{P}(\mathcal{C} | M) \mathbb{P}(M | \alpha, \beta) dM \quad (12)$$

The integral in the right hand part of the sum can be done for each parent-child group since  $M$  is allowed to vary in a scale coming from the same distribution. To calculate this integral, analytical computation is actually possible up to some extent in the sense that we can reduce the integral to an expression involving only library functions in a standard software like MATLAB. There are five distinct values of the CRP probabilities. Their difference depends on the exponent term of  $M$  and a constant. Thus, we have the following form:

$$\begin{aligned} & \int_0^\infty \mathbb{P}(\mathcal{C} | M) \mathbb{P}(M | \alpha, \beta) dM \\ & \propto \int_0^\infty \frac{M^k}{(M+1)(M+2)(M+3)} M^{\alpha-1} e^{-M\beta} dM \\ & = \int_0^\infty \frac{M^{k+\alpha-1} e^{-M\beta}}{(M+1)(M+2)(M+3)} dM \\ & = \int_0^\infty \frac{M^{k+\alpha-1} e^{-M\beta}}{2(M+3)} dM - \int_0^\infty \frac{M^{k+\alpha-1} e^{-M\beta}}{(M+2)} dM \\ & + \int_0^\infty \frac{M^{k+\alpha-1} e^{-M\beta}}{2(M+1)} dM \end{aligned} \quad (13)$$

It will suffice to solve the integral below with  $k^* = k + \alpha - 1$  from above since

each term in the sum is of the same form.

$$\begin{aligned}
& \int_0^\infty \frac{M^{k^*} e^{-M\beta}}{(M+j)} dM \\
&= \int_j^\infty \frac{(t-j)^{k^*} e^{-\beta(t-j)}}{t} dt \\
&= e^{\beta j} \int_j^\infty \sum_{r=0}^{k^*} \binom{k^*}{r} t^{r-1} j^{k^*-r} (-1)^{k^*-r} e^{-\beta t} dt \\
&= e^{\beta j} \sum_{r=0}^{k^*} \binom{k^*}{r} j^{k^*-r} (-1)^{k^*-r} \int_j^\infty t^{r-1} e^{-\beta t} dt \\
&= e^{\beta j} (-j)^{k^*} \int_j^\infty \frac{e^{-\beta t}}{t} dt \\
&+ e^{\beta j} \sum_{r=1}^{k^*} \binom{k^*}{r} (-j)^{k^*-r} \int_j^\infty t^{r-1} e^{-\beta t} dt \tag{14}
\end{aligned}$$

The integral in the first term is the exponential integral, and the integral in the second term is the incomplete gamma function, both of which are available as standard library functions in most mathematical software. Even though this cannot be solved analytically, using standard library functions is much faster than a numerical integration of the above expression. Substituting expression (14) into (13), we optimize for  $\beta$  and restrict  $\alpha$  to the integers since it is used in the sum. The probability of each configuration is proportional to (10) giving a method of sampling from this discrete distribution.

Both of these methods above are shown for one parent-child group. Since the smoothing parameter distributions do not involve the data, these will be the same for each parent-child group. Although outlined above for one parent-child group, obtaining smoothing parameters separately in each of these groups for a scale is not possible due to the severe lack of data. Instead, we will utilize all parent-child groups in a given scale to obtain smoothing parameters for that scale.

In the first scale of the data, there would be only one parent-child group: however, take the second scale. There are four different parent-child groups. The following is applicable to either method of finding smoothing parameters,

but uses the notation of the first method based on finding an  $M$  for notational simplicity. Denote  $\mathcal{C}_1$  as the configuration in the first group, and let

$$P_{(1,1)} = P(\mathcal{C}_{(1,1)}|M)P(x_{2,(1,1)}, x_{2,(1,2)}, x_{2,(2,1)}, x_{2,(2,2)}|\mathcal{C}_{(1,1)}, x_{1,(1,1)}). \quad (15)$$

Similarly define  $P_{(1,2)}$ ,  $P_{(2,1)}$ , and  $P_{(2,2)}$ . To find an optimal  $M$  for the entire scale, the following sum would have to be maximized:

$$\sum_{\mathcal{C}_{(1,1)} \in \mathcal{C}} \sum_{\mathcal{C}_{(1,2)} \in \mathcal{C}} \sum_{\mathcal{C}_{(2,1)} \in \mathcal{C}} \sum_{\mathcal{C}_{(2,2)} \in \mathcal{C}} P_{(1,1)}P_{(1,2)}P_{(2,1)}P_{(2,2)} \quad (16)$$

This would take  $15^{4^{2-1}}$  operations since there are 15 different configurations in each of the 4 parent-child groups, and 2 is the scale of data. However, (16) can be factored into:

$$\sum_{\mathcal{C}_{(1,1)} \in \mathcal{C}} P_{(1,1)} \times \sum_{\mathcal{C}_{(1,2)} \in \mathcal{C}} P_{(1,2)} \times \sum_{\mathcal{C}_{(2,1)} \in \mathcal{C}} P_{(2,1)} \times \sum_{\mathcal{C}_{(2,2)} \in \mathcal{C}} P_{(2,2)}, \quad (17)$$

which only takes  $4^{2-1} \times 15$  operations. More generally, there will  $15^{4^{l-1}}$  operations when doing the computation as in (16), where  $l$  is the scale of data; however, because of the factorization in (17), the number of operations reduces to  $4^{l-1} \times 15$ . This a substantial gain in computation efficiency when the  $L$  can be as high as 10.

Instead of summing over all possible configurations for the entire scale of data, each parent-child group can be summed over configurations and then multiplied by all other parent-child groups. The log of this product is used to optimize for a value of  $M$  or  $\alpha$  and  $\beta$  across a scale of the data.

$$\sum_{z=1}^{4^{l-1}} \log \left( \sum_{\mathcal{C}_z \in \mathcal{C}} P(\mathcal{C}|M)P(x_1, x_2, x_3, x_4|\mathcal{C}, x) \right),$$

where  $z$  represents the parent-child groups in scale  $l$ .

The two methods based on (9) and (10) respectively performed almost the

same, but we prefer the second due to an increase in speed. In some cases, one did better than the other, but they were never far apart. The smoothing method based on choosing  $M$  directly is more sensitive to the choice of parameters than choosing hyper-parameters  $(\alpha, \beta)$  in the prior for  $M$ . Additionally, the idea of allowing  $M$  to vary in a scale is more appealing.

Once the smoothing parameters are chosen, regardless of which method is used to find them, a discrete distribution is obtained and used to generate an observation from the posterior of the configuration given the data and smoothing parameters in each parent-child group. Once a configuration is obtained, the split probabilities can be sampled from the appropriate Dirichlet distribution for each parent-child group. The split probabilities for all parent-child groups are found in a given scale and then the next finer scale split probabilities can be obtained. This recursive pattern continues until the finest scale or pixel scale split probabilities are found and thus intensity estimates are obtained. The whole process is repeated to generate an independent sample from the posterior distribution, and then this sample is averaged to obtain a smoothed image for the object of interest.

It is important to note that most other methodologies allow for the user to vary the smoothing parameters themselves, but without knowing the underlying models used, it is difficult to understand what these numbers represent. Having this as a built-in feature, we feel, was more attractive to a researcher who is not involved in the actual development of the image reconstruction methodology and is more interested in its application.

## 2.5 Cycle Spinning

Cycle spinning is a common practice that removes artifacts due to the quad partitioning when transforming the data into multiple scales. As seen in many other methodologies and discussed in Coifman and Donoho (1995), cycle spinning takes away these staircase-like artifacts from doing a Haar wavelet transform or a multi-scale factorization, as done in this paper.

This technique involves circularly shifting the image to the left or right and up or down so that the top left cell in the matrix of pixel-counts is different. This shift is made before the image reconstruction and it is inverted afterwards. Just as iterations of the independent sample are averaged to obtain an estimate, the different shifts are averaged after inverting the image back to its original position. There does not seem to be any benefit of many samples from one shift; therefore, each shift is sampled once giving better results with more shifts. If all possible shifts are made then the method becomes translation invariant, but this requires implementing the process at each pixel, which becomes computationally huge. Willett and Nowak (2004) has also proposed another method of obtaining a translation invariant estimate using traditional undecimated wavelet thresholding, but this not used in our methodology because there are sampled values in each iteration.

### 3 Consistency

The Bayesian model of this image reconstruction technique has been developed. To prove this method works, theoretical and empirical properties also need to be investigated. Consistency is a basic, but very important theoretical property of an estimator. If it were possible to expose the image for an infinite amount of time, would the solution image converge to the actual object? The actual object is more of a relative percent intensity since, the longer we expose the detector to the object, the larger in magnitude the overall intensity. Therefore, it is the relation of all of the intensities that make up an object that give it its true representation. Instead of needing the intensity to be converging to the true intensity, we need the split probabilities to converge to their true values, which will then preserve this relative magnitude of intensity values. Note that, given the number of photons the model is multinomial.

The idea of posterior consistency says that the posterior distribution of  $\rho$  is consistent at  $\rho_0$ , the true value of this parameter; that is, it concentrates

in neighborhoods of  $\rho_0$  as numbers of observations increase. It is well known that when the observations are from a well-behaved finite dimensional family, multinomial in this case, and there is positive probability in the prior for all neighborhoods of any possible value of  $\rho$ , then posterior consistency holds true. For instance, in our case, this follows from the results of Schwartz (1965). Since in a multinomial model, parameter value can be tested with exponentially small error probability uniformly against the complement of any neighborhood and since in our prior specification, a nonsingular component is always present, Schwartz's theorem applies giving consistency. In fact, posterior probability of the complement of any neighborhood is exponentially small in  $n$ , which in our case, is the total number of photons.

Other methodologies, such as the earlier Bayesian multi-scale methodologies, will also have this property of consistency. In addition, there are the issues of detecting structures and removing artifacts. In our setting, a structure can be defined by equality among neighboring values of the split parameters at any level. An artifact is formed when actually neighboring split parameters are not equal, but the processing method sets them equal. For our method, consistency rules out this possibility when the photon count is very large. In fact, the probability of an artifact showing up decays exponentially with the sample size.

The other issue, namely detection of a structure, is more involved. Since equal value of neighboring split probabilities can be approximated arbitrarily closely by unequal values, consistency cannot imply that structure will be detected eventually. Each equality among values induces a submodel. After incompatible models are ruled out by consistency, one can compare the compatible models among which the unique model with the lowest dimension describes the structure in the target image. Logarithms of model probabilities in the Bayesian setting can be approximated by the Bayesian Information Criterion (BIC) given in Schwarz (1978). The BIC criterion has two components. One term coming from the log likelihood function at the maximum likelihood

estimate under the model, and the other term equal to the negative of the number of parameters in the model multiplied by  $(1/2) \log n$ , where  $n$  stands for the sample size. For all compatible models, the difference between the first term for the two models is the log-likelihood ratio test statistic for testing nullity when true, and hence converges in distribution to  $(1/2)\chi_d^2$ , where  $d$  is the difference in the dimension of the model under consideration to that of the true one. In particular, the difference is stochastically bounded. The difference of the second term is  $-(d/2) \log n$ . Since under our allocation all models get positive probability, the effect of difference in prior probabilities of the models is asymptotically negligible. Hence, the probability of every compatible redundant model will decay like an integral power of  $1/\sqrt{n}$ . Thus, a structure present in the image will eventually show up in the posterior sampling.

## 4 Empirical Results and Simulations

Although, our method shows theoretical promise for reconstructing Poisson images, empirical evidence is very useful in determining the performance. In this section, the new Bayesian multi-scale model is compared to platelets (Willett and Nowak 2003), the original Bayesian multi-scale model (Kolaczyk 1999), a translation-invariant Haar wavelet transform for Poisson data (Kolaczyk 1997), and the multi-scale complexity regularization (MSCR) introduced in Kolaczyk and Nowak (2004). These four methods will be compared using multiple test images so that the underlying truth will be known after each method attempts to reconstruct the image.

An important distinction between images is the total amount of photons or photon flux that is observed in an image. There are many situations in which an image is limited in its photon count, but in some situations, the image has an abundance of photons. Both of these examples will be analyzed.

The image of Saturn that is used to compare methods was also used in Willett (2006). Three different scenarios are compared: two with images and

tables and a third with only a table. The first is a photon-limited image with a max intensity level of 0.3. The next is the same image with a max intensity level of 10. The third image has a max intensity of 500, as to reveal how these methods compare for an abundance of photons. The intensity image is used to take a Poisson sample that emulates what would be seen in practice if looking at an X-ray image, or any image with Poisson noise.

The following tables compare mean absolute error, mean squared error, and the computing speed. Each of these could determine which methodology should be used in a given situation. All simulations were run using the Matlab mathematical software on a PC with 1GB of RAM and a Pentium 4 processor.

Table 1 and figure 1 represent the photon-limited type of observation. The Bayesian CRP model performs quite well in this scenario. It is a clear improvement over the earlier Bayesian MAP method, and is just behind the multi-scale complexity method. Even though the MSCR method outperforms the Bayesian CRP for these metrics, there are some definite artifacts that occur in parts of the image that are more dramatic than what is seen in the Bayesian CRP method. The computing speed for an image of this size and intensity is longer for the Bayesian CRP method than other methods, with the exception of platelets. All three of the fast methods are using the fast translation-invariant procedure derived from Willett and Nowak (2004). [The code for each of these methods, except the Bayesian CRP, was obtained from Dr. Rebecca Willett.] All of these methods perform quite well considering the observed image.

As the intensity increases, all of the methods obtain better results with respect to the image quality and the metrics of comparison. The difference in metrics starts to become closer and closer between all of the methods. In fact, more noticeable than the difference in metrics in table 2 is the difference in image quality in figure 2. There are only small differences in the processed images when comparing methods, but these small differences can be very noticeable when comparing them to the original image. All methods take the same amount of computation time when increasing the total amount of photons except for

platelets. Its computation time will depend on the photon flux.

Sometimes true images have much higher intensities. For certain supernova remnants, images have been obtained with intensities most likely ranging into the 3000 level. To compare the methods in this scenario in a reasonable time setting, we have used a max intensity of 500. The images are not shown because differences are hard to see with the naked eye. Computing time does not depend on the photon flux for any of the methods except for platelets. When the photon flux becomes relatively large, computation time starts to grow exponentially as seen in table 3.

When implementing all of the methods other than the Bayesian CRP, there is a smoothing parameter that must be chosen. Since the true intensity image is available, a number of different choices are compared to obtain the smallest error with the true image. This would not be possible in a real application. Again, the Bayesian CRP method is completely data driven in that it chooses smoothing parameters based on the observed image only and does not use the true image. Therefore, even though the metrics conclude significant superiority of platelets (and slight superiority of MSCR), the conclusion may differ if one is not allowed to look at the image to choose the smoothing parameters for other methods.

## 5 Astronomical Images

X-ray astronomy is a well-researched field. The methodology described in this paper is particularly well suited in this type of imaging since X-ray images are known to have poisson noise. Each individual pixel in an image represents a count of photons. Each photon can also have a different energy level, but the counts themselves make a poisson representation of the underlying object. With imaging devices like the Chandra X-Ray Observatory, there are many X-rays of all sorts of objects in space. Some are very large and others tend to be small point sources. Regardless of their size, there is always some degree of noise in

these types of images. It is also very common in astronomy to have very limited photon-count images, where the maximum number of photons from every pixel is very small, such as 10. In these situations, many pixels will have counts of 0.

Scientists would like to be able to observe the object exactly as it produces these photons in space. Of one particular interest in X-ray astronomy is observing supernova remnants. A supernova remnant is the energy that remains after a star explodes. There are many different types of supernova remnants, each having different shapes and sizes. Obtaining a good representation of these actual remnants in terms of X-rays allows astronomers to determine what type of remnants they are, as well as other aspects of the remnants that give details to their origin. Two images of supernova remnants were reconstructed with this methodology. They are known as Kepler's Supernova Remnant, a very well known remnant, and G1.9, the youngest supernova remnant discovered thus far.

G1.9 is shown in figure 3 and represents an image with very limited photon counts. The observed image on the left has a max photon count of 10 in a pixel. The smoothed version on the right is the result of applying the Bayesian CRP method to reconstruct the image. This image has a much more defined shape and also minimizes the background noise.

Kepler's supernova remnant is a much larger image with a very large photon flux and is shown in figure 4. The observed image on the left does not look as granular as an observed image for photon-limited astronomy, but when examining the image for minor details, an image reconstruction technique is useful. This image provides another example of images that are commonly seen in X-ray astronomy. Essentially a more popular object in space will get more exposure time, thus giving it a larger photon flux.

There are many other images in X-ray astronomy that could benefit from methodology that is presented here. Many times older methodology that does not produce reasonable results is still used since it is readily available. Typically images are  $1024 \times 1024$  pixels if the full detector is used, at least this is the case

with the Chandra Observatory. Other common uses of Poisson imaging include medical X-rays. These techniques would work for any image that has photon counts in pixels or Poisson noise. The images in this section were obtained from Dr. Stephen Reynolds at North Carolina State University, and can also be found on the Chandra Observatory website (SAO 2009).

## 6 Conclusions

The new Bayesian multi-scale model introduced in this paper using the Chinese Restaurant Process is a very useful competitor that should be considered when smoothing an image with Poisson noise, which include astronomical and medical images. It has been shown through theoretical results and empirical simulations that this methodology has many positive properties. In §2, all of the statistical model is underlined and explained. Key differences to earlier Bayesian methodologies include using a different prior that incorporates combining pixels together by using the Chinese Restaurant Process alongside a Dirichlet prior for the split probabilities after they have been tied together in different groups. This allows for a better smoothing mechanism that ties neighboring pixels smoothing the image. Independent samples are generated from this Bayesian model as opposed to more cumbersome MCMC techniques.

Theoretical consistency properties are given in section §3 and provide another distinct improvement over earlier Bayesian models, namely the model selection consistency. These results are more closely related to statistical models, but do show that this model is consistent in its posterior as well as model selection consistency, which are very appealing in a Bayesian framework.

In empirical simulations, the Bayesian CRP method was compared to four of the leading methods for smoothing images with Poisson noise. Each of these methods have smoothing parameters that must be chosen by the practitioner to adjust the level of smoothing. The Bayesian CRP method outperforms the Bayesian MAP methodology and the Haar wavelets with corrected

Poisson thresholds. It is very close to the multi-scale complexity algorithm and is consistently beaten by the platelets methodology. However, MSCR and platelets choose smoothing parameters knowing the true underlying intensity. The Bayesian CRP method chooses smoothing parameters based on the data, which is more sensible since the true image is not available except in simulations. Further, the computing time for platelets becomes very long with a large photon flux.

The Bayesian CRP methodology we introduced in this paper requires that every pixel a priori has a positive intensity parameter, or that the  $\rho$  values are always positive numbers. This is due to generating the split probabilities from a Dirichlet distribution as well as how a Poisson model is defined. In other words, the current methodology does not allow the true intensity to be exactly zero, which is often a sensible value. Since a zero intensity value may be inconsistent with the data, we will need to redefine the model to allow for background noise. Research is underway to incorporate this possibility of zero intensity in the prior, which is expected to increase the contrast in the images processed by the Bayesian CRP method.

# A Chinese Restaurant Process

The Chinese Restaurant Process (CRP) is a sampling scheme that is very closely related to the Dirichlet process prior used in nonparametric Bayesian statistics. It is a method of randomly assigning objects to groups or a discrete-time process that partitions integers  $(1, \dots, N)$  at time  $N$  into separate groups. A more thorough explanation of the Chinese Restaurant Process can be obtained from Teh et al. (2006) and Pitman (1995). This sampling scheme is described by the following analogy.

Suppose there are an infinite number of tables in a restaurant. These tables can never be too full. Each customer that comes in will sit at a table. These customers are data points. When a customer arrives they will either sit at a new table or sit at another table with other customers creating a cluster. This clustering is very important in the method used in this paper. The more customers that sit at each table the higher the probability a new customer will sit there, however, there is always some probability that a new customer will sit at a new table. This probability is related to the single parameter of the CRP. In this paper, we denote this as  $M$ .

At time  $N + 1$ , the integer has the following probability distribution

$$\begin{aligned} \text{P(joins group } k) &= \frac{n_k}{M + N} \\ \text{P(forms a new group)} &= \frac{M}{M + N} \end{aligned}$$

where  $n_k$  is the number of integers in group  $k$  at time  $N$ . The parameter  $M$  decides how likely a new group is formed. Larger values of  $M$  will make more groups.

The Dirichlet process prior is related to the CRP by the following. Suppose  $X_1, X_2, \dots, X_n | P \sim P$  and  $P \sim \text{DP}(M, G)$ , where  $\text{DP}(M, G)$  stands for the Dirichlet process prior with  $G$  as the center measure and  $M$  as the total mass or precision parameter.

The following properties of a Dirichlet process prior are well known:

$$X_1 \sim G \text{ and } X_2|P, X_1 \sim P,$$

$$P|X_1 \sim \text{DP}(M+1, \frac{M}{M+1}G + \frac{1}{M+1}\delta_{X_1}),$$

$$X_2|X_1 \sim \frac{M}{M+1}G + \frac{1}{M+1}\delta_{X_1}.$$

Thus with probability  $\frac{1}{M+1}$ ,  $X_2$  will replicate  $X_1$ . With probability  $\frac{M}{M+1}$ ,  $X_2$  will be a new draw from distribution  $G$ . Continuing on in a similar manner

$$X_{N+1}|X_1, \dots, X_N \sim \frac{M}{M+N}G + \sum_{i=1}^N \frac{1}{M+N}\delta_{X_i}.$$

The CRP can be thought of as the grouping of the indices of  $X$ . If  $M$  is larger, then the probability of obtaining a new group is higher. Likewise, the more  $X$ 's that are the same, the more likely there will be more of the same.

Most importantly for this paper, the CRP allows neighboring pieces of block-pixels to be equal to each other in the multi-scale representation of images, meaning the number of photons split up between the lower scale from the higher scale can be equivalent. If this splitting is data driven and is also controlled by the value of  $M$ , then different degrees of smoothing can be used in different scales of the image by using different values for  $M$ . This is explained thoroughly in the paper.

## References

- Coifman, R. R. and Donoho, D. L. (1995), “Translation-Invariant De-Noising,” in *Wavelets and Statistics*, eds. Antoniadis, A. and Oppenheim, G., pp. 125–150.
- Donoho, D. L. (1999), “Wedgelets: nearly minimax estimates of edges,” *The Annals of Statistics*, 27, 859–897.
- Esch, D. N., Connors, A., Karovska, M., and van Dyk, D. A. (2004), “An Image Restoration Technique with Error Estimates,” *The Astrophysical Journal*, 610, 1213–1227.
- Kolaczyk, E. D. (1997), “Nonparametric Estimation of Gamma-Ray Burst Intensities Using Haar Wavelets,” *The Astrophysical Journal*, 483, 340–349.
- (1998), “Wavelet Shrinkage Estimation of Certain Poisson Intensity Signals Using Corrected Thresholds,” *Statistica Sinica*, 9, 119–135.
- (1999), “Bayesian Multiscale Models for Poisson Processes,” *Journal of the American Statistical Association*, 94, 920–933.
- Kolaczyk, E. D. and Nowak, R. D. (2004), “Multiscale likelihood analysis and complexity penalized estimation,” *Annals of Statistics*, 32, 500–527.
- Nowak, R. D. and Kolaczyk, E. D. (2000), “A Statistical Multiscale Framework for Poisson Inverse Problems,” *IEEE Transactions On Information Theory*, 46, 1811–1825.
- Pitman, J. (1995), “Exchangeable and partially exchangeable random partitions,” *Probab. Th. Rel. Fields*, 102, 145–158.
- SAO (2009), “Chandra X-RAY Observatory,” <http://chandra.harvard.edu>.
- Schwartz, L. (1965), “On Bayes Procedures,” *Z. Wahrsch. Verw. Gebiete*, 4, 10–26.

- Schwarz, G. (1978), “Estimating the Dimension of a Model,” *Annals of Statistics*, 6, 461–464.
- Teh, Y. W., Jordan, M. I., Beal, M. J., and Blei, D. M. (2006), “Hierarchical Dirichlet Processes,” *Journal of American Statistical Association*, 101, 1566–1581.
- Willett, R. M. (2006), “Multiscale Analysis of Photon-Limited Astronomical Images,” in *Proceedings of Statistical Challenges in Modern Astronomy (SCMA) IV*.
- Willett, R. M. and Nowak, R. (2004), “Fast Multiresolution Photon-Limited Image Reconstruction,” in *Proc. IEEE Int. Sym. Biomedical Imaging — ISBI '04*, 15-18 April, Arlington, VA, USA.
- Willett, R. M. and Nowak, R. D. (2003), “Platelets: A Multiscale Approach for Recovering Edges and Surfaces in Photon-Limited Medical Imaging,” *IEEE Transactions On Medical Imaging*, 22, 332–350.

Table 1: Simulation Results for Saturn Image for max intensity 0.3

<b>Flux=3802 Max 4 photons per pixel</b>			
Method	MAD	MSE	Time
Bayesian CRP (100 iterations)	.179	.0066	1 Min.
Platelets (100 iterations)	.153	.0055	10 Min.
Bayesian Multiscale MAP	.300	.0222	< 1 Sec.
TIPSH	.188	.0074	< 1 Sec.
Multi-scale Complexity	.155	.0062	< 1 Sec.

Table 2: Simulation Results for Saturn Image for max intensity 10

<b>Flux=195973 Max 22 photons per pixel</b>			
Method	MAD	MSE	Time
Bayesian CRP (100 iterations)	.058	.031	1 Min.
Platelets (100 iterations)	.048	.021	111 Min.
Bayesian Multiscale MAP	.120	.128	< 1 Sec.
TIPSH	.066	.040	< 1 Sec.
Multi-scale Complexity	.054	.031	< 1 Sec.

Table 3: Simulation Results for Saturn Image for max intensity 500

<b>Flux=9819764 Max 559 photons per pixel</b>			
Method	MAD	MSE	Time
Bayesian CRP (100 iterations)	.0189	.184	1 Min.
Platelets (100 iterations)	.0149	.109	27.7 Hrs.
Bayesian Multiscale MAP	.0428	.829	< 1 Sec.
TIPSH	.0197	.190	< 1 Sec.
Multi-scale Complexity	.0182	.170	< 1 Sec.

Figure 1: Image of Saturn with max intensity 0.3.

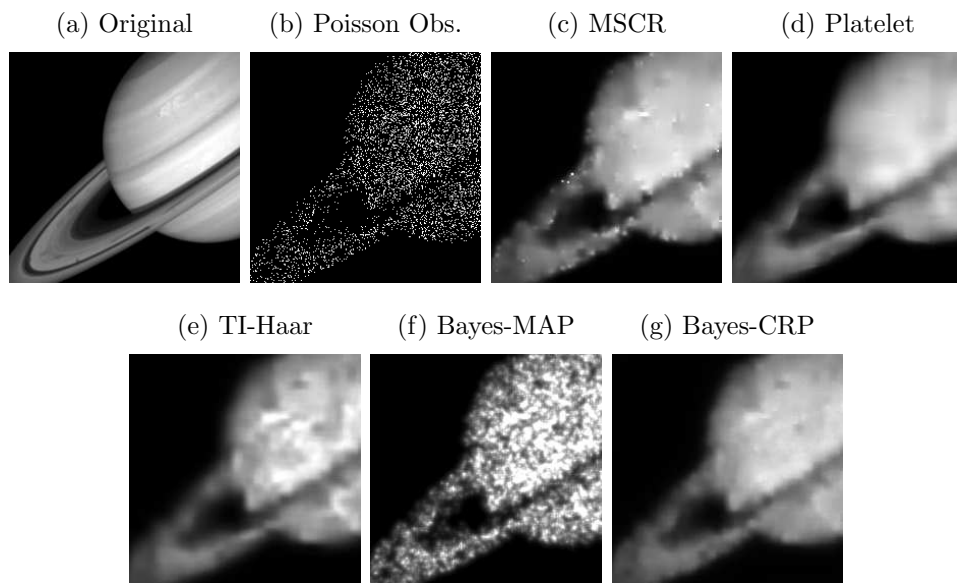


Figure 2: Image of Saturn with max intensity 10.

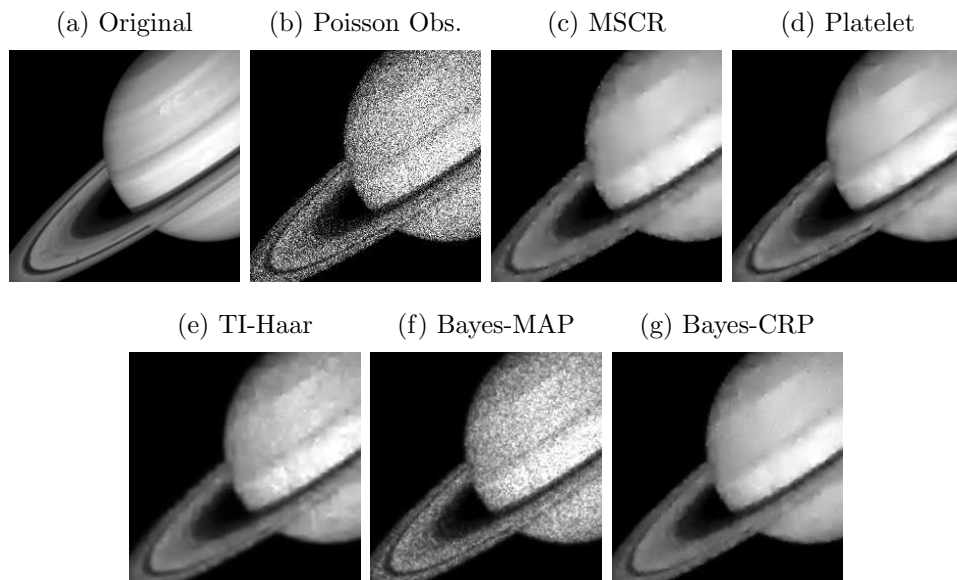


Figure 3: X-ray image of G1.9 from the Chandra X-ray observatory. The original image is on the left and smoothed image is on the right. All of the astronomical images are being viewed with DS9.

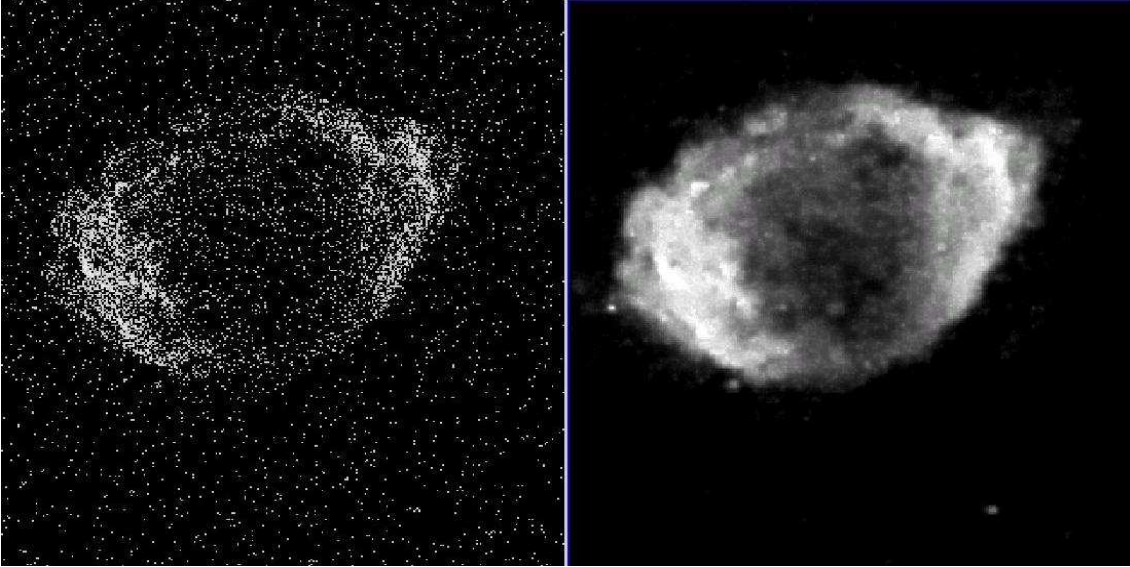


Figure 4: X-ray image of Kepler Supernova Remnant from the Chandra X-ray observatory. The original image is on the left and smoothed image is on the right. All of the astronomical images are being viewed with DS9.

