FULL TITLE
*ASP Conference Series, Vol. **VOLUME**, **YEAR OF PUBLICATION***
NAMES OF EDITORS

Multiscale Analysis of Photon-Limited Astronomical Images

Rebecca Willett

Duke University

Abstract. Many astronomical studies rely upon the accurate reconstruction of spatially distributed phenomena from photon-limited data. These measurements are inherently "noisy" due to low photon counts. In addition, the behavior of the underlying photon intensity functions can be very rich and complex, and consequently difficult to model *a priori*. Nonparametric multiscale reconstruction methods overcome these challenges and facilitate characterization of fundamental performance limits. In this paper, we review several multiscale approaches to photon-limited image reconstruction, including wavelets combined with variance stabilizing transforms, corrected Haar wavelet transforms, multiplicative multiscale innovations, platelets, and the *à trous* wavelet transform. We discuss the performance of these methods in simulation studies, and detail statistical analyses of their performances.

1. Photon-limited astronomy

Many imaging modalities involve the detection of light or higher energy photons, and often the random nature of photon emission and detection is the dominant source of noise in imaging systems. Such cases are referred to as *photon-limited* imaging applications, since the relatively small number of detected photons is the factor limiting the signal-to-noise ratio. In many cases, the intensity underlying the photon-limited observations may be distorted by the point spread function or other physical properties of the imaging system. Using these inherently noisy and distorted observations to perform quantitative inference on the underlying astrophysical phenomenon is a challenging problem affecting many researchers in the statistical and astronomical communities.

The data collected by these imaging systems are usually assumed to obey a spatial Poisson distribution involving a two-dimensional intensity image that describes the probability of photon emissions at different locations in space. The mean and variance of a Poisson process are equal to the intensity. The intensity/mean is the "signal" of interest and the variability of the data about the mean can be interpreted as "noise." Thus, as the intensity varies spatially as a function of astrophysical structure and function, so does the signal-to-noise ratio. In this sense it could be said that the noise in photon-limited imaging is signal-dependent.

Many investigators have considered the use of wavelet representations for image denoising, deblurring, and other forms of image reconstruction because of the theoretical near-optimality and practical efficacy of wavelets in a variety of image processing contexts; for examples, see (Mallat 1998; Starck et al. 1998; Aldroubi and Unser 1996). The procedure for classic wavelet denoising via hard thresholding is the following: compute the wavelet transform of the noisy image, set wavelet coefficients with magnitude less than some threshold to zero, and compute the inverse wavelet transform. Wavelet denoising via soft thresholding is very similar, except each coefficient is either set to zero or shrunk depending upon its magnitude. The basic idea behind these methods is that wavelet bases form a very parsimonious representation of many images of interest, so that the bulk of the image's energy in concentrated in just a few wavelet coefficients, which as a result have very high magnitudes. Most noise, however, does not share this property, and has its energy distributed relatively evenly among all the wavelet coefficients. Thus, by

thresholding or shrinking wavelet coefficients, we remove a significant component of the noise without a large loss in image detail. Initial robust statistical analyses of this procedure were based on the assumption of additive white Gaussian noise. However, in the context of photon-limited imaging most wavelet-based approaches are based on Gaussian or other simplifying approximations to the Poisson likelihood. This is due to the fact that it is very difficult in general to apply wavelets (as well as more recent innovations such as complex wavelets (Kingsbury 1999) and curvelets (Candès and Donoho 1999)) to Poisson data due to the spatially varying noise variance.

Ignoring the effects of non-Gaussian noise can be a significant source of error, but more precise noise models can sometimes lead to highly non-linear computational problems. Gaussian approximations are usually only reasonable if the numbers of detected photons are sufficiently large (so that the Poisson data, possibly after a suitable transformation, is roughly Gaussian distributed). Binning or aggregating observations over pixels of sufficiently large area to increase photon count levels forces an immediate sacrifice of spatial resolution. In addition, taking advantage of the wealth of theoretical, algorithmic, and experimental expertise developed in photon-limited imaging in the past two decades, we observe that methods which retain the Poisson likelihood criterion as the fundamental tool for statistical inquiry are quite advantageous.

A number of researchers have dedicated considerable time and energy to the development of photon-limited image reconstruction methods, only a few of which are highlighted in this paper. This review of multiscale methods for photon-limited image estimation is by no means comprehensive. Instead, it is meant to emphasize the similarities of and differences between some seminal and recently proposed techniques. A variety of additional techniques are described with examples by Starck and Murtagh (2002) and Besbeas et al. (2004), including multiscale methods for astronomical image compression, detection, and estimation of multichannel data. A significant proportion of these works focus on studying the statistical distribution of wavelet coefficients of photon-limited images and using these distributions to develop improved wavelet shrinkage techniques (Besbeas et al. 2004; Bijaoui and Jammal 2001; Beran and Dümbgen 1998; Antoniadis and Sapatinas 2001). Sardy et al. (2004) attack the problem in a complexity regularization framework by penalizing the Poisson likelihood with the ℓ_1 norm of the wavelet coefficients.

Rather than attempting to describe every multiscale Poisson image analysis method in the literature, we focus on four primary classes of approaches and detail a few examples in each class. These five classes are (1) variance stabilizing transforms, (2) Poisson-specific Haar wavelet thresholds, (3) Multiplicative Multiscale Innovations (MMI) models, (4) platelets, and (5) the *à trous* wavelet transform. In particular, this paper is structured as follows: In Section 2., we describe the mathematical formulation of the photon-limited image estimation problem, and provide an empirical example of the effect of using traditional wavelet-based methods in this context. We then explore the usage of variance stabilizing transforms such as the Anscombe transform in Section 3.. Next, in Section 4. we begin our discussion of multiscale methods developed explicitly for photon-limited data analysis, starting with the corrected Haar wavelet thresholds developed by Kolaczyk (1999a). This is followed in Section 5. by a discussion of Multiscale Multiplicative Innovation (MMI) models and their extensions (Timmermann and Nowak 1999; Esch et al. 2004; Kolaczyk and Nowak 2004). In Section 6., we examine platelets (Willett and Nowak 2003) which, unlike any of the other methods discussed in this paper, are designed explicitly to handle the two-dimensional structures often found in images. Finally, in Section 7., we study the application of the *à trous* wavelet transform to photon-limited images, which results in smoother representations of extended image structures and performs very well in practice.

It is important to note that many of the estimation (i.e. denoising) methods described in this paper can easily be incorporated into a deblurring or other image reconstruction context, as described

by Nowak and Kolaczyk (2000). They demonstrate that Poisson inverse problems can be solved by an Expectation-Maximization (EM) method in which the E-step consists of a standard Richardson-Lucy (Richardson 1972; Lucy 1974) or Shepp-Vardi (Shepp and Vardi 1982) iteration, and the M-step consists of Poisson intensity estimation. Algorithmically this procedure can be performed for any Poisson intensity estimation method, though for the resulting procedure to be a formal EM method, the Poisson intensity estimation must be formulated in a likelihood context. Thus, in order to best contrast the methods under study in terms of photon flux preservation, error performance, and artifacts, we display empirical results for direct estimation with the understanding that extensions to inverse problems are in several cases immediate.

2. Problem Formulation

In this paper we consider the problem of reconstructing an intensity image $\boldsymbol{\lambda}$, where $\boldsymbol{\lambda} = \{\lambda_i\}_{i=0}^{N-1}$ and λ_i is the intensity at the i^{th} pixel, and there are a total of N pixels. The reconstruction is performed using the data $\boldsymbol{x} = \{x_i\}_{i=0}^{N-1}$, where x_i is the number of photons detected by the i^{th} detector element. Assume that the data are independent observations of a Poisson random variable with intensity $\boldsymbol{\lambda}$; i.e.

$$x_i \sim \textit{Poisson}(\lambda_i).$$

The *photon flux*, $\sum_i x_i$, is denoted as n.

To see the challenge of using traditional wavelet-based reconstruction methods for this problem when n is very small, consider the following experiment, which is revisited throughout this paper. An original intensity image, $\boldsymbol{\lambda}$, with intensities ranging from zero to 0.2, is displayed in Figure 1(a); the Poisson observations of this intensity image, $\boldsymbol{x}$, are displayed in Figure 1(b). Note that most of the observations consist of either zero or one photon counts, with a mean of 0.06 photons per pixel and a total photon flux of $n = 3,961$.

The wavelet denoising via hard thresholding method described above can be used to form an estimate of the underlying intensity in this example. For this (and all future) experiments, we will also employ cycle spinning (Coifman and Donoho 1995) to generate visually appealing, translation invariant results with low errors. In general, cycle spinning (or averaging over shifts) can yield significant improvements in the quality of intensity estimates. The process entails circularly shifting the raw data, denoising, and then shifting the estimate back to its original position, and repeating this procedure for each possible shift. This procedure can be performed very computationally efficiently by using undecimated wavelets. Cycle spinning is especially important in the context of Haar-based estimates; while the Haar system is very convenient for photon-limited imaging, as detailed below, Haar wavelets lack the smoothness of other wavelets. This effect is mitigated with cycle spinning.

Figures 1(c-d) display the result of wavelet estimation via hard thresholding for undecimated (translation invariant) Haar and undecimated Daubechies D6 wavelets, respectively. The threshold level was clairvoyantly set for each image to minimize the ℓ_1 error between $\boldsymbol{\lambda}$ and the estimate $\widehat{\boldsymbol{\lambda}}$, $\sum_i |\lambda_i - \widehat{\lambda}_i|$; setting the threshold in this manner for real problems would not be possible, and so this experiment can be regarded as a "best case" scenario. In both cases, most features are significantly distorted. In addition, it is clear that artifact levels are intensity dependent, which results directly from the fact that the variance of a Poisson random variable is equal to its intensity. One helpful feature of this method, however, is that photon flux is preserved; ensuring that the photon flux of the estimate equals the total number of observed photons is a critical component of estimation methods used in many astronomical contexts.

This experiment is repeated for the XMM simulation image used to study photon-limited imaging in (Starck and Murtagh 2006). The original intensity underlying the data is unavailable, but consists of sources of multiple sizes and intensities and hence is a useful benchmark in many astronomical applications. The observations are displayed on a log scale in Figure 2(a), and the wavelet reconstructions

are displayed in Figure 2(b-c). Zoomed-in versions of the results are in Figure 3. While wavelet-based techniques appear effective for bright sources, they are unable to faithfully reconstruct fainter sources. A number of artifacts are also visible.

3. Variance stabilizing transformations

The Poisson statistics of the observations can be altered by using a *variance stabilizing transformation*, such as the Anscombe transform (Anscombe 1948) or the Fisz transform (Fisz 1955; Fryźlewicz and Nason 2001). In this review we focus on the Anscombe transform, though the Fisz transform approaches share many similar properties. In particular, if X is a Poisson random variable with intensity λ, let the Anscombe transform of X, denoted Y, be defined as $Y \equiv \mathcal{A}(X) = 2\sqrt{X + \frac{3}{8}}$. It can be shown that

$$Y \xrightarrow{\mathcal{D}} \mathcal{N}(2\sqrt{\lambda}, 1),$$

meaning Y converges (as λ increases) in distribution to a normal random variable with mean $2\sqrt{\lambda}$ and unit variance. Note that the variance of X varies spatially with the intensity λ, while the variance of Y is constant. The additive white Gaussian noise assumption following the Anscombe transform is accurate when each photon count x_i is approximately thirty or more.

Based on this, several researchers have proposed using the Anscombe transformation in the following manner (e.g. see Donoho 1993; Starck and Murtagh 2002):

1. Set $y_i = \mathcal{A}(x_i)$ for each pixel location $i = 1, \ldots, N$, where N is the number of pixels in the image.

2. Estimate $\mu_i = \mathbb{E}[y_i]$ assuming an additive white Gaussian noise model and using an image estimation procedure such as wavelet hard or soft thresholding (Donoho 1992; Donoho and Johnstone 1998); denote this estimate $\widehat{\boldsymbol{\mu}}$.

3. Set $\widehat{\lambda}_i = \mathcal{A}^{-1}(\mu_i) \equiv \left(\frac{\widehat{\mu}_i}{2}\right)^2 - \frac{3}{8}$, where $\mathcal{A}^{-1}$ denotes the inverse Anscombe transform.

In many cases, wavelet estimation is a popular signal and image processing tool because it is theoretically near-minimax optimal and practically highly effective. Specifically, Donoho (1992) and Donoho and Johnstone (1998) show that is it possible to choose a hard or soft wavelet threshold such that

$$\mathbb{E}\left[\|\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}\|_2^2\right] \preceq \left(\frac{\log N}{N}\right)^{2\alpha/(2\alpha+d)}$$

when μ is in $\mathcal{B}^\alpha$, a Besov space with smoothness parameter $\alpha > 0$, and d is the dimension of the data. (For sequences a_N and b_N let the notation $a_N \preceq b_N$ imply there exists some $C > 0$ such that $a_N \leq C b_N$ for all N. The notation $a_N \succeq b_N$ is defined similarly.) The Besov space is a large family of images including smooth images, such as polynomials of degree at most α (DeVore 1998). When $\alpha = 1$, this performance bound also holds for *piecewise* smooth images with edges and boundaries. Multiresolution analysis in general has a key advantage over many competing methods precisely because it performs well even when the images are not smooth everywhere, but have jumps or sharp changes.

This bound tells us how quickly the error decays as we increase the number of observations. The rate quoted above is within a logarithmic factor of the fastest possible rate achievable by any estimator for functions in this Besov space. That is,

$$\inf_{\widetilde{\boldsymbol{\mu}}_N} \sup_{\boldsymbol{\mu} \in \mathcal{B}^\alpha} \mathbb{E}[\|\widetilde{\boldsymbol{\mu}}_N - \boldsymbol{\mu}\|_2^2] \succeq N^{-2\alpha/(2\alpha+d)}$$

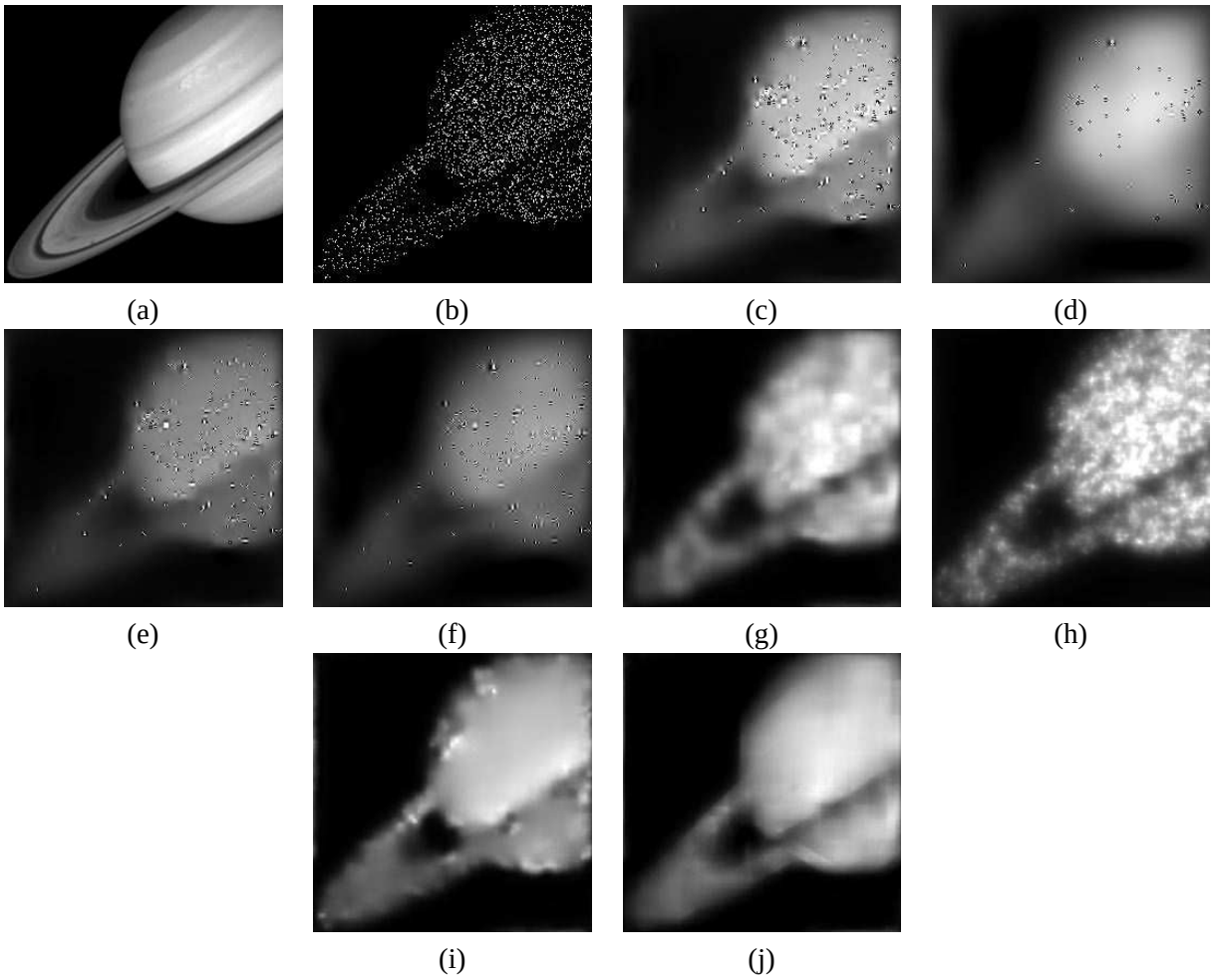

Figure 1. Multiscale Poisson intensity estimation. (a) Original intensity image, with intensities in the range $[0, 0.2]$. Mean intensity is 0.06. (b) Poisson observations. Maximum number of photons/pixel is seven. (c) Result of applying wavelet estimation to the observations, using undecimated Haar wavelets and hard thresholding. $\frac{1}{n}\|\boldsymbol{\lambda} - \widehat{\boldsymbol{\lambda}}\|_1 = 0.325$. (d) Result of applying wavelet estimation to the observations, using undecimated Daubechies D6 wavelets and hard thresholding. $\frac{1}{n}\|\boldsymbol{\lambda} - \widehat{\boldsymbol{\lambda}}\|_1 = 0.360$. (e) Result of applying wavelet estimation to the Anscombe transform of the observations and then computing the inverse Anscombe transform, using undecimated Haar wavelets and hard thresholding. $\frac{1}{n}\|\boldsymbol{\lambda} - \widehat{\boldsymbol{\lambda}}\|_1 = 0.465$. (f) Result of applying wavelet estimation to the Anscombe transform of the observations and then computing the inverse Anscombe transform, using undecimated Daubechies D6 wavelets and hard thresholding. $\frac{1}{n}\|\boldsymbol{\lambda} - \widehat{\boldsymbol{\lambda}}\|_1 = 0.462$. (g) Result of estimating the intensity using corrected wavelet thresholds, using undecimated Haar wavelets and hard thresholding. $\frac{1}{n}\|\boldsymbol{\lambda} - \widehat{\boldsymbol{\lambda}}\|_1 = 0.198$. (h) Result of applying MAP estimation with the MMI model and cycle spinning. $\frac{1}{n}\|\boldsymbol{\lambda} - \widehat{\boldsymbol{\lambda}}\|_1 = 0.245$. (i) Result of applying complexity regularization with the MMI model and cycle spinning. $\frac{1}{n}\|\boldsymbol{\lambda} - \widehat{\boldsymbol{\lambda}}\|_1 = 0.173$. (j) Result of estimating the intensity using platelet estimation, cycle spun over 100 random shifts. $\frac{1}{n}\|\boldsymbol{\lambda} - \widehat{\boldsymbol{\lambda}}\|_1 = 0.163$. For each estimate, any parameters and/or thresholds were set to minimize the ℓ_1 error between the estimate and the true intensity image.

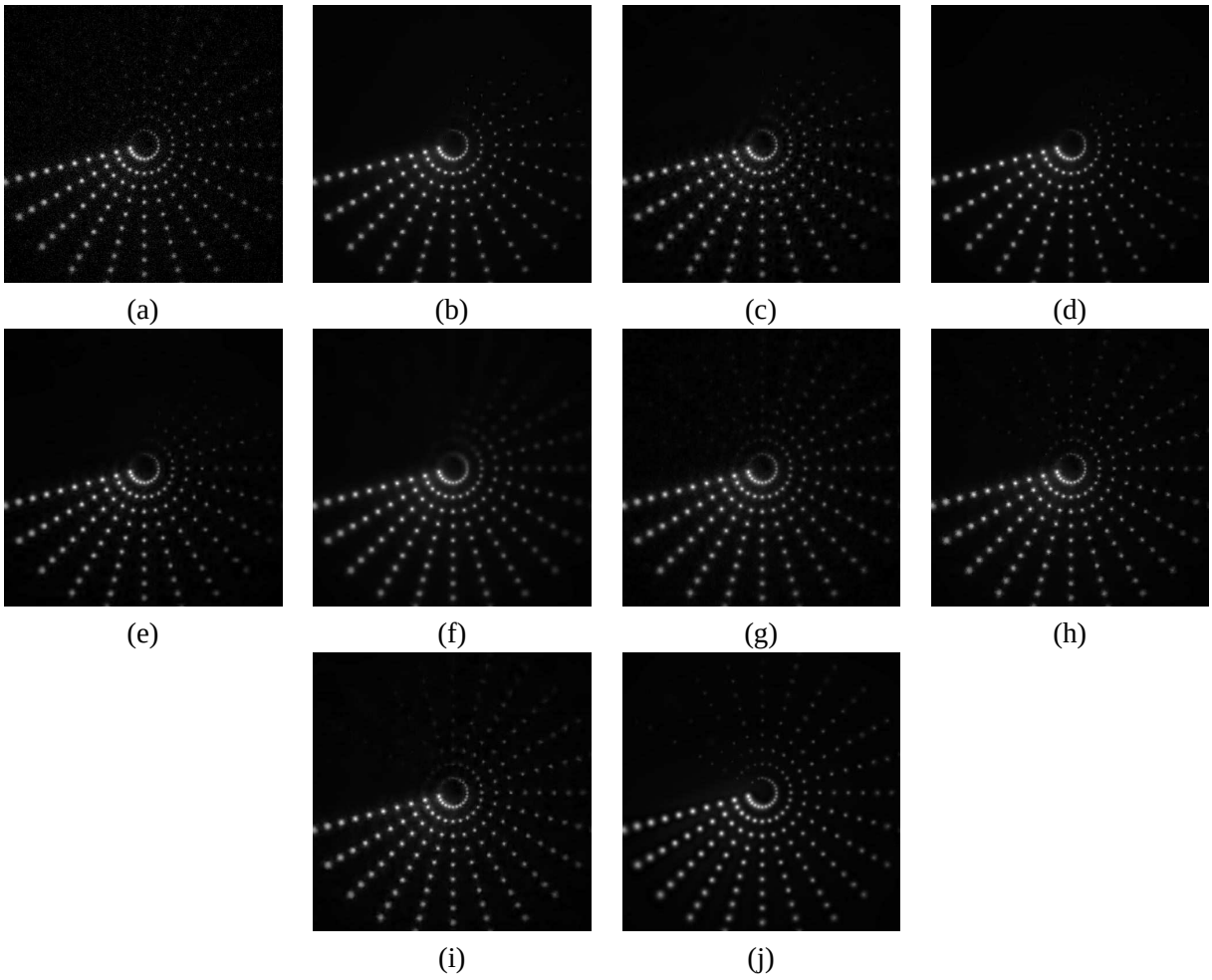

Figure 2. Multiscale Poisson intensity Estimation. (a) Poisson observations. Mean number of photons/pixel is 0.38, and maximum number of photons per pixel is 186. (b) Result of applying wavelet estimation to the observations, using undecimated Haar wavelets and hard thresholding. (c) Result of applying wavelet estimation to the observations, using undecimated Daubechies D6 wavelets and hard thresholding. (d) Result of applying wavelet estimation to the Anscombe transform of the observations and then computing the inverse Anscombe transform, using undecimated Haar wavelets and hard thresholding. (e) Result of applying wavelet estimation to the Anscombe transform of the observations and then computing the inverse Anscombe transform, using undecimated Daubechies D6 wavelets and hard thresholding. (f) Result of estimating the intensity using corrected wavelet thresholds, using undecimated Haar wavelets and hard thresholding. (g) Result of applying MAP estimation with the MMI model and cycle spinning. (h) Result of applying complexity regularization with the MMI model and cycle spinning. (i) Result of estimating the intensity using platelet estimation, cycle spun over 100 random shifts. (j) Result of estimating the intensity using the *à trous* wavelet transform and variance stabilized coefficient thresholding. For each estimate, any parameters and/or thresholds were set to yield high visual quality.

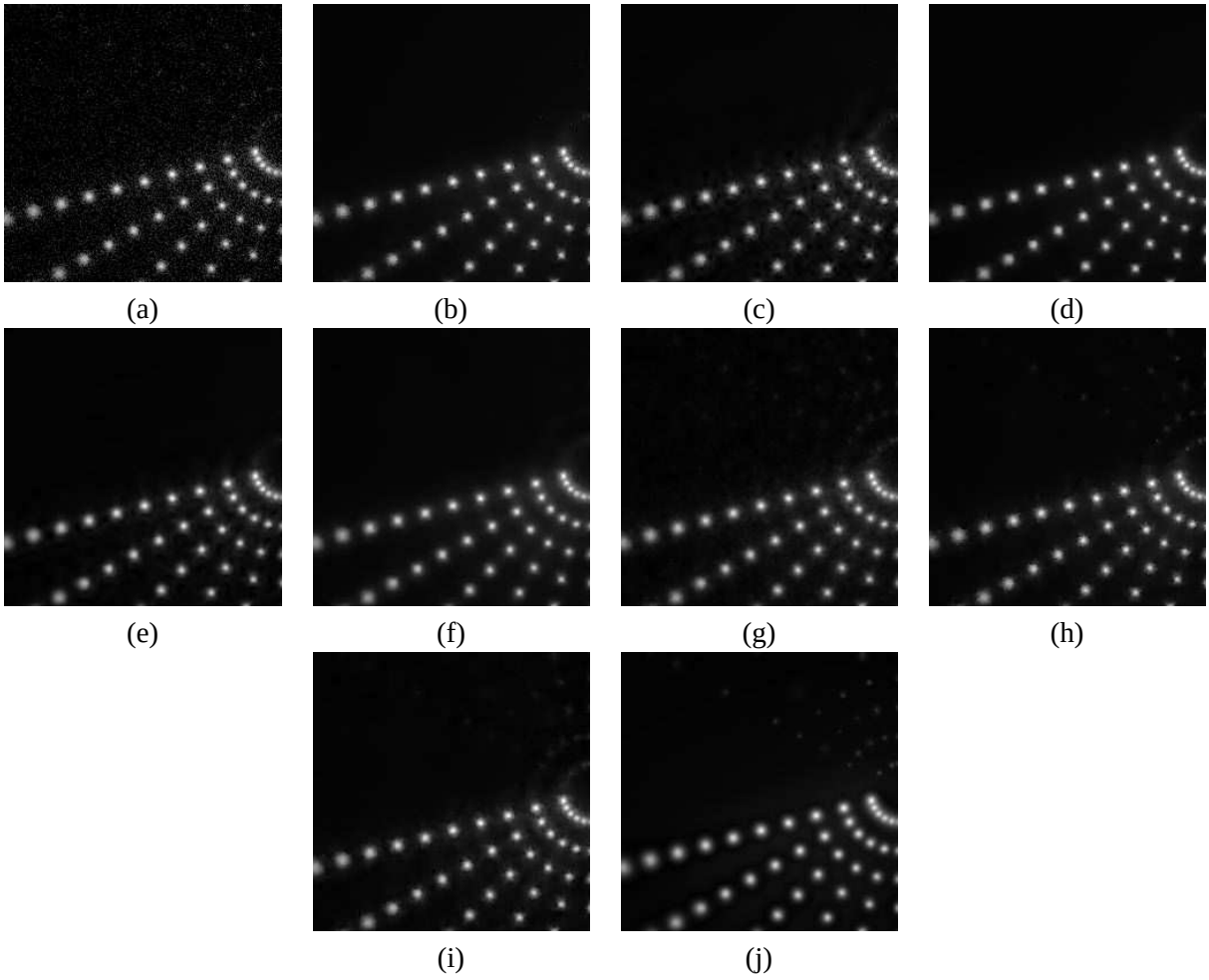

Figure 3. Zoomed-in images from Figure 2. Note that methods designed using the Poisson likelihood reconstruct faint sources much more effectively than more traditional approaches. (a) Poisson observations. Mean number of photons/pixel is 0.38, and maximum number of photons per pixel is 186. (b) Result of applying wavelet estimation to the observations, using undecimated Haar wavelets and hard thresholding. (c) Result of applying wavelet estimation to the observations, using undecimated Daubechies D6 wavelets and hard thresholding. (d) Result of applying wavelet estimation to the Anscombe transform of the observations and then computing the inverse Anscombe transform, using undecimated Haar wavelets and hard thresholding. (e) Result of applying wavelet estimation to the Anscombe transform of the observations and then computing the inverse Anscombe transform, using undecimated Daubechies D6 wavelets and hard thresholding. (f) Result of estimating the intensity using corrected wavelet thresholds, using undecimated Haar wavelets and hard thresholding. (g) Result of applying MAP estimation with the MMI model and cycle spinning. (h) Result of applying complexity regularization with the MMI model and cycle spinning. (i) Result of estimating the intensity using platelet estimation, cycle spun over 100 random shifts. (j) Result of estimating the intensity using the *à trous* wavelet transform and variance stabilized coefficient thresholding. For each estimate, any parameters and/or thresholds were set to yield high visual quality.

These bounds provide astronomers with a mechanism for estimating the accuracy of the estimate of $\boldsymbol{\mu}$. However, as noted by Kolaczyk and Dixon (2000), since the Anscombe transform and its inverse are very nonlinear operations, it is difficult to directly translate these bounds into bounds on the accuracy of the estimate of $\boldsymbol{\lambda}$. Performance bounds for methods based on variance stabilizing transforms are currently in development by Brown et al. (2006).

Using the Anscombe transform to estimate intensities does not always lead to photon flux preservation. In particular, if the number of photon observed is n, then a photon flux preserving estimate would have the property that $\sum_i \widehat{\lambda}_i = n$. Wavelet denoising via coefficient thresholding or shrinking, without the use of a variance stabilizing transform, would ensure this property, but combined with the nonlinear Anscombe transform, the photon flux of the estimate may be very different from the photon flux of the observations. Another difficulty associated with the Anscombe transform occurs when photon counts are very small. In the extreme case, imagine a high-resolution photon-limited system in which every observation is either zero or one. Clearly, the Anscombe transform of this image remains binary valued and has very non-Gaussian statistics. This means that traditional wavelet estimation methods, which were developed under Gaussian noise assumptions, behaves suboptimally and often produces undesirable artifacts.

The results of applying wavelet estimation in conjunction with the Anscombe transform are displayed in Figure 1(e-f) and Figures 2 and 3 (d-e). Again, these images correspond to the undecimated Haar and undecimated D6 wavelets, respectively. The photon flux of the estimate is significantly smaller than the photon flux of the source. In addition, as pointed out by Kolaczyk (1999a), this yields significant over-smoothing of the image.

4. Corrected Haar wavelet thresholds

The fact that the variance of the noise is proportional to the intensity in photon-limited systems is a major stumbling-block for most wavelet-based image reconstruction methods, resulting in the intensity dependent artifacts we observed in the previous section. This observation suggests that, rather than using a single threshold for wavelets at all scales, we should threshold wavelet coefficients differently at different scales. Noting this, Kolaczyk (1999a) and Kolaczyk and Dixon (2000) developed a method for calibrating wavelet threshold levels in a scale-dependent manner using the Poisson statistics of the observations.

This approach is derived from a hypothesis testing framework, where a Haar wavelet coefficient should be thresholded when it is consistent with the known background (the null hypothesis) and not thresholded when it is inconsistent with the background. This consistency is determined by evaluating the probability that the coefficient exceeds some threshold under the null hypothesis.

For the purpose of this paper, assume the background is a constant denoted λ_0. Next, consider the statistics of the wavelet coefficients of an additive white Gaussian noise contaminated background (i.e. the coefficient statistics under the null hypothesis). If the noise has variance σ^2, then the wavelet coefficients $\{w_{j,k}\}_{j,k}$ of the observations are independent and identically distributed (iid) Gaussian random variables with zero mean and variance σ^2. Furthermore it can be shown that

$$\mathbb{P}\left[\max w_{j,k} \leq \sigma\sqrt{2\log N}\right] \longrightarrow 1 \quad \text{as } N \longrightarrow \infty \tag{1}$$

where $w_{j,k}$ is the scale-j wavelet coefficient at the k^{th} shift. This means that large wavelet coefficients are highly unlikely to correspond to noisy observations of a constant background. This suggests thresholding coefficients (at all scales) with magnitudes less than $\sigma\sqrt{2\log N}$. Contrast this with the distribution of the wavelet coefficients of Poisson observations with intensity λ_0. These coefficients, denoted $d_{j,k}$, are also zero-mean and homoscedastic (i.e. have the same variance), but they are distinctly non-Gaussian.

Kolaczyk (1999a) designed scale dependent thresholds t_j such that

$$\mathbb{P}\left[\max_k d_{j,k} \leq t_j\right] \longrightarrow 1 \quad \text{as } N \longrightarrow \infty$$

at a rate similar to that in (1). Kolaczyk and Dixon (2000) argue for setting the threshold at level j to

$$t_j = \frac{2^j}{\sqrt{N}}\left[\log N_j + \sqrt{\log^2 N_j + 2\lambda_j \log N_j}\right]$$

where $N_j = 4^j$ is the number of image wavelet coefficients at scale j and $\lambda_j = N/2^j\lambda_0$ is the background rate per coefficient at scale j. This threshold is selected to limit the false detection rate (FDR), which is related to the probability of determining a wavelet coefficient is significant when in fact it is consistent with the background. Despite these helpful statistical properties, it is not immediately clear how the corrected thresholds impact the rates of convergence discussed in the previous section; to date, this type of theoretical property has not been directly established for this method.

The results of applying Poisson threshold correction to our test data set, see Figure 1(g) and Figures 2 and 3(f). For this simulation, we set

$$t_j = C\frac{2^j}{\sqrt{N}}\left[\log N_j + \sqrt{\log^2 N_j + 2 \cdot 2^{(J-j)}\lambda_0 \log N_j}\right]$$

and searched over multiple values of C and λ_0 to find the pair which minimized the ℓ_1 error between λ and $\widehat{\lambda}$; for the XMM simulation image, C was chosen to yield a visually appealing result. Clearly the corrected threshold result outperforms techniques we considered in the previous section based on non-corrected thresholds and variance stabilizing transforms. Furthermore, because this is a wavelet shrinkage technique, it is naturally photon flux preserving.

5. Multiplicative Multiscale Innovations

As an alternative to the Kolaczyk (1999a) work on corrected thresholds, Timmermann and Nowak (1999); Kolaczyk (1999b) independently developed a *Multiplicative Multiscale Innovations* (MMI) model for photon-limited image intensity estimation, which are related to multiplicative cascades (Kolmogorov 1941) and multifractals (Mandelbrot 1982). This approach was further developed into the fully Bayesian EMC2 method by Esch et al. (2004). Later, Nowak and Kolaczyk (Nowak and Kolaczyk 2000; Kolaczyk and Nowak 2004) expanded upon this framework together in the context of inverse problems and complexity regularization. This section details these techniques.

5.1. The MMI model

The basic idea of the MMI model is to perform estimation by examining statistics of unnormalized Haar scaling coefficients rather than normalized wavelet coefficients. Let $J = \log_2 N$ and set

$$\begin{aligned} x_{J,k,l} &\equiv x_{k,l}, \quad k,l = 0,\ldots,2^J-1 \\ x_{j,k,l} &= x_{j+1,2k,2l} + x_{j+1,2k+1,2l} + x_{j+1,2k,2l+1} + x_{j+1,2k+1,2l+1}, \\ &\qquad k,l = 0,\ldots,2^j-1,\ 0 \leq j \leq J-1. \end{aligned}$$

As before, j refers to the resolution, with $j = J$ being the finest scale. Here, k and l refer to two-dimensional locations (as opposed to the index i used earlier in this paper). The $\{x_{j,k,l}\}$ are the unnormalized Haar scaling coefficients of $\boldsymbol{x}$. Since each $x_{j,k,l}$ is the unweighted sum of independent Poisson

random variables, it also obeys a Poisson distribution. Specifically, define a similar multiscale analysis of the intensity $\boldsymbol{\lambda}$:

$$\begin{aligned}\lambda_{J,k,l} &\equiv \lambda_{k,l}, \quad k,l = 0,\ldots,2^J-1 \\ \lambda_{j,k,l} &= \lambda_{j+1,2k,2l} + \lambda_{j+1,2k+1,2l} + \lambda_{j+1,2k,2l+1} + \lambda_{j+1,2k+1,2l+1}, \\ &\qquad k = 0,\ldots,2^j-1,\ 0 \le j \le J-1.\end{aligned}$$

Then $x_{j,k,l}$ follows a Poisson distribution with intensity parameter $\lambda_{j,k,l}$. This simple statistical characterization facilitates an analysis which is far more tractable than analyzing typical normalized wavelet coefficients, which have complicated distributions in the presence of Poisson noise. Now let

$$\rho^{(1)}_{j,k,l} \equiv \frac{\lambda_{j+1,2k,2l}}{\lambda_{j,k}}, \qquad \rho^{(2)}_{j,k,l} \equiv \frac{\lambda_{j+1,2k+1,2l}}{\lambda_{j,k}},$$
$$\rho^{(3)}_{j,k,l} \equiv \frac{\lambda_{j+1,2k,2l+2}}{\lambda_{j,k}}, \quad \text{and} \quad \rho^{(4)}_{j,k,l} \equiv 1 - \rho^{(1)} - \rho^{(2)} - \rho^{(3)};$$

the collection of these parameters $\{\rho^{(i)}_{j,k,l}\}$, which can be regarded as "splitting factors", are the multiplicative multiscale innovations.

For clarification, refer to Figure 4. In this graph, $x_{j,k,l}$ denotes the observed $(j,k,l)^{th}$ unnormalized Haar scaling coefficient, $\lambda_{j,k,l}$ its expected value, and $\rho_{j,k,l}$ the associated multiplicative multiscale innovations. The innovations are displayed on the edges of the graph because of the relationship described above.

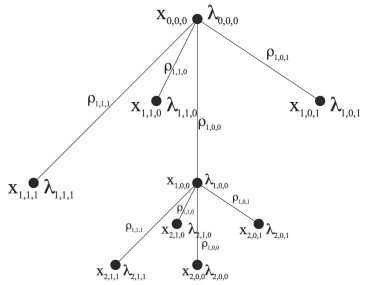

Figure 4. Multiplicative multiscale innovations displayed on a quadtree. In this figure, $x_{j,k,l}$ denotes the observed $(j,k,l)^{th}$ unnormalized Haar scaling coefficient, $\lambda_{j,k,l}$ its expected value, and $\rho_{j,k,l}$ the associated multiplicative multiscale innovations. The innovations are displayed on the edges of the graph because of their multiplicative nature.

This framework can be used for intensity estimation in several ways. In the following subsections, we discuss maximum *a posteriori* (MAP), fully Bayesian, and complexity regularization techniques.

5.2. MAP estimation

One way to estimate the unnormalized Haar scaling coefficients of the intensity, $\{\lambda_{j,k,l}\}$, from the observed coefficients $\{x_{j,k,l}\}$, is to place a Bayesian prior distribution on the MMI parameters $\{\rho_{j,k,l}\}$ and the total intensity parameter $\lambda_{0,0,0}$. Both Timmermann and Nowak (1999) and Nowak and Kolaczyk (2000) use a Gamma prior for $\lambda_{0,0,0}$ and a Dirichlet (or Beta in one dimension) prior distribution for the MMI parameters because they have the proper support and are conjugate priors (facilitating closed-form estimates). In Nowak and Kolaczyk (2000), the final estimates, based on the Poisson statistics of the

observations and the prior probability models, are written as follows:

$$\widehat{\lambda}_{0,0,0} = \frac{\gamma + x_{0,0,0} - 1}{\delta + 1} \qquad \widehat{\rho}^{(1)}_{j,k,l} \equiv \frac{x_{j+1,2k,2l} + \alpha_j - 1}{x_{j,k} + 4(\alpha_j - 1)},$$
$$\widehat{\rho}^{(2)}_{j,k,l} \equiv \frac{x_{j+1,2k+1,2l} + \alpha_j - 1}{x_{j,k} + 4(\alpha_j - 1)}, \qquad \widehat{\rho}^{(3)}_{j,k,l} \equiv \frac{x_{j+1,2k,2l+2} + \alpha_j - 1}{x_{j,k} + 4(\alpha_j - 1)}, \text{ and}$$
$$\widehat{\rho}^{(4)}_{j,k,l} \equiv 1 - \widehat{\rho}^{(1)} - \widehat{\rho}^{(2)} - \widehat{\rho}^{(3)}; \qquad 0 \le k \le 2^j - 1,\ 0 \le j \le J - 1$$

Here, γ and δ are parameters of the Gamma prior distribution, and $\{\alpha_j\}$ are the parameters of the Dirichlet prior distribution at each scale j. (The Dirichlet prior is restricted to be symmetric, and hence only has one parameter.) In Nowak and Kolaczyk (2000), these parameters are assumed to be known, or are regarded as tuning parameters to be set by a user. A more Bayesian perspective is possible, as described in Section 5.3.

5.3. EMC2

Esch et al. (2004) extended the work of Nowak and Kolaczyk (2000) by further developing the Bayesian framework described in Section 5.2. In particular, above we noted that $\{\alpha_j\}_j$ were parameters which could be used to impact the MAP estimate. Esch et al. (2004), rather than estimating these *hyperparameters* directly, placed additional prior distributions (*hyperpriors*) on them. For example, a Gamma or a similar distribution may be used for the prior distribution on each hyperparameter α_j. The estimate of the image is then set to be equal to the expected value of the posterior distribution for each pixel, which is computed using Markov-Chain Monte Carlo (MCMC) methods; hence, the technique is called EMC2 for Estimation with MCMC.

This technique exhibits two particularly useful features. First, the method gives the user a means for formally incorporating prior information about the intensity into the estimation procedure. For example, when observing a source in front of a known or somewhat known background, the prior information can be used to help shape the hyperpriors and potentially improve the estimation of the source. It can also be used to incorporate information about the underlying physics, such as the "burstiness" of the source. Second, the variance of the posterior distribution can be computed for each pixel along with the expected value; this per-pixel variance is to some extent a measure of the accuracy of the estimate of that pixel's intensity.

Notwithstanding these helpful features, selecting hyperpriors can be a challenging task. While EMC2 yields the optimal estimate for a given set of hyperpriors, it is unclear precisely how changes in the hyperprior selection impact the estimate. For example, error performance bounds for EMC2 for different hyperpriors, such as those available with wavelet analysis and some of the methods described below, have currently not been developed. Finally, numerically computing the expectation and variance of the posterior can be computationally demanding.

5.4. Complexity regularization

An alternative to the MAP approach described above is to circumvent the use of explicit prior probability models by using complexity regularization. Here the key idea is to choose the estimate which is the best balance between fidelity to the data and simplicity of the model. In general, a model is considered "simpler" when more of the MMI parameters $\rho_{j,k}$ equal one-fourth, which is equivalent to having the corresponding wavelet coefficient equal to zero. The complexity measure can be regarded as a negative log prior on the intensity.

This approach has two advantages over the MAP method described above. First, it consistently outperforms MAP estimation in our simulations, suggesting that the complexity-based prior is a better

model for the images considered in our experiments. Second, it allows us to derive error performance bounds similar to those associated with wavelet estimation and prove that it is near-minimax optimal for certain broad classes of intensities. This type of theoretical analysis may be possible in the context of the MAP or EMC2 procedures outlined above, but it has not been accomplished to date. It should be noted, however, that the theoretical results derived in the context of complexity regularization do *not* detract from alternative multiscale methods such as those described above; rather, the simplicity of the complexity regularization framework facilitate an analysis which strongly *supports* the efficacy of the multiscale approach.

Kolaczyk and Nowak (2004) propose estimating $\boldsymbol{\lambda}$ according to

$$\begin{aligned} \widehat{\boldsymbol{\lambda}} &= \underset{\boldsymbol{\lambda}' \in \Gamma}{\operatorname{argmin}} \left\{ -\log p(\boldsymbol{x}|\boldsymbol{\lambda}') + 2\mathrm{pen}(\boldsymbol{\lambda}') \right\} \\ \mathrm{pen}(\boldsymbol{\lambda}') &\equiv \gamma \cdot \log n \cdot \# \left\{ \rho_{j,k,l}(\boldsymbol{\lambda}') \neq 1/4 \right\} \end{aligned}$$

where

$$p(\boldsymbol{x}|\boldsymbol{\lambda}) \equiv \prod_{i=1}^{N} \frac{e^{-\lambda_i} \lambda_i^{x_i}}{x_i!},$$

γ is a tuning parameter, $\# \left\{ \rho_{j,k,l}(\boldsymbol{\lambda}') \neq 1/4 \right\}$ counts the number of MMI coefficients which are not trivial ($1/4$), and Γ is a finite class of potential models. It is demonstrated that when Γ consists of intensities corresponding to pruned quad-trees, so that if $\rho_{j,k,l}$ is set to $1/4$, then $\rho_{j+1,2k,2l}$, $\rho_{j+1,2k+1,2l}$, and $\rho_{j+1,2k,2l+1}$ must also be $1/4$, this minimization can be computed very efficiently and exhibits the following error performance bound. When n (the total number of observed photons) is close to N and $\boldsymbol{\lambda}$ is in a Besov space with smoothness parameter $0 < \alpha < 1$:

$$\mathbb{E}\left[\|\widehat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}\|_2^2 \right] \preceq \left(\frac{\log N}{N} \right)^{2\alpha/(2\alpha+d)}.$$

While this bound does not hold for a wide range of α's because of the underlying piecewise constant nature of the estimate (corresponding to Haar wavelets), it nevertheless is true for intensities of photon-limited observations unlike the earlier analysis in the Anscombe transform domain.

Note that the restriction that all $\boldsymbol{\lambda}' \in \Gamma$ correspond to pruned quad-trees may be a key component of its effectiveness in our simulations. As noted by Donoho (1997), quadtrees necessarily dictate certain hereditary relationships on the Haar coefficients. Our experiments have revealed that wavelet coefficient thresholding can be more robust to noise when the thresholding rule has a *hereditary constraint*, i.e. when a coefficient can only be thresholded if all its descendants are also thresholded. Previous studies (e.g. Crouse et al. (1998)) have also demonstrated that consideration of parent-child relationships can result in improved signal and image estimation.

5.5. Cycle spinning

As mentioned above, cycle spinning can result in significant improvements in multiscale image reconstruction. In wavelet-based methods, cycle spinning can be performed in a very computationally efficient manner using an undecimated wavelet transform. With MMI estimates computed by tree pruning, however, cycle spinning becomes a slightly more complicated operation. In particular, cycle spun tree pruning estimates for Poisson data can be complicated because thresholding decisions are based not only on the magnitude of the coefficients (as in traditional wavelet denoising), but also on the coefficients of the node's descendants (Kolaczyk and Nowak 2004).

A naïve implementation can easily require $O(N^2)$ operations, while estimation via undecimated wavelet transforms only requires $O(N \log N)$ operations. However, Willett and Nowak (2004) show that

hereditary estimates of Poisson intensities can be computed with the same computational complexity as traditional undecimated wavelet thresholding techniques (Lang et al. 1996). Using this method, only $O\left(N \log N\right)$ operations are required to calculate the translation invariant multiscale estimate of an image.

5.6. Experiments

The MMI models described in this section is demonstrated in the context of MAP estimation and complexity regularization in Figure 1(h-i) and Figure 2 and 3(g-h). For imaging, the prior distribution takes the form of a Dirichlet prior, which is a multidimensional analog of the beta distribution. In this simulation, we let α_{J-1} be a free (tuning) parameter and then set $\alpha_j = \alpha_{j+1}/2$; it was observed that this procedure yielded a low ℓ_1 error. The resulting MAP estimate is certainly superior to the raw wavelet and Anscombe transform based methods discussed earlier, and comparable to the modified thresholds proposed by Kolaczyk and Dixon (2000).

The result of complexity regularization is a very accurate estimate. As mentioned above, there are two possible reasons to explain the difference in performance: First, the prior distribution implicitly imposed by the complexity regularization framework may be a better match for the image under consideration. Second, the nature of the tree-pruning method induces a hereditary constraint on the multiresolution analysis, which is analogous to considering parent-child relationships in conventional wavelet estimation.

6. Platelets

The complexity regularization scheme above involved fitting constant models to each leaf node in a pruned quadtree representation of the intensity. This can be an effective tool in many situations, but it has some limitations. First, the error performance bounds hold for intensities in Besov spaces with smoothness parameter α for $0 < \alpha < 1$. In contrast, wavelet-based methods, in the presence of Gaussian noise, have near-minimax optimal rates of convergence for larger α corresponding to smoother images when a wavelet with $\lceil \alpha \rceil$ vanishing moments, where $\lceil \cdot \rceil$ denotes the ceiling operation. Achieving near-minimax optimality for larger degrees of smoothness requires higher-order model fits. Second, wavelet-based methods in general are not well-suited to the estimation of images containing boundaries or edges. This is because a large number of wavelets with large magnitudes and highly correlated supports are needed to accurately estimate smooth boundaries.

Platelets were developed by Willett and Nowak (2003) to address these two challenges. Platelets are localized atoms at various locations, scales and orientations that can produce highly accurate, piecewise linear approximations to images consisting of smooth regions separated by smooth boundaries. Platelets generalize Donoho's wedgelets (Donoho 1999), and like complex wavelets, steerable pyramids, and curvelets, platelets are capable of concisely representing edges at various scales, locations and *orientations*. Platelet approximations can significantly outperform conventional wavelet and wedgelet approximations (with respect to the number of terms required to approximate images to within a given tolerance). Platelets can be incorporated into the complexity regularization scheme described above to produce near-minimax optimal intensity estimates for images composed of smooth surfaces separated by smooth boundaries.

One of the key aspects of platelet analysis is that one need not restrict the image decomposition to the dyadic square partitions associated with pruning quadtree representations. The wedgelet partition (Donoho 1999) is based on a recursive dyadic square partition of the image in which the final nodes are allowed to terminate with a wedge instead of a square. The wedge split is defined by a line connecting two points on two sides of the square. The points are not arbitrary; rather, they are chosen from a finite set of vertices spaced evenly apart around the perimeter of the square. This restriction is crucial because it means that the resulting "dictionary" of wedgelet elements is finite and easily computed.

Furthermore, instead of approximating the image on each piece of the partition by a constant, we can approximate it with a planar surface. In many applications it is beneficial to have this added flexibility. Image gradients, or smooth transitions between regions of varied intensities, encode information about light emission as well as surface geometry. Each planar fit requires three parameters, compared with the one parameter for piecewise constant approximation. Additional parameters must be used to describe the location and orientation of the optimal wedgelet split for each leaf in the quad tree. Although each platelet has more parameters per term, for images of sufficient smoothness many fewer platelets than constant blocks are needed to approximate an image to within a certain error. Thus, platelet approximations may require far fewer parameters for a given level of accuracy.

Similarly to our previous discussion of complexity regularization, we compute the estimate as

$$\begin{aligned} \widehat{\boldsymbol{\lambda}} &= \operatorname*{argmin}_{\boldsymbol{\lambda}' \in \Gamma_P} \left\{ -\log p(\boldsymbol{x}|\boldsymbol{\lambda}') + 2\mathrm{pen}(\boldsymbol{\lambda}') \right\} \\ \mathrm{pen}(\boldsymbol{\lambda}') &\equiv \gamma \cdot \log n \cdot \# \left\{ \{\text{PlateletCoefficients}\}(\boldsymbol{\lambda}') \right\} \end{aligned}$$

where the penalty is proportional to the number of platelet coefficients associated with the estimate $\boldsymbol{\lambda}'$. The family of intensities from which we draw our estimate, Γ_P, consists of intensities corresponding to wedgelet partitions with platelet models fit to the leaves. Willett and Nowak (2005) show that this estimator exhibits the following error performance bound when n (the total number of observed photons) is close to N and $\boldsymbol{\lambda}$ is composed of Hölder-α smooth surfaces separated by Hölder-α smooth boundaries for $0 < \alpha < 2$:

$$\mathbb{E}\left[\|\widehat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}\|_2^2 \right] \preceq \left(\frac{\log N}{N} \right)^{2\alpha/(2\alpha+d)}.$$

This class of piecewise Hölder-α images includes piecewise polynomial images of degree α. The fact that platelets can perform effectively on piecewise smooth images is a critical advantage over traditional wavelet-based methods, which cannot adequately approximate boundaries and edges in images.

As demonstrated in (Willett and Nowak 2005, 2003), platelets can be highly effective for photon-limited observations of piecewise smooth intensities. This efficacy comes at a computational price, however. A platelet estimate of an image with N pixels can be calculated with $O(N^{4/3})$ calls to a convex minimization routine to compute the maximum likelihood platelet fits to each leaf node. If platelet coefficients are calculated using least squares, then the overall computation complexity is $O(N^{7/3})$ operations. The method can be sped up by further restricting the search space Γ_P, but this can impact the quality of the estimate. Castro et al. (2004) propose a computational technique to dramatically reduce the computational burden, but it remains slower than the cycle-spun Haar estimates described above. In many applications, the increase in accuracy is worth the computational cost, especially considering that the algorithm is parallelizable and fast hardware implementations are feasible. However, many astronomical images consist of point sources and/or extended emissions without edges or boundaries.

The empirical performance of platelet estimation is displayed in Figure 1(j) and Figures 2 and 3(i). Because of the highly nonlinear process of choosing the optimal wedgelet partition, the cycle spinning cannot be performed as efficiently as it was in the previous sections, and instead must be performed by explicitly averaging over shifts. For this simulation, we averaged over one hundred randomly chosen shifts. The result is the best estimate presented in this paper, both in terms of ℓ_1 error and visual perception. It exhibits few spurious artifacts and improved estimation in edge regions of the image, and it is photon flux preserving.

7. *À trous* wavelet denoising

The pioneering work of Starck and Murtagh (2002) developed the use of the *à trous* wavelet transform for astronomical image analysis. Introduced by Holschneider et al. (1989) and described in detail in

Starck and Murtagh (2002), the *à trous* transform is a multiscale redundant transform in which the wavelet function is defined as the difference between the scaling function at two consecutive scales; i.e. $\frac{1}{2}\psi(\frac{t}{2}) = \phi(t) - \frac{1}{2}\psi(\frac{t}{2})$. Then the intensity image at the i^{th} pixel can be written as

$$\lambda_i = c_{0,i} + \sum_{j=0}^{J-1} w_{j,i},$$

where j denotes scale and $j = J$ corresponds to the finest (pixel) scale, $w_{j,i}$ is the i^{th} *à trous* wavelet coefficient at scale j, and $c_{0,i}$ is the i^{th} scaling coefficient. This framework facilitates a very intuitive understanding of each coefficient, which makes the transform an effective tool not only for intensity estimation, but also for source detection and other common problems in astronomical data analysis.

However, unlike the Haar wavelet based methods, such as the MMI model, described above, each coefficient is not proportional to the sum or ratio of photon counts, and hence the distribution of the coefficients can be difficult to characterize theoretically. When photon counts are sufficiently large, the *à trous* method can used in conjunction with a variance stabilizing transform. When photon counts are very low, however, the combination of the *á trous* transform and a variance stabilizing transform is subject to some of the challenges discussed earlier in this paper. To address this issue, Slezak et al. (1993); Starck and Murtagh (2006) present methods for performing a variance stabilizing transform on each of the *à trous coefficients* (as opposed to the observations as described above), and then using the empirical distribution of these coefficients to distinguish "significant" coefficients.

One of the key advantages of this procedure is that it can be used to perform low-count Poisson intensity estimation with wavelets which are much smoother than the Haar wavelets required by several of the methods describe above. While performance bounds for this method have not yet been derived, the performance bounds described above for related methods support this procedure. In addition, it is highly effective in practice, as displayed in Figures 2 and 3j. This result, contributed by Jean-Luc Starck, was computed using a cubic spline wavelet. While the estimate is more smooth than Haar wavelet based estimates, it exhibits some dark ringing artifacts around the sources due to the shape of the wavelet function.

8. Conclusions

This review has presented an overview of several distinctive multiscale methods for estimating inhomogeneous Poisson image intensities. We have seen that classical wavelet thresholding, while simple to implement and effective for very high photon counts, yields poor performance for very low photon counts. Even when variance stabilizing transforms such as the Anscombe transform are used in conjunction with classical wavelet thresholding techniques, photon counts must exceed about thirty per pixel for satisfactory results; furthermore, photon flux is not preserved. From here it is clear that more specialized techniques, which account for the Poisson statistics of the observations, are necessary. Using a variance stabilizing transform in the *à trous* wavelet domain appears to be a highly effective tool which facilitates smooth wavelet representations, but thresholding here is more computationally complex than in the Haar-based methods. Haar wavelets are a particularly convenient multiscale construct in the context of photon limited imaging. The corrected Haar thresholds are an effective and computationally efficient tool, particularly when the goal is to distinguish a source from a known background; however, it is difficult to extend this framework to smoother wavelet families. Multiplicative Multiscale Innovation (MMI) models are also closely related to Haar wavelet analysis, and can be used in conjunction with MAP estimation, more formal Bayesian approaches, and complexity regularization. The Bayesian approach in general provides the user with a high degree of flexibility which can be used to incorporate prior knowledge of the inten-

sity. Platelets were also shown to be an effective tool for photon-limited multiscale analysis, particularly in the presence of smooth surfaces and boundaries.

A particularly nice feature of the MMI models and platelets is that they lend themselves readily to theoretical performance bounds. These bounds provide support for the observed practical effectiveness of these and many other methods. While similar bounds have not been explicitly established for all the other methods described or referenced in this paper, the existing bounds can certainly be interpreted to lend strong support to Bayesian multiscale methods in general. While complexity regularization has not been presented here in a formally Bayesian framework, it is implicitly a Bayesian method, with the complexity penalty playing the role of a negative log prior probability. In general, the form of the prior distribution or the complexity term dictates the class of images for which the method performs optimally. In this paper we have focused on the perspective that images of interest are smooth or piecewise smooth, and seen that several commonly used (explicit or implicit) priors lead to excellent performance in very relevant images.

One might question, in light of this review, which method to apply for his/her imaging problem. Note that the images used in this paper's experiments were deliberately chosen to highlight some of the differences between the methods studied; these differences are more or less prevalent depending on the structure being imaged. For the Saturn image, which consisted of smooth regions separated by smooth boundaries, platelets performed very well and exhibited a near-optimal theoretical performance characterization. However, platelet estimation has a non-negligible computational cost in its current implementations. As noted above, it is parallelizable and hardware implementations are possible, but for images void of edges and boundaries alternative methods may be preferable. For example, platelets would not perform significantly better than any of the other methods presented in this paper for images consisting primarily of compact sources and extended and complex emission. The *à trous* method combined with a variance stabilizing transform on the coefficients and an empirical Bayes thresholding scheme yielded visually appealing results on the XMM simulation image, despite some ringing artifacts. The complexity regularized quadtree pruning method yielded less smooth results, but nevertheless worked very well in all simulations and experiments conducted in the preparation of this review.

Software

Software for the complexity regularized quadtree pruning method is available online at http://www.ee.duke.edu/∼willett/software.html.

Acknowledgments

The author would like to thank Robert Nowak, Eric Kolaczyk, and David van Dyk for helpful feedback and discussions during the preparation of this document.

References

A. Aldroubi and M. Unser. *Wavelets in Medicine and Biology*. CRC Pr., Boca Raton FL, 1996.

F. J. Anscombe. The tranformation of poisson, binomial, and negative-binomial data. *Biometrika*, 15:246–254, 1948.

A. Antoniadis and T. Sapatinas. Wavelet shrinkage for natural exponential families with quadratic variance functions. *Biometrika*, 88:805–820, 2001.

R. Beran and L. Dümbgen. Modulation of estimators and confidence sets. *Ann. Statist.*, 26:1826–1856, 1998.

P. Besbeas, I. De Feis, and T. Sapatinas. A comparative simulation study of wavelet shrinkage estimators for poisson counts. *Internat. Statist. Rev*, 72(2):209–237, 2004.

A. Bijaoui and G. Jammal. On the distribution of the wavelet coefficient for a poisson noise. *Signal Processing*, 81:1789–1800, 2001.

L. D. Brown, T. Cai, R. Zhang, L. H. Zhao, and H. Zhou, 2006. Personal communication.

E. Candès and D. Donoho. *Curves and Surfaces*, chapter Curvelets: A Surprisingly Effective Nonadaptive Representation For Objects with Edges. Vanderbilt University Press, Nashville, TN, 1999.

R. Castro, R. Willett, and R. Nowak. Coarse-to-fine manifold learning. In *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing — ICASSP '04*, 17-21 May, Montreal, CA, 2004.

R. Coifman and D. Donoho. Translation invariant de-noising. in *Lecture Notes in Statistics: Wavelets and Statistics*, New York: Springer-Verlag:125–150, 1995.

M. Crouse, R. Nowak, and R. Baraniuk. Wavelet-based statistical signal processing using hidden markov models. *IEEE Transactions on Signal Processing*, 46(4):886–902, 1998.

R. A. DeVore. Nonlinear approximation. *Acta Numerica*, 7:51–150, 1998.

D. Donoho. De-noising via soft-thresholding. Technical Report Statistics 409, Stanford University, 1992.

D. Donoho. Nonlinear wavelet methods for recovery of signals, densities, and spectrafrom indirect and noisy data. In *Proceedings of Symposia in Applied Mathematics, American Mathematical Society*, pages 173–205, 1993.

D. Donoho. Cart and best-ortho-basis selection: A connection. *Annals of Stat.*, 25:1870–1911, 1997.

D. Donoho. Wedgelets: Nearly minimax estimation of edges. *Ann. Statist.*, 27:859 – 897, 1999.

D. Donoho and I. Johnstone. Minimax estimation via wavelet shrinkage. *Annals of Statistics*, 26(3):879–921, 1998.

D. N. Esch, A. Connors, M. Karovska, and D. A. van Dyk. An image restoration technique with error estimates. *The Astrophysical Journal*, 610:1213–1227, 2004.

M. Fisz. The limiting distribution of a function of two independent random variables and its statistical application. *Colloquium Mathematicum*, 3:138–146, 1955.

P. Fryźlewicz and G. Nason. Poisson intensity estimation using wavelets and the fisz transformation. Technical Report 01/10, Department of Mathematics, University of Bristol, United Kingdom, 2001.

M. Holschneider, R. Kronland-Martinet, J. Morlet, and P. Tchamitchian. *Wavelets, Time-Frequency Methods and Phase-Space*, chapter Real-time algorithm for signal analysis with the help of the wavelet transform, pages 289–297. Springer, Berlin, 1989.

N. Kingsbury. Image processing with complex wavelets. *Phil. Trans. Royal Society London A*, 357:2543–2560, 1999.

E. Kolaczyk. Wavelet shrinkage estimation of certain poisson intensity signals using corrected thresholds. *Statistica Sinica*, 9:119–135, 1999a.

E. Kolaczyk and D. Dixon. Nonparametric estimation of intensity maps using haar wavelets and poisson noise characteristics. *The Astrophysical Journal*, 534(1):490–505, 2000.

E. Kolaczyk and R. Nowak. Multiscale likelihood analysis and complexity penalized estimation. *Annals of Stat.*, 32:500–527, 2004.

E. D. Kolaczyk. Bayesian multi-scale models for poisson processes. *Journal of the American Statistical Association*, 94:920–933, 1999b.

A. N. Kolmogorov. The local structure of turbulence in compressivle viscous fluid for very large reynolds number. *Dokl. Akad. Nauk SSSR*, 30:9–13, 1941.

M. Lang, H. Guo, J. E. Odegard, C. S. Burrus, and R. O. Wells. Noise reduction using an undecimated discrete wavelet transform. *IEEE Signal Processing Letters*, 3(1):10–12, 1996.

L. B. Lucy. An iterative technique for the rectification of observed distributions. *Astron. J.*, 79:745–754, 1974.

S. Mallat. *A Wavelet Tour of Signal Processing*. Academic Press, San Diego, CA, 1998.

B. B. Mandelbrot. *The Fractal Geometry of Nature*. W. H. Freeman and Co., San Francisco, 1982.

R. Nowak and E. Kolaczyk. A multiscale statistical framework for Poisson inverse problems. *IEEE Trans. Info. Theory*, 46:1811–1825, 2000.

W. Richardson. Bayesian-based iterative method of image restoration. *J. Opt. Soc. of Am.*, 62:55–59, 1972.

S. Sardy, A. Antoniadis, and P. Tseng. Automatic smoothing with wavelets for a wide class of distributions. *Journal of Computational and Graphical Statistics*, 13(2):399–421, 2004.

L. A. Shepp and Y. Vardi. Maximum likelihood reconstruction for emission tomography. *IEEE Trans. Med. Imaging*, 1:113–122, 1982.

E. Slezak, V. de Lapparent, and A. Bijaoui. Objective detection of voids and high-density structures in the first cfa redshift survey slice. *Astrophysical Journal, Part 1*, 409:517–529, 1993.

J. Starck, F. Murtagh, and A. Bijaoui. *Image Processing and Data Analysis: The Multiscale Approach*. Cambridge Univ. Press, 1998.

J.-L. Starck and F. Murtagh. *Astronomical Image and Data Analysis*. Springer, Berlin, 2002.

J.-L. Starck and F. Murtagh. *Astronomical Image and Data Analysis*, 2nd edition. In press, 2006.

K. Timmermann and R. Nowak. Multiscale modeling and estimation of Poisson processes with application to photon-limited imaging. *IEEE Transactions on Information Theory*, 45(3):846–862, 1999.

R. Willett and R. Nowak. Fast multiresolution photon-limited image reconstruction. In *Proc. IEEE Int. Sym. Biomedical Imaging — ISBI '04*, 15-18 April 2004, Arlington, VA, USA, 2004.

R. Willett and R. Nowak. Multiscale poisson intensity and density estimation. Submitted to *IEEE Trans. Info. Th.*. Available at http://www.ee.duke.edu/∼willett/papers/WillettInfoTh2005.pdf, 2005.

R. Willett and R. Nowak. Platelets: a multiscale approach for recovering edges and surfaces in photon-limited medical imaging. *IEEE Transactions on Medical Imaging*, 22(3), 2003.