The Astrophysical Journal, 610:1213–1227, 2004 August 1

AN IMAGE RESTORATION TECHNIQUE WITH ERROR ESTIMATES

David N. Esch
Center for Health Quality, Outcomes, and Economics Research, Department of Veterans Affairs,
200 Spring Road, Bedford, MA 01730; esch@stat.harvard.edu

Alanna Connors
Eureka Scientific, 2452 Delmer Street, Suite 100, Oakland, CA 84602-3017

Margarita Karovska
Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138

and

David A. van Dyk
Department of Statistics, University of California, Irvine, CA 92697-1250
Received 2004 January 12; accepted 2004 April 6

ABSTRACT

Image restoration including deconvolution techniques offers a powerful tool to improve resolution in images and to extract information on the multiscale structure stored in astronomical observations. We present a new method for statistical deconvolution, which we call expectation through Markov Chain Monte Carlo (EMC2). This method is designed to remedy several shortfalls of currently used deconvolution and restoration techniques for Poisson data. We use a wavelet-like multiscale representation of the true image to achieve smoothing at all scales of resolution simultaneously, thus capturing detailed features in the image at the same time as larger scale extended features. Thus, this method smooths the image, while maintaining the ability to effectively reconstruct point sources and sharp features in the image. We use a principled, fully Bayesian model–based analysis, which produces extensive information about the uncertainty in the fitted smooth image, allowing assessment of the errors in the resulting reconstruction. Our method also includes automatic fitting of the multiscale smoothing parameters. We show several examples of application of EMC2 to both simulated data and a real astronomical X-ray source.

Subject headings: methods: data analysis — techniques: high angular resolution

1. INTRODUCTION

Recent high angular resolution ground- and space-based observations have produced unprecedented views of various astronomical objects at wavelengths ranging from γ-rays to radio waves. These observations contain many complex components with different spatial scales and a wide range of contrast levels. Spatial characterization of these components is extremely important for understanding the physical characteristics of these sources. The multiscale structures in the images of astronomical sources, however, often cannot readily be studied directly, because of either the limited resolution and noise in the images or the low contrast of the small-scale structures when compared to the large-scale features. Even the most spectacular high angular resolution images are blurred by the atmosphere, the point-spread function (PSF) of the telescope, or by instrumental effects including the limited sizes of detector pixels.

Early techniques implicitly assumed known, constant, Gaussian white noise—from Wiener filters, which often overblurred locally sharp features, to autocorrelation-based techniques and CLEAN algorithms (Kaaresen 1997), which worked for point sources. More robust maximum-likelihood, maximum-entropy, and general probabilistic-based techniques such as Richardson-Lucy (Richardson 1972; Lucy 1974) and the EM algorithm (Dempster et al. 1977) still often assumed a single uniform scale (usually the pixel size) for features in the "true" image. These techniques tend to sharpen pointlike sources but produce grainy restoration of extended features. Even with a prior or regularizer added, as in maximum entropy method (MEM; Frieden 1972; Gull & Skilling 1991) and MEMSYS (Skilling & Bryan 1984), the result can be unrealistically grainy (but see Knödelseder et al. 1996).

More recently, there has been a rush to use wavelets and other multiscale methods that could reconstruct both sharp and broad features across an image. First efforts to adapt them to Poisson statistics either did not include the instrument responses (Kolaczyk 1999; Timmermann & Nowak 1999; Kolaczyk & Dixon 2000) or very cleverly and exhaustively approximated the effects of Poisson statistics on thresholds (Murtaugh et al. 1995). However, these algorithms lacked the elegance of "orthogonal factorization" that made wavelets so key in the Gaussian regime. (See Starck & Murtagh 2002 for a nice review.)

One final shortcoming of most of the previous commonly used deconvolution techniques is that they do not produce reliable convergence (the "stopping" problem) and uncertainty information. Thus, the user cannot easily evaluate the accuracy of the results. This is a serious shortcoming, especially in the case of images characterized by low count statistics.

Statistical modeling and estimation provide the perfect framework for solving these problems and for quantifying the errors in the solution, especially in low-count imaging situations. We present a new method for statistically accounting

for a PSF by deconvolving a blurred source image. The technique can be thought of as an extension of two previous deconvolution techniques: the Richardson-Lucy/EM technique and a Nowak-Kolaczyk multiscale technique explicitly for Poisson data that is similar to the Haar wavelet. Our method is ideally suited for restoration of images of sources containing multiscale structures including extended emission. The statistical model contains smoothing parameters that are assigned a common hyperprior distribution and fit with the rest of the model parameters. Although these parameters are automatically fit by the model, the range over which they are fitted is regulated by this hyperprior distribution. One sets this range from calibrations prior to the analysis, rather than during restoration. We call this method expectation through Markov chain Monte Carlo (EMC2). It includes regularization and is based on a fully Bayesian formulation that allows for a consistent top-to-bottom approach to model construction, model fitting, and error estimation. As a result, the output includes error information that can be used to construct confidence statements.

The article is organized into the following sections. In § 2 we summarize the main elements of the statistical Bayesian formulation, as well as our multiscale representation of the data. The EMC2 technique, our Markov chain Monte Carlo–based method for fitting intensities to data and mapping out uncertainties, is described in § 3. Examples of both simulated and real data are in § 4, and conclusions are given in § 5.

2. STATISTICAL BACKGROUND

2.1. *Bayesian Statistical Models*

Statistical analysis of an astronomical image composed of pixel counts begins with a "statistical model" that is a summary of both a "source" or "astrophysical" model for the image and the stochastic processes involved in data collection and recording. The source model describes the spatial structure in the relative intensities of photon emission in the source; this model is of principle scientific interest. In addition to the source model, the statistical model describes such instrumental effects as the PSF and other processes that effect the observed data, such as variability in the observed source counts (e.g., Poisson variability), the exposure map, and background contamination of the counts.

The statistical model is expressed in terms of a set of unknown parameters, which are of scientific interest since they directly describe the image. The variability in the observed data is represented by a probability distribution, $p(\boldsymbol{X}|\Lambda)$, where $\boldsymbol{X}$ represents the array of pixel counts and Λ represents the astrophysical model or image; Λ is the array of relative pixel intensities of the source emission across the image and is scaled to preserve the total expected counts. Statisticians refer to Λ as a "model parameter."

A standard mode of statistical inference "restores" the image by computing the value Λ that maximizes $p(\boldsymbol{X}|\Lambda)$ as a function of Λ with $\boldsymbol{X}$ fixed at the values of the observed counts. This estimate of Λ is known as the maximum likelihood estimate (MLE) and

$$L(\Lambda|\boldsymbol{X}) = p(\boldsymbol{X}|\Lambda) \tag{1}$$

is called the "likelihood function." [Although the likelihood function and $p(\boldsymbol{X}|\Lambda)$ are equal, the likelihood is viewed as a function of the parameter given the observed data, while $p(\boldsymbol{X}|\Lambda)$ is viewed as a function of the data for a given value of the parameter.] Note that the MLE is the value of Λ that maximizes $L(\Lambda|\boldsymbol{X})$.

A Bayesian statistical analysis is based not only on the likelihood function, but also on a second component known as the "prior distribution." The prior distribution quantifies information or knowledge about the likely values of the model parameter, or structures in the model parameter, that are known *prior* to data collection. Such prior information may include an expectation that the true image is smooth, i.e., the image intensities in nearby or adjacent regions of the image should be similar. This information is quantified through a (prior) probability function that we denote $p(\Lambda)$. In the methods that we describe in §§ 3.2 and 3.3 $p(\Lambda)$ primarily quantifies smoothness constraints rather than information as to the actual parameter values. We illustrate such prior distributions below, with details given in § 3.2. The prior distributions themselves are often parametrized with a number of parameters; the values of these parameters may substantially affect the resulting model fit or image restoration. In such cases we may wish to fit these parameters rather than fix them at a set of arbitrary values. We can achieve this by assigning another prior distribution to these parameters. This second prior distribution is then known as a "hyperprior" distribution, and the parameters in the prior distribution on which the hyperprior distribution is imposed are called "hyperparameters." For this application, two hyperparameters govern the total intensity of the restored image, and the rest govern the degree of smoothing in the final restoration; thus, we refer to the latter subset of hyperparameters as "smoothing parameters."

Inference in Bayesian statistics is based on the "posterior distribution," which is the distribution of the unknown model parameters of interest, given the observed data. The posterior distribution is calculated using Bayes theorem and is proportional to the product of the likelihood, prior distribution, and perhaps hyperprior distribution:

$$p(\Lambda|\boldsymbol{X}) = \frac{p(\boldsymbol{X}|\Lambda)p(\Lambda)}{p(\boldsymbol{X})} \propto p(\boldsymbol{X}|\Lambda)p(\Lambda), \tag{2}$$

or

$$p(\Lambda,\ \boldsymbol{A}|\boldsymbol{X}) = \frac{p(\boldsymbol{X}|\Lambda)p(\Lambda|\boldsymbol{A})p(\boldsymbol{A})}{p(\boldsymbol{X})} \propto p(\boldsymbol{X}|\Lambda)p(\Lambda|\boldsymbol{A})p(\boldsymbol{A}), \tag{3}$$

where $\boldsymbol{A} = \{\alpha_1, \ldots, \alpha_K\}$ represents the smoothing parameters. In practice, we need only calculate the posterior distribution up to a proportionality constant, since many model fitting routines do not require normalized distributions. From a Bayesian perspective $p(\Lambda|\boldsymbol{X})$ is a complete summary of the information for Λ available in the data and in the prior distribution. When hyperparameters are included in the model, $p(\Lambda|\boldsymbol{X})$ can be computed via $\int p(\Lambda,\ \boldsymbol{A}|\boldsymbol{X})\, d\boldsymbol{A}$. Since $p(\Lambda|\boldsymbol{X})$ is a probability function, its mean and its mode are natural summaries and candidates for the reconstructed image. The mode is called the "maximum a posteriori" (MAP) estimate and denoted $\widehat{\Lambda_{\mathrm{MAP}}}$; the posterior mean (or posterior expectation) is denoted $E(\Lambda|\boldsymbol{X})$. In low-count images the shape of $p(\Lambda|\boldsymbol{X})$ is often skewed toward zero for many of the pixel intensities (see Fig. 1). In this case the mode of the distribution may be zero and a less desirable summary of the distribution

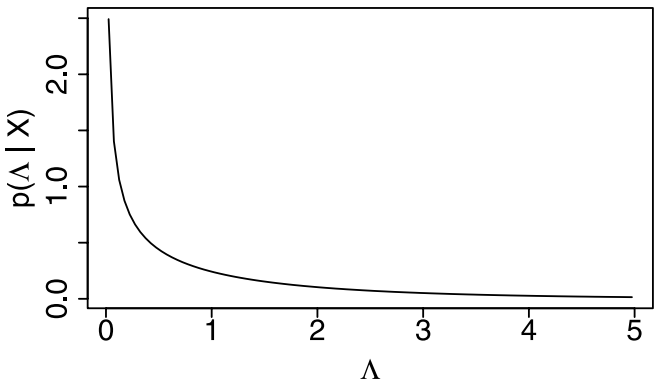

FIG. 1.—Skewed distribution. Posterior distributions of the intensity parameters of low-count Poisson processes can be skewed toward zero. In this case the MAP estimate is zero and is a questionable summary of the distribution.

than the mean. Thus, our methods use $E(\Lambda|\boldsymbol{X})$ to summarize $p(\Lambda|\boldsymbol{X})$; this is our reconstructed image. The uncertainty in the reconstructed image can be computed by quantifying the variability of $p(\Lambda|\boldsymbol{X})$. This is how our method computes error bars or error maps.

Computing $p(\Lambda|\boldsymbol{X}) = \int p(\Lambda,\ \boldsymbol{A}|\boldsymbol{X})\, d\boldsymbol{A}$, $E(\Lambda|\boldsymbol{X})$, and the variability of $p(\Lambda|\boldsymbol{X})$ are all computationally challenging tasks. A useful strategy is to simulate a Monte Carlo sample from $p(\Lambda,\ \boldsymbol{A}|\boldsymbol{X})$. In this way we obtain many simulated values of Λ and $\boldsymbol{A}$ from $p(\Lambda,\ \boldsymbol{A}|\boldsymbol{X})$. These simulated values can be used to summarize $p(\Lambda,\ \boldsymbol{A}|\boldsymbol{X})$. For example, we can compute $E(\Lambda|\boldsymbol{X})$ by averaging the simulated values of Λ. This is an example of Monte Carlo integration. A similar technique can be used to quantify the variability of $p(\Lambda|\boldsymbol{X})$ when computing error bars and error maps.

In order to simulate values from $p(\Lambda,\ \boldsymbol{A}|\boldsymbol{X})$ we employ the method of Markov chain Monte Carlo (or MCMC; Metropolis et al. 1953; Hastings 1970). This method constructs a Markov chain with stationary distribution equal to $p(\Lambda,\ \boldsymbol{A}|\boldsymbol{X})$. For example, starting with an initial value of $\boldsymbol{A} = \boldsymbol{A}^{(0)}$, in principle we might simulate $\Lambda^{(1)}$ from $p(\Lambda|\boldsymbol{A}^{(0)},\ \boldsymbol{X})$ and then update the smoothing parameters by simulating $\boldsymbol{A}^{(1)}$ from $p(\boldsymbol{A}|\Lambda^{(1)},\ \boldsymbol{X})$. Iterating between these two steps, we would obtain the Markov chain $\{(\Lambda^{(t)},\ \boldsymbol{A}^{(t)},\ t = 1,\ 2, \ldots\}$. This is an example of the so-called Gibbs sampler. We use a more involved Gibbs sampler to construct a Markov chain with stationary distribution equal to $p(\Lambda,\ \boldsymbol{A}|\boldsymbol{X})$ in § 3. Basic Markov chain theory states that after a suitable burn-in period the chain will deliver simulated values from $p(\Lambda,\ \boldsymbol{A}|\boldsymbol{X})$. Unfortunately, the simulated values obtained are correlated with adjacent values in the chain, and great care must be taken to be sure the burn-in period is adequate. Such practical issues are discussed in a detailed introduction to MCMC methods in astrophysics that is given in van Dyk et al. (2001). They provide an overview of Bayesian methods and computation in astrophysics and develop an MCMC sampler for spectral analysis of astronomical sources (see also van Dyk et al. 2003). Because the reconstructed image is the posterior expectation that is computed using a Monte Carlo sample obtained via MCMC, we refer to our method as expectation through MCMC, or EMC2.

2.2. *Multiscale Image Representation*

Normally we use a pixel-based representation of the image, i.e., the set of pixel intensities $\Lambda = \{\Lambda_i,\ i = 1, \ldots, N\}$, which are stored in an array. The EMC2 procedure also uses a second representation of Λ, i.e., a transformation of Λ. This transformed multiscale representation, rather than storing each pixel intensity, stores the total of all pixel intensities, and a series of "four-way split proportions" that, when multiplied together in the correct sequence, give back the pixel intensities.

The elegance of this multiscale representation lies in that it imposes structure on the image by assigning prior distributions to each split proportion, or to groups of split proportions. This allows us to achieve regularization or smoothing at multiple scales in the image. We express the multiscale representation of the $2^k \times 2^k$ array, Λ as

$$\Lambda_i = G \prod_{k=1}^{K} D_{k,\, l_{k(i)},\, m_{k(i)}}, \tag{4}$$

where G is the overall intensity and $D_{k,\, l_{k(i)},\, m_{k(i)}}$ is the split proportion at scale k corresponding to the group of pixels containing pixel i. The first subscript of $D_{k,\, l_{k(i)},\, m_{k(i)}}$ represents the scale, the second indexes which four-way split is being referenced, and the third subscript indexes the quadrants of this split. Thus, for an 8×8 image, k varies from 1 to 3; each Λ_i is the product of G and three split probabilities. For $k = 1$ there is a single vector of length four of split probabilities, so for $k = 1$, $l_{k(i)} = 1$ for all i and $m_{k(i)}$ runs from 1 to 4 indexing the quadrants of the image. For $k = 2$, there are four vectors of split probabilities, each of length four, so for $k = 2$, $l_{k(i)}$ varies from 1 to 4 and $m_{k(i)}$ again runs from 1 to 4 indexing the components of each vector of split probabilities, i.e., the subquadrants of each quadrant of the image. Likewise, for $k = 3$, $l_{k(i)}$ varies from 1 to 16. More generally, $l_{k(i)}$ runs from 1 to 4^{k-1}, and $m_{k(i)}$ runs from 1 to 4 for all i and all k. It is advantageous to notate the split proportions in this way because in § 3.2 we assign a probability distribution to the vectors $\boldsymbol{D}_{kl} \equiv \{D_{klm},\ m = 1, \ldots, 4\}$.

Figure 2 shows a schematic view of the multiscale representation, in which each layer is shown as a group of pixels, with 4 times as many pixels as its parent layer, and one-quarter as many pixels as its child layer. The top layer is 1 pixel for the entire image, and the lowest layer is the image in its full resolution. The intensities of 2 pixels in each layer are shown using the multiscale notation.

This multiscale representation is wavelet-like in nature; it is quite similar to the two-dimensional Haar wavelet in that it is square in shape, and each scale partitions its parent scale into four parts. However, unlike the Haar wavelet, there are no negative components to the image, and the four parts of each split are kept separately, as opposed to being kept as north-south, east-west, northeast-southwest, and northwest-southeast pairs, as in the two-dimensional Haar wavelet.

Our representation also keeps the entire multiscale representation of the image, rather than pruning back the parts of the multiscale tree for which the data do not provide evidence of differences in intensities. We do this because we prefer the restoration to reflect the full resolution of the instrument across the entire field of view, and we do not have any a priori reason to prefer higher resolutions in certain parts of the image. Our statistical model assumes that the image is well represented by a square pixilation, and for most purposes this assumption is reasonable. As discussed in § 5, it is possible to use a grid that has a finer resolution than the data by using a higher resolution PSF or to use a partition that does not split

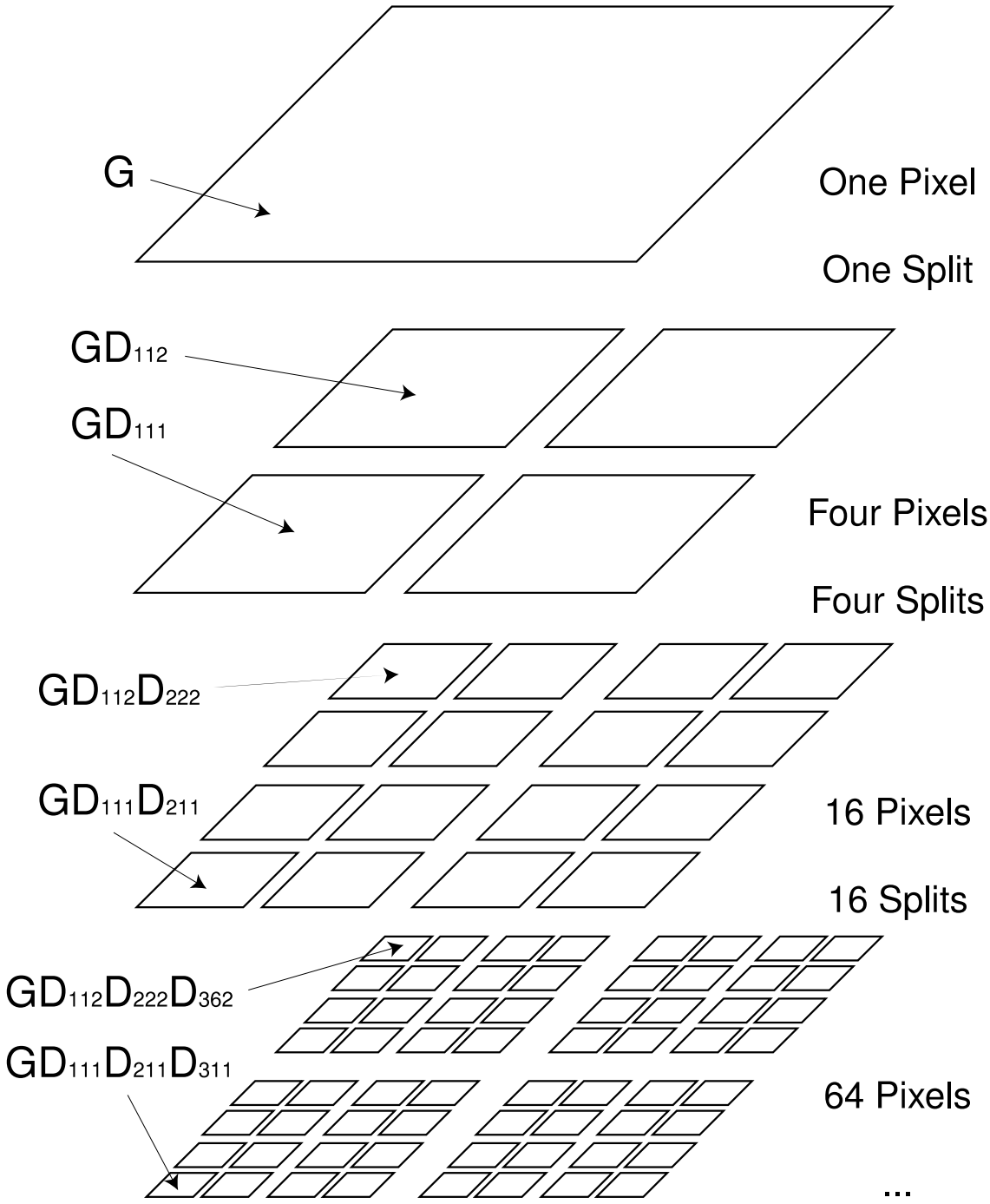

FIG. 2.—Multiscale image representation. The top layer is the total intensity, and each subsequent layer splits the blocks of its parent layer into four parts. The total intensities of the cells indicated with arrows are shown on the left.

along a grid (e.g., Willett & Nowak 2002). Although both of these strategies might lead to finer detail in the reconstructed image, we do not pursue them here. Rather, we focus on developing methods that provide error maps while requiring as little user intervention as possible.

3. THE EMC2 RESTORATION TECHNIQUE

The statistical model used by the EMC2 procedure is comprised of several layers: a likelihood function, a prior distribution, and a hyperprior distribution, each of which is discussed in detail in the following sections. In addition to describing the statistical model, we describe the model fitting technique, summarize the algorithm, and describe the outputs.

3.1. *The Likelihood*

To model the relative intensity of photon emission across the image, we begin by overlaying a grid of pixels on the source image. The photon counts originating from each of these pixels cannot actually be observed because of instrumental effects such as the PSF and other data degradation such as the effects of the exposure map and of instrumental background contamination. In the statistics literature, such unobservable

quantities are called "missing data." We emphasize that the term "missing data" does not imply that we expected to observe these data or that they were somehow lost, but rather the missing data are an elegant statistical construct that helps clarify model assumptions and simplify statistical computation. Once we recognize that "true values" of quantities recorded with measurement error can be treated as "missing data" a large set of statistical tools designed to handle missing data can be employed (see, e.g., van Dyk et al. 2001).

Inclusion of the exposure map for the detector area is important if the area of interest falls on a region over which the exposure map varies substantially. The exposure map accounts for the fact that not all parts of the detector have equal capacity to detect photons; the sections of the chip where detector arrays are joined are, in particular, unable to detect photons. The dithering of the telescope axis somewhat compensates for this effect in that all regions of the sky have nonzero effective exposure time; however, because of the dark regions of the detector, some areas of the image have effectively less exposure time than others, and resulting intensities in the image will be correspondingly less by this well-calibrated known factor. Since the PSF varies considerably across the image for larger images, and we have not yet written computer code to handle variable PSFs, we are limited to smaller images, for which the exposure map is unlikely to vary significantly. Thus, we further assume that the exposure map is constant across the image, and we accordingly suppress the exposure map in the notation of the statistical model. The exposure map will be included in the fitting algorithm in future revisions of the computer code.

We denote the missing pixel counts as $\boldsymbol{X} = \{X_i,\ i \in \mathcal{I}\}$ where i indexes the set of pixels, $\mathcal{I}$; we emphasize that X_i are not contaminated by instrumental background or blurred by the PSF and are therefore unobservable missing data. We refer to these missing pixel counts as "ideal counts." We model the ideal counts as independent Poisson[1] variables,

$$X_i \overset{\mathrm{d}}{\sim} \mathrm{Poisson}(\Lambda_i) \tag{5}$$

where Λ_i is the expected count in pixel i, and $\Lambda = \{\Lambda_i,\ i \in \mathcal{I}\}$ describes the relative intensity of photon emission across the image. Here Λ is the quantity of primary interest.

Given that we do not know the ideal counts $\boldsymbol{X}$, we express the model in terms of the observed pixel counts, which we denote as $\boldsymbol{Y} = \{Y_i,\ i \in \mathcal{I}\}$. These observed counts are distorted by the PSF and contaminated by background counts. In particular, a photon originating in a region of the source corresponding to pixel i has probability P_{ij} of being detected in pixel j. We use the multinomial[2] $(X_i,\ \boldsymbol{P}_i)$ distribution to model the pixel counts resulting from an ideal count of X_i convolved with the PSF, where $\boldsymbol{P}_i = \{P_{ij},\ j \in \mathcal{I}\}$ is the PSF. Here $\sum_j P_{ij} \leq 1$, with equality only if photons originating in pixel i of the source are recorded somewhere on the detector with certainty. In modeling the intensity parameters, we also consider instrumental background contamination, which occurs when a subset of the recorded counts are not due to photons arriving from the source of interest. By including instrumental background in the model as a component of the Poisson intensities, we avoid simply subtracting off the background and ending up with negative "background-corrected" counts. Statistically, we may write

$$\boldsymbol{Y} | \boldsymbol{X} \overset{\mathrm{d}}{\sim} \sum_{i \in \mathcal{I}} \mathrm{Multinomial}(X_i, \boldsymbol{P}_i) + \mathrm{Poisson}(\Lambda^B), \tag{6}$$

where $\boldsymbol{Y} = \{Y_j,\ j \in \mathcal{I}\}$ is the set of observed pixel counts, $\Lambda^B = \{\Lambda^B_i,\ i \in \mathcal{I}\}$ is the vector component of the intensity due to background contamination, and $\mathrm{Poisson}(\Lambda^B)$ is an array of independent Poisson variables with intensities given by Λ^B.

We may combine equation (6) with equation (5) via

$$p(\boldsymbol{Y} | \Lambda,\ \Lambda^B) = \sum_X p(\boldsymbol{Y} | \boldsymbol{X},\ \Lambda,\ \Lambda^B) p(\boldsymbol{X} | \Lambda,\ \Lambda^B)$$

to obtain

$$Y_j | \Lambda, \Lambda^B \overset{\mathrm{d}}{\sim} \mathrm{Poisson}\left[\left(\sum_{i \in \mathcal{I}} P_{ij} \Lambda_i\right) + \Lambda^B_j\right]. \tag{7}$$

The model given in equations (5) and (6) is exactly the same as that given in equation (7), but with an intermediate stage added, involving the missing data, $\boldsymbol{X}$. This additional layer in the model facilitates fitting Λ.

Equation (7) represents the distribution of the observed data given the model parameters and, thus, is the building block of the likelihood function. In particular, since the observed counts $\boldsymbol{Y}$ are independent, we can write the likelihood function as

$$L(\Lambda,\ \Lambda^B | \boldsymbol{Y}) \equiv L(\Lambda | \boldsymbol{Y}) \propto \prod_{j \in \mathcal{I}} p(Y_j | \Lambda), \tag{8}$$

where $p(Y_j | \Lambda)$ is the Poisson distribution given in equation (7); in equation (8) we suppress Λ^B because we assume the background intensities are small compared to the source intensity. Thus, our restored images include background intensities rather than modeling them out.

Inclusion of instrumental background in EMC2 presents additional challenges. The calibration information for *Chandra* instrumental background is an observation of counts, which are relatively small, and closely follow the shape of the detector chips. These counts are presumably Poisson observations with unknown intensities. It is these intensities that must be estimated in order to be used in the model described above. The primary focus of the work described here has been to include the smoothing model and uncertainty information on the restorations, so the details of inclusion of instrumental background are left for future work.

3.2. *The Prior Distribution*

The prior distribution is the second layer of the statistical model. In this layer we use the multiscale representation for the image given in equation (4). We specify a gamma[3] prior distribution for the total intensity G. The gamma distribution

[1] A random variable X is said to follow a Poisson distribution with parameter or intensity Λ if $\Pr(X = x) = e^{-\Lambda}\Lambda^x/x!$. In this case $E(X) = \Lambda$ and we often write $X \overset{\mathrm{d}}{\sim} \mathrm{Poisson}(\Lambda)$ (read as X is distributed as Poisson with intensity Λ). This representation conditions on the intensity parameter, Λ, which in turn may vary.

[2] The multinomial distribution is a random vector allocating a fixed number of counts independently to a series of bins, each with assigned probabilities. Thus, a multinomial (n, π) random variable will be a vector of the same length as π of nonnegative integers whose sum is n.

[3] The gamma distribution is a continuous probability distribution for positive real numbers, with probability distribution $p(x|a,\ b) = b^a x^{a-1} e^{-bx}/\Gamma(a)$, with mean a/b and variance a/b^2. Its flexibility makes it useful for modeling positive quantities.

is restricted to positive values, as is desirable for Poisson intensities, and is both flexible in terms of shape and mathematically tractable, and thus is an ideal choice for modeling the total intensity. We use Dirichlet[4] prior distributions for the four-way split proportions D_{klm}. Variables that follow the Dirichlet distribution are nonnegative vector quantities, restricted to sum to 1. Dirichlet distributions are also mathematically tractable when used as a prior distribution that is combined with a multinomial likelihood; conditional on the other model components, the resulting posterior distribution for the split proportions is also a Dirichlet distribution.[5]

Thus, the building blocks of the prior distribution, the second layer of the statistical model, are

$$G \overset{\mathrm{d}}{\sim} \mathrm{gamma}(\gamma_0,\ \gamma_1),\ \text{and} \tag{9}$$

$$\boldsymbol{D}_{kl} \equiv \{D_{klm},\ m = 1, \ldots, 4\} \overset{\mathrm{d}}{\sim} \mathrm{Dirichlet}(\alpha_k,\ \alpha_k,\ \alpha_k,\ \alpha_k),$$
$$k = 1, \ldots, K, \quad l = 1, \ldots, 4^{k-1}. \tag{10}$$

The effect of the Dirichlet prior distribution can be expressed in terms of the resulting posterior distribution. In particular, the posterior distribution obtained with $\alpha_j = n$, where n is a positive integer, is the same as the posterior distribution obtained with $\alpha_j = 0$, and n counts in each quadrant of the split in the ideal data $\boldsymbol{X}$. Hence, the prior distribution can be thought of as adding an equal number of ideal data counts to all four corners of each split at its given resolution in the multiscale representation of the image. This has the effect of regularizing the splits and, together, smoothing the image.

The complete prior distribution is simply the product of the probability densities corresponding to equation (9) and a copy of equation (10) for each $\boldsymbol{D}_{kl}$ in the multiscale image representation given in equation (4). $\boldsymbol{A} = \{\alpha_k,\ k = 1, \ldots, K\}$ and $\{\gamma_0,\ \gamma_1\}$ are the hyperparameters of the statistical model, and $\boldsymbol{A} = \{\alpha_k,\ k = 1, \ldots, K\}$ are the smoothing parameters. As discussed in Nowak & Kolaczyk (2000), the settings of $\{\gamma_0,\ \gamma_1\}$ only affect the total flux of the fitted restoration and not the pixels' relative intensities, so we fix these parameters so that the posterior mean flux of the restoration is equal to the observed flux, e.g., γ_0 is set to the total observed counts, and γ_1 is set to 1. Thus, total flux will be preserved, on average, by the statistical model.

Setting the smoothing parameters, $\boldsymbol{A}$, requires considerably more care: Different settings affect ratios of pixel intensities and can produce quite different results. Larger values produce more smoothing at a given resolution. At each scale we choose all four parameters of each Dirichlet prior distribution to be equal, because this in effect shrinks the reconstruction toward a smooth image. At different levels of resolution we allow the smoothing parameters to be different. This allows for differing degrees of smoothing at different scales. This and other issues related to the smoothing parameters are discussed in Nowak & Kolaczyk (2000) in detail, and their suggestions are quite relevant to our procedure as well. Timmermann & Nowak (1999) discuss a similar model for image restoration, in which a mixture of beta distributions is used as the prior for a one-dimensional multiscale analysis similar to ours. This would correspond to using a mixture of Dirichlet distributions. In this mixture model, several distributions with varying sharpnesses of their central peaks are combined to give a distribution with substantial probability mass near an equal split, but also with substantial probability mass spread through the entire range of splits. Such a mixture could be incorporated into the EMC2 framework and is worth investigating, since it could have the effect of adapting the splits spatially to better account for sharp edges or smoother areas in the image.

Our prior distribution has an effect on the restoration analogous to the "complexity penalties" and "regularizers" used in, e.g., Willett & Nowak (2002), and in many other image analysis techniques; it also provides a measure of "shrinkage," in which the fitted pixels intensities are shrunk toward mean values of groups of pixels.

3.3. *The Hyperprior Distribution*

The complete model, with components as described in equations (4), (5), (6), (9), and (10), is dependent on the values of $\boldsymbol{A}$, the smoothing parameters. In order to compensate for our uncertainty about the proper values for the smoothing parameters $\boldsymbol{A}$, we add a common hyperprior distribution for these parameters. The choice of smoothing parameter density has a somewhat paradoxical effect on the fitted image: In order to specify a less informative prior distribution on the image restoration, i.e., to let the data influence the outcome more with less regularization, we must put a more informative prior distribution on the smoothing parameters, restricting them to smaller numerical values. In particular, less informative Dirichlet prior distributions correspond to smaller values of α_k; thus, we would like a hyperprior distribution that prefers smaller values of α_k. In fact, without a hyperprior distribution that sharply tapers off away from zero, the fitted values of the smoothing parameters are so large as to favor nearly uniformly smooth reconstructed images.

We have considered several densities for the hyperprior distribution on the smoothing parameters. The best choices are a compromise between two opposing forces. We wish for the image not to be too smooth, in order to retain as much of the sharp detail as possible. Hence, most of the probability mass for the hyperprior distribution should be near zero, preferably encouraging or at least allowing arbitrarily small positive values for these smoothing parameters. On the other hand, when there is too much probability near zero in the hyperprior distribution, we have encountered numerical problems, in which a sampled value for a smoothing parameter is numerically zero. When this situation occurs, the Markov chain enters a persistent state in which the intensity in any cells with zero counts will be zero. We know this to be wrong, so we avoid zero-valued draws for the smoothing parameter. This means choosing a density whose probability is not too concentrated near zero. We have had the best empirical results with the density

$$p(\alpha_k) \propto \exp\left(-\delta\alpha^3/3\right). \tag{11}$$

Several possible choices for $p(\alpha_k)$ are shown in Figure 3, including the one in equation (11).

[4] The Dirichlet distribution is a continuous probability distribution on a vector of positive values, such that the vector sums to 1. The probability density function is given by $p(\boldsymbol{x}|\alpha) = \Gamma(\alpha_1 + \ldots + \alpha_k)x_1^{\alpha_1 - 1} \cdots x_k^{\alpha_k - 1}/\Gamma(\alpha_1) \cdots \Gamma(\alpha_k)$. It is a multidimensional generalization of the beta distribution, and is convenient for modeling a set of probabilities.

[5] It has been shown by Morris (1983) that the worst possible error resulting from using the wrong prior distribution is minimized by choosing the Dirichlet prior for a multinomial likelihood. Since we do not have any specific information here to guide us to a different choice of prior distribution, we reduce the bias which might be caused by a bad choice of prior distribution by choosing the Dirichlet distribution.

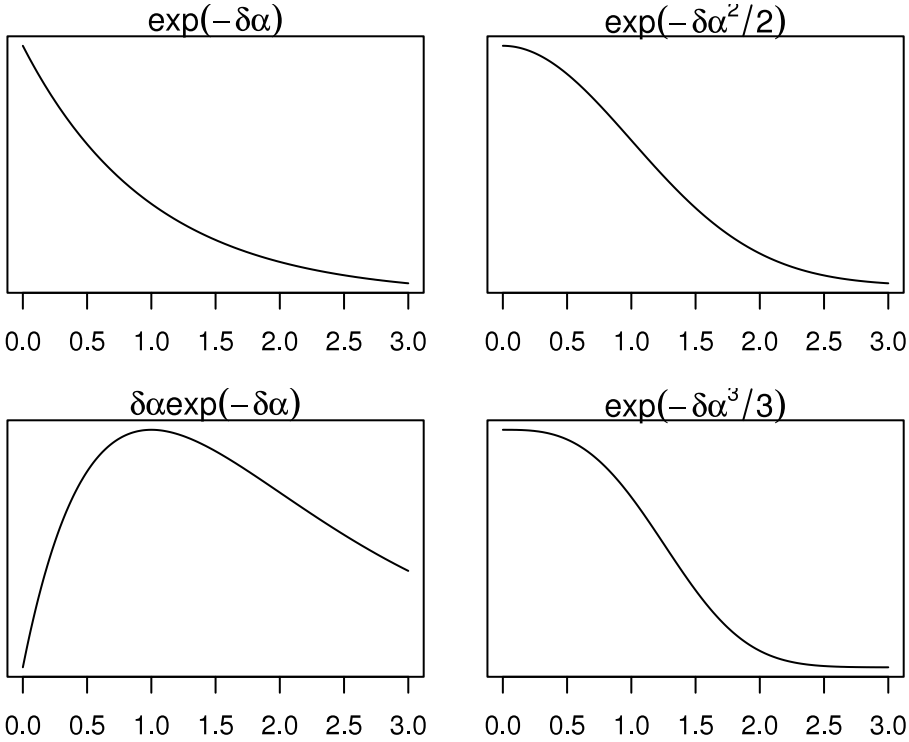

FIG. 3.—Unnormalized densities for the smoothing hyperprior. The EMC2 procedure currently uses the density proportional to the function in the lower right corner, for reasons explained in the text. For simpler comparisons, the scale parameter δ has been set to unity in all four panels.

We have so far described the components of the statistical model. We now proceed to fit the model, in order to generate the restored image, which is represented by the fitted values of Λ.

3.4. *Fitting the Model With Markov Chain Monte Carlo*

Many image restoration techniques compute a value of Λ to optimize a particular objective function. Such fitting methods produce a single restoration, often called the "best fit," e.g., an MLE or a MAP estimate. The best fit may not be a useful summary of the objective function. Indeed, it is the multimodal character of the likelihood function that encourages users to stop the Richardson-Lucy algorithm before it converges. A single best-fit reconstruction cannot be expected to adequately summarize the multimodel or highly skewed form that typical objective functions take in high-dimensional low-count problems. (A highly skewed objective function is illustrated in Fig. 1.) Standard techniques for computing error bars for best-fit estimates are also not well suited to such ill-behaved objective functions. Such techniques generally use second derivative matrices evaluated at the best fit and are based on Gaussian approximations to the objective function.

Because of the difficulties inherent in optimization-based fits, we apply a Monte Carlo technique to explore the entire posterior distribution, rather than attempting to summarize the distribution by a mode. Upon convergence, such algorithms create a chain of draws from the posterior distribution, which can subsequently be used in various ways to quantify the uncertainty of the restoration.

3.4.1. *Cycle Spinning*

In this section we address issues associated with the multiscale representation of the image and the statistical model that we propose. For example, in the restored image the multiscale representation has a tendency to produce artifacts in the form of checkerboard-like patterns. This is because some of the pixels are closer than others to the edges of the coarser resolution splits. To remedy this, we shift the multiscale representation to a different location in the image at each iteration. The shifting can be done systematically or randomly. Since a systematic coverage of a typical image would require many more iterations than possible in a typical run of EMC2, we randomly select a new origin at each iteration of the procedure. The general technique of shifting the multiscale basis is known as cycle spinning and is described in detail by Coifman & Donoho (1995). Cycle spinning can be obtained as a principled fitting algorithm under a minor modification of the statistical model: for an N by N image there are N^2 possible prior distributions as described in § 3, each centered at a different lattice point in the image. The prior distribution that results in randomly ordered cycle spinning is an equal-weighted mixture of all of these models. This modification to the statistical model adds the desirable property of translation invariance to the image restoration. Thus, cycle spinning is not simply an ad hoc procedure but a principled model-based approach to adding translation invariance and reducing visible artifacts in the image restoration.

3.4.2. *Convergence*

Our algorithm constructs a Markov chain with stationary distribution equal to the posterior distribution of interest. The Markov chain must be initiated at some starting value and may require a significant number of iterations before delivering draws from the posterior distribution. Once such convergence is achieved, the Markov chain must be run long enough to obtain a Monte Carlo sample that is sufficiently large for Monte Carlo evaluation of the posterior distribution. Since the starting values for the Markov chain are not necessarily a

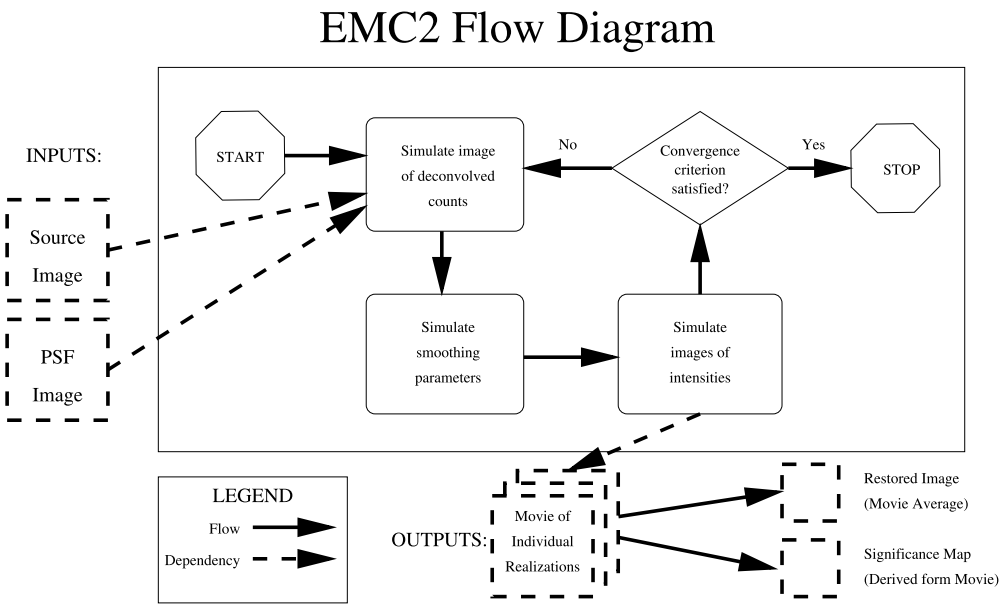

FIG. 4.—EMC2 flow diagram.

representative sample from the posterior distribution under the statistical model, we discard some iterations from the beginning of the chain. To minimize the potential for making inference based on samples that are not from the correct distribution, we suggest implementing a formal procedure to assess convergence. One such procedure uses two or more parallel chains of draws with different starting values. Specifically, we use $\hat{R}$ from Gelman & Rubin (1992) to assess convergence of the chains. This statistic compares the total variability of the samples from all of the chains to the sum of the variabilities within each separate chain. The ratio of these two variabilities is computed for one or more scalar quantities, e.g., parameters. Since the chains should coincide upon convergence, this ratio should tend toward one as the chains converge. We thus declare convergence when the $\hat{R}$ statistics for all of the quantities of interest are close to 1, e.g., below 1.1, or some other suitable criterion. A more thorough explanation of the use of $\hat{R}$ to assess convergence of MCMC samples used in spectral analysis is given in van Dyk et al. (2001).

Since each step of each chain is relatively expensive, we compute $\hat{R}$ using just two chains. We can keep track of the value of $\hat{R}$ for all of the smoothing parameters, as well as some or all of the pixel intensities. We declare convergence when the maximum $\hat{R}$ reaches a suitably low value, near 1.0, e.g., 1.07, a more stringent criterion than the 1.1 recommended in Gelman & Rubin (1992). We have found it necessary to require a stronger criterion, because the qualitative features of the image can still be changing when $\hat{R}$ is equal to 1.1 for some pixel intensities. After we declare convergence in the chains, it remains to run the chains until the Monte Carlo error has been reduced to a suitably low (user-specified) level.

The examples shown in § 4 were run for 500 burn-in iterations followed by 2000 iterations used for the restorations. This is conservative and was done to ensure proper convergence and low Monte Carlo error. With properly calibrated convergence diagnostics as described here, fewer iterations are required, and thus the restoration could be performed in a shorter time.

3.5. *EMC2 Flow Diagram and Outputs*

A flow diagram of the EMC2 technique is illustrated in Figure 4. The steps of the algorithm are as follows:

1. We start with the array of the observed counts and locations, and the PSF obtained from observations of point sources or using a simulator of the imaging system (e.g., the telescope and detectors). We form initial values for the intensities either with a uniform image or the raw counts with a small positive constant added to avoid zero-intensity pixels. For the purpose of convergence testing, we use both of these initial values to start two separate Markov chains.
2. For each of the two chains, we recreate the missing image true counts, according to the following probabilities: If an event is observed in pixel j, the probability that it originated from pixel i is $\Lambda_i P_{ij} / \sum_k \Lambda_k P_{kj}$. This step samples $\boldsymbol{X}$ given $\boldsymbol{Y}$ and Λ from its conditional posterior distribution, calculated under the model given in equations (5) and (6), and is represented by the first box in the flow diagram in Figure 4.
3. We then convert the two ideal images (i.e., $\boldsymbol{X}$) to their multiscale representations. Each of these multiscale representations is centered randomly on the image. Then we draw the smoothing parameters at each scale for each of the two chains from their conditional posterior distributions, given $\boldsymbol{X}$ and $\boldsymbol{Y}$. These are the $\boldsymbol{A} = \{\alpha_k,\ k = 1, \ldots, K\}$ parameters that appear in the prior distributions in equation (10). The hyperprior, shown in equation (11), is imposed on these parameters. This step is represented by the second box in the flow diagram.
4. Then we draw images of intensities Λ in the same multiscale representations from their conditional posterior distributions, given $\boldsymbol{X}$ and $\boldsymbol{A}$, the smoothing parameters. This step samples the parameters from the prior distributions shown in equations (9) and (10). Their conditional posterior distributions are also gamma and Dirichlet. The parameters for the gamma posterior distribution are γ_0 plus the total counts and $\gamma_1 + 1$; the parameters of the Dirichlet posterior distributions at the k^{th} scale are α_k plus the corresponding multiscale count totals from the ideal images. This step is shown in the third box in the flow diagram.

5. We then test for convergence using a method described in Gelman & Rubin (1992). Essentially, a formal hypothesis test is conducted to detect differences in means between the two chains. If detected, differences signify lack of convergence, since the two chains should converge to the same distributions and hence the same means. If convergence has not been achieved, we repeat the cycle (steps 2–4) with the updated intensities. If convergence has been achieved, we save the intensities in a file and then check whether we have enough post-convergence iterations to reduce the Monte Carlo error to a desirably small level. If so, we stop, and if not, we repeat the cycle (steps 2–4).

The primary output of EMC2, after convergence, is a sequence of images of intensities produced in each iteration. This sequence of images is a representative correlated sample from the posterior distribution of the true image and thus is the basis for inference under the statistical model. We can form a principal estimate of the true image by averaging all of the images in this sequence. This is a Monte Carlo approximation of the posterior mean.

To compile a significance map, we calculate the standard deviation of each pixel intensity and rescale the averaged image, dividing by the standard deviations of the pixels. In the resulting significance map, each pixel value is a test statistic for the hypothesis that the pixel intensities are zero under the assumption that those intensities are Gaussian; for high pixel intensities these test statistics make sense. However, because of the Poisson nature of the observed counts, the test statistics are less meaningful in low-intensity areas of the image, which are precisely the areas where questions of significance apply. A more useful test statistic would measure the probability that the intensity in the area in question does not arise from instrumental background. This probability can be calculated as the proportion of draws whose intensities are greater than a background intensity estimate. The background intensity might also be a sample from a procedure similar to that of EMC2, but without the PSF, since background files are normally provided as images of counts and are typically quite low-count, with many pixels containing zero counts. This area is currently under research and will be incorporated into EMC2 at a later date.

The output images can be viewed in a movie, which by itself provides much information to the researcher about the restoration. Some pixels in the movie are persistent, whereas others flicker between very low and moderately low intensities. The persistent pixels are thus more statistically significant, because under the model we are more certain that their intensities are not near zero, whereas the flickering pixels have some posterior probability of near-zero intensity under the statistical model.[6]

The three-dimensional array of output images can be the basis for many other types of inference. For any image feature that can be assigned a quantitative summary we can calculate a posterior predictive *P*-value (PPP-value) to assess statistical significance (Protassov et al. 2002; Gelman et al. 1996). Regions of the image could be defined and compared with other regions to extract information regarding observed features in the restoration.

Since the model generally conserves total flux, it should be noted that pairs of pixels in the image generally tend to correlate negatively, especially when they are located nearby. This can be seen by the splitting mechanism: If intensity is increased in one pixel, it must be decreased in an adjacent pixel to preserve the sum. This correlation structure is important in making principled inferences but is also very difficult to summarize in a concise and informative manner.

The movie can also be sorted pixel-by-pixel, providing a useful summary of each pixel's marginal posterior density. Sorting pixel intensities in this way ignores the between-pixel correlation and thus should not be used to make inference about features. However, it provides a useful summary of each pixel's marginal distribution, thus showing areas of the image in which the posterior distribution of the intensity is concentrated on large or small values.

4. EXAMPLES

4.1. *Simulation of an Extended Source image*

Figure 5 shows a simulation from a familiar image of Abraham Lincoln. The image was first digitized and reduced to four levels of gray, as shown in Figure 1*a*. The colors were coded from zero to three, shown ranging from light to dark in this figure; each pixel was then considered to contain that number of counts. Each count was then displaced randomly from its original location according to a *Chandra* PSF, shown in Figure 1*b*, to obtain the simulated data set shown in Figure 1*c*. The EMC2 restoration of this image is shown in Figure 1*d*. These images are shown on the log scale but with different color maps to illustrate features in the images. EMC2 does indeed preserve flux, despite the fact Figure 1*a* is darker than the data set in Figure 1*c* or the restorations in Figures 1*d* and 1*f*. The statistical model has been designed to preserve total flux, and this preservation has been empirically verified. The restoration is quite good: Much of the signal has been extracted from the noise, and the smooth areas match those of the original image fairly well in intensity. The sharp edges of the ideal image are not well restored. Considering that EMC2 is designed to smooth extended areas of the image, and that Nyquist-like considerations would lead one to expect a degradation in resolution; however, the results are still very good. Figure 1*e* shows the significance map, thresholded at 1 standard deviation. The signifance map is quite similar in appearance to the restoration; pixels with higher intensity in the restoration tend to have larger values in the significance map. This is characteristic of Poisson distributions, whose standard deviations are equal to the square roots of their means. In Figure 1*f*, we see the Richardson-Lucy restoration of the simulation, stopped at 100 iterations. The comparison of this restoration with the EMC2 restoration shows one of the relative strengths of EMC2—the ability to restore extended smooth material. The Richardson-Lucy restoration aggregates the intensity into clumps and the smooth material in the ideal image shows as grainy material in the restoration. EMC2 does a better job at restoring the smooth areas, although the sharpness of the borders is necessarily sacrificed in order to obtain this smoothness.

Figure 6 shows four different termination points of the Richardson-Lucy algorithm on the same image and PSF. We can see that the proper choice of termination point involves a compromise between smoothing, which is seen somewhat in Figure 1*a* at 10 iterations, and resolving features in the image. Despite the smoothing obtained by stopping the algorithm at 10 iterations, the restoration already has a grainy character and the features are not yet fully resolved. Later in the algorithm,

[6] Several examples of movies are posted at http://davidesch.150m.com/EMC2movies.html.

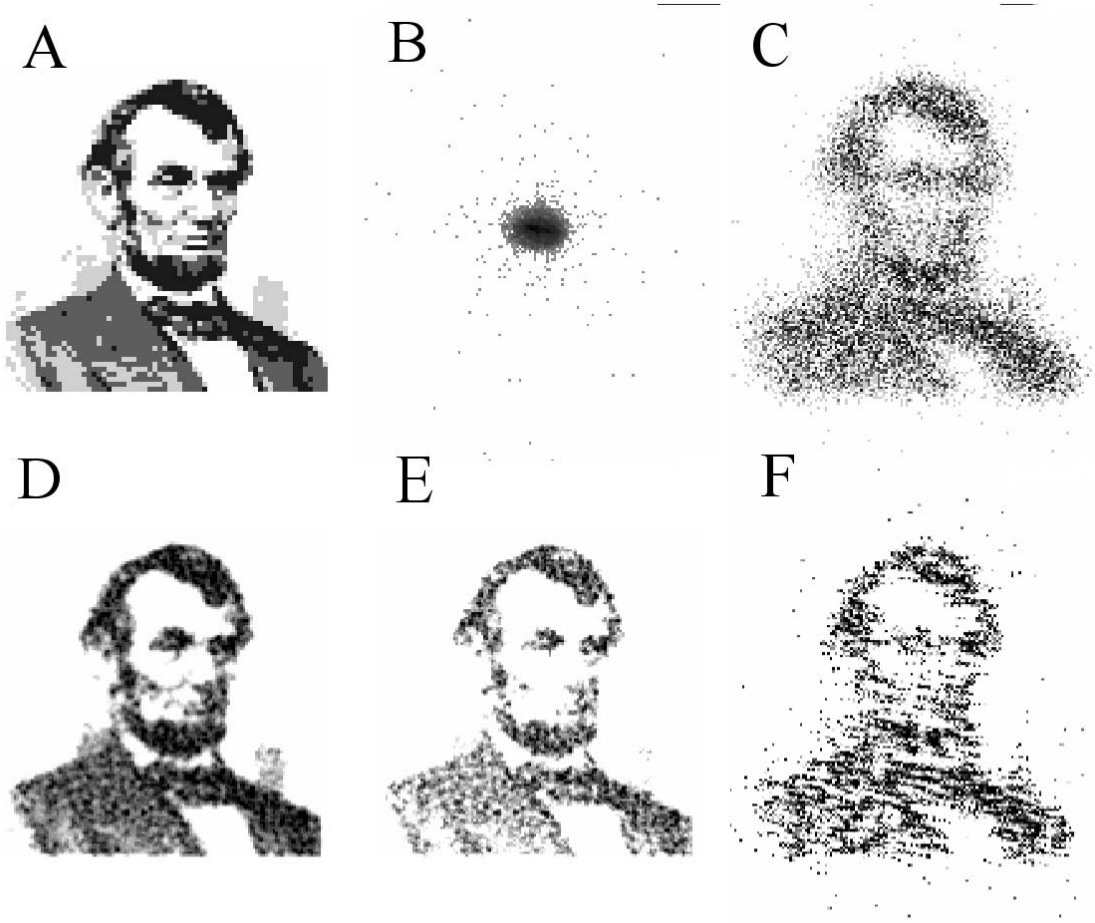

FIG. 5.—Simulation of an extended source. (A) the original image used for the simulation; (B) the PSF; (C) the convolved simulation used in the analyses; (D) the EMC2 primary reconstruction; (E) the EMC2 significance map, thresholded at one standard deviation; and (F) the Richardson-Lucy restoration, stopped at 100 iterations. The EMC2 reconstruction was performed with 500 burn-in iterations and 2000 subsequent iterations were used for the restoration.

as the procedure nears convergence, the features are well resolved, but the extended material is eroded into grainy material. In Figures 1*b*, 1*c*, and 1*d*, at 20, 50, and 100 iterations, there is a steady increase in the resolution of sharp features and a steady decline in the smoothness of extended material. Because of this compromise, we feel that the restorations from all four termination points shown here are inferior to the EMC2 restoration shown in Figure 5.

Because the Richardson-Lucy algorithm can take advantage of the fast Fourier transform (FFT), each iteration of Richarson-Lucy is considerably less computationally complex than an iteration of EMC2. EMC2 would have complexity on the same order as an implementation of Richardson-Lucy, which used the sliding-cell convolution method rather than the FFT. We are currently researching ways to decrease the complexity of the EMC2 algorithm and improve its speed and scalability to larger images. It should also be remembered that iterations have different meanings in the two algorithms. Each iteration of EMC2 draws another sample from the posterior, whereas each iteration of Richardson-Lucy moves the restoration one step closer to the MLE.

4.2. *Restoration of Chandra Images*

In this section we present several examples of applying the EMC2 technique to low-count statistics astronomical images. We first show an example of a simulated *Chandra* observation of a binary source and then an example of a more complex extended source containing multiscale structures.

Chandra produces sharper images than any other X-ray telescope to date (FWHM$\sim$0.″3 on axis). This high-spatial resolution results from the innovative design of this observatory, in particular the high-resolution mirror assembly (Van Speybroeck et al. 1997), but also including the guidance systems and the focal plane detectors (Weisskopf et al. 2002). In addition to the photon positions, *Chandra* observations provide information about their number, energy, and time of arrival. As a result, *Chandra* provides a unique opportunity for high spatial/spectral resolution studies of the X-ray emission, at scales ranging from sub-arcsecond to several arcminutes.

Although *Chandra* images of numerous sources reveal multiscale structures with unprecedented detail and clarity, it is often difficult to study small-scale structures because of the low count statistics and the image blurring by the off-axis HRMA (High Resolution Mirror Assembly) PSF. The size and the shape of the PSF changes significantly as a function of off-axis angle and the spectral energy distribution of the source. Simulating the *Chandra* PSF as a function of off-axis angle and energy for each observed source is the first step in multiscale analysis of *Chandra* images (Karovska et al. 2001).

The examples presented in this paper are based on simulations using the *Chandra* PSF simulator ChaRT/MARX

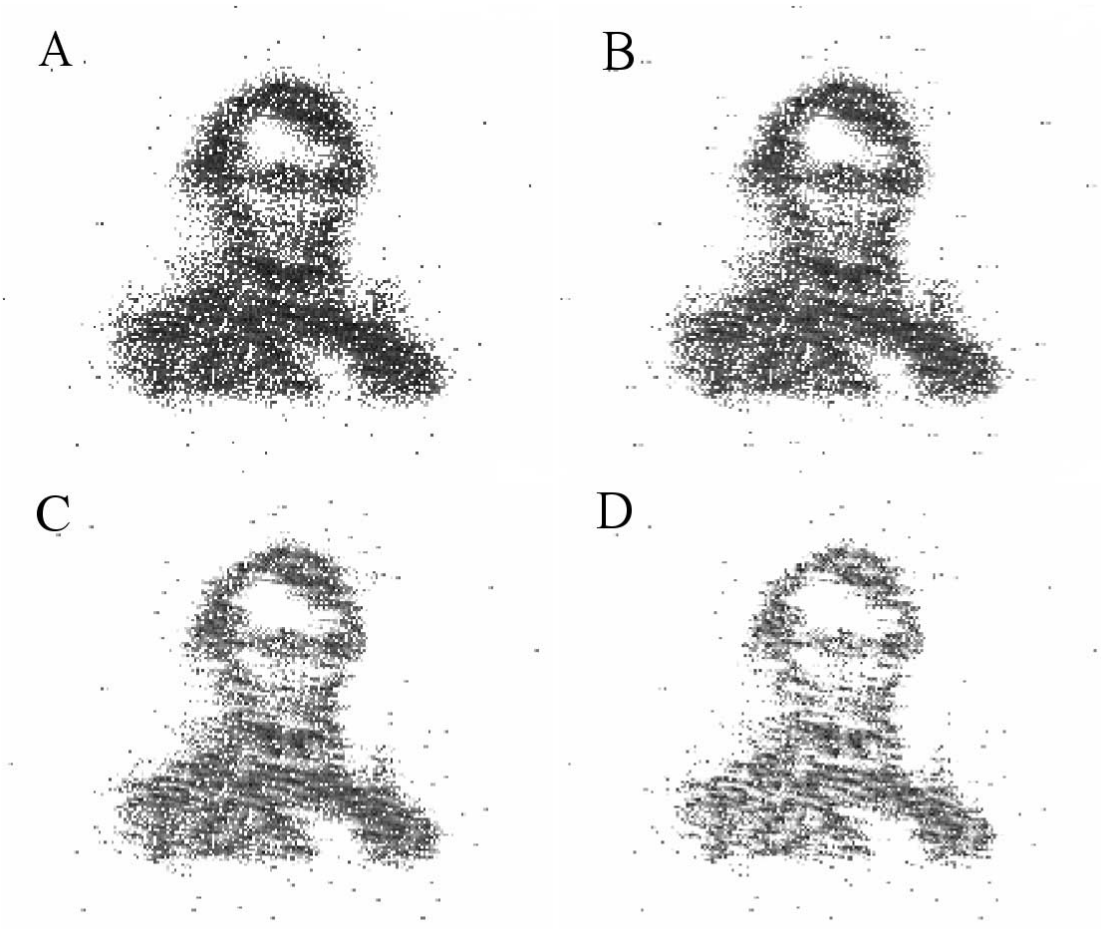

FIG. 6.—Comparing termination points of the Richardson-Lucy deconvolution. The Algorithm was stopped at (*a*) 10, (*b*) 20, (*c*) 50, and (*d*) 100 iterations.

(Carter et al. 2003) and archival *Chandra* observations made with the imaging detectors: the Advanced CCD Imaging Spectrometer (ACIS-S) and the High-Resolution Camera (HRC-I).

In Figure 7 we show a simulated *Chandra* observation using the HRC detector ($0\farcs13$ pixel^{-1}) of a binary source with pointlike components separated by $\sim 2''$ (16 detector pixels) at $5'$ off-axis angle. The brightness ratio between the components is 10:1, with 5000 counts for the source on the left and 500 counts for the source on the right. To obtain this image, we convolved these two point sources with a blurring function composed of the *Chandra* telescope off-axis PSF and the detector blurring due to the pixel size of $0\farcs13$. The PSF was simulated using the *Chandra* Ray Tracer, ChaRT (Carter et al. 2003) and MARX (Wise et al. 1997) software that simulates the effects of the detectors. Figure 7*b* shows the simulated PSF.

The EMC2 restorations of the source and significance maps are shown in Figures 7*c*, 7*d*, 7*e*, and 7*f* . Two sets of restorations are shown, both with hyperprior density function proportional to $\exp(-\delta\alpha^3/3)$. Figures 7*c* and 7*d* have $\delta = 10$, and Figures 7*e* and 7*f* have $\delta = 1000$. These restorations are intended to show the effect of the hyperprior distribution on the outcome of the restoration. We can see clearly that the $\delta = 1000$ restorations are preferable, because the $\delta = 10$ produces too much smoothing in the restoration. Otherwise, the results are good, and the simulation noise in the data makes it impossible for the intensity in the restored image to shrink down to 2 pixels. We have shown the restored images on the log scale to illustrate these "leftovers," but the pixels adjacent to the points have restored intensity values far less than the two points themselves. On the linear scale the images appear as two lit pixels against a black background. The significance maps in Figures 7*d* and 7*f* again show patterns very similar to the restorations. The hyperprior using $\delta = 1000$ produced good results, and it is used in all of the other EMC2 restorations shown in the paper.

In our second example we show the results of applying the EMC2 technique to a *Chandra* image of an extended source containing multiscale structures. NGC 6240 is a starburst galaxy at a distance of 144 Mpc containing two supermassive black holes separated by $\sim 2''$ ($\sim$ 1000 pc). This active galaxy has been observed using various multiwavelength techniques. High spatial and spectral resolution X-ray studies of the central region surrounding the black holes are very important for understanding the phenomena associated with post-merger systems evolution and nuclear activity (Lira et al. 2002; Komossa 2003).

In Figure 8 the EMC2 posterior mean reconstruction of the NGC 6240 central region is shown with the original data. Figures 8*c* and 8*d* are the significance maps, thresholded at 1 and 3 standard deviations, respectively. The EMC2 reconstruction shows multiscale structures including diffuse emission features and subarcsecond scale structures associated with the central sources. Clearly visible are several arcs and looplike structures that appear to be originating from the

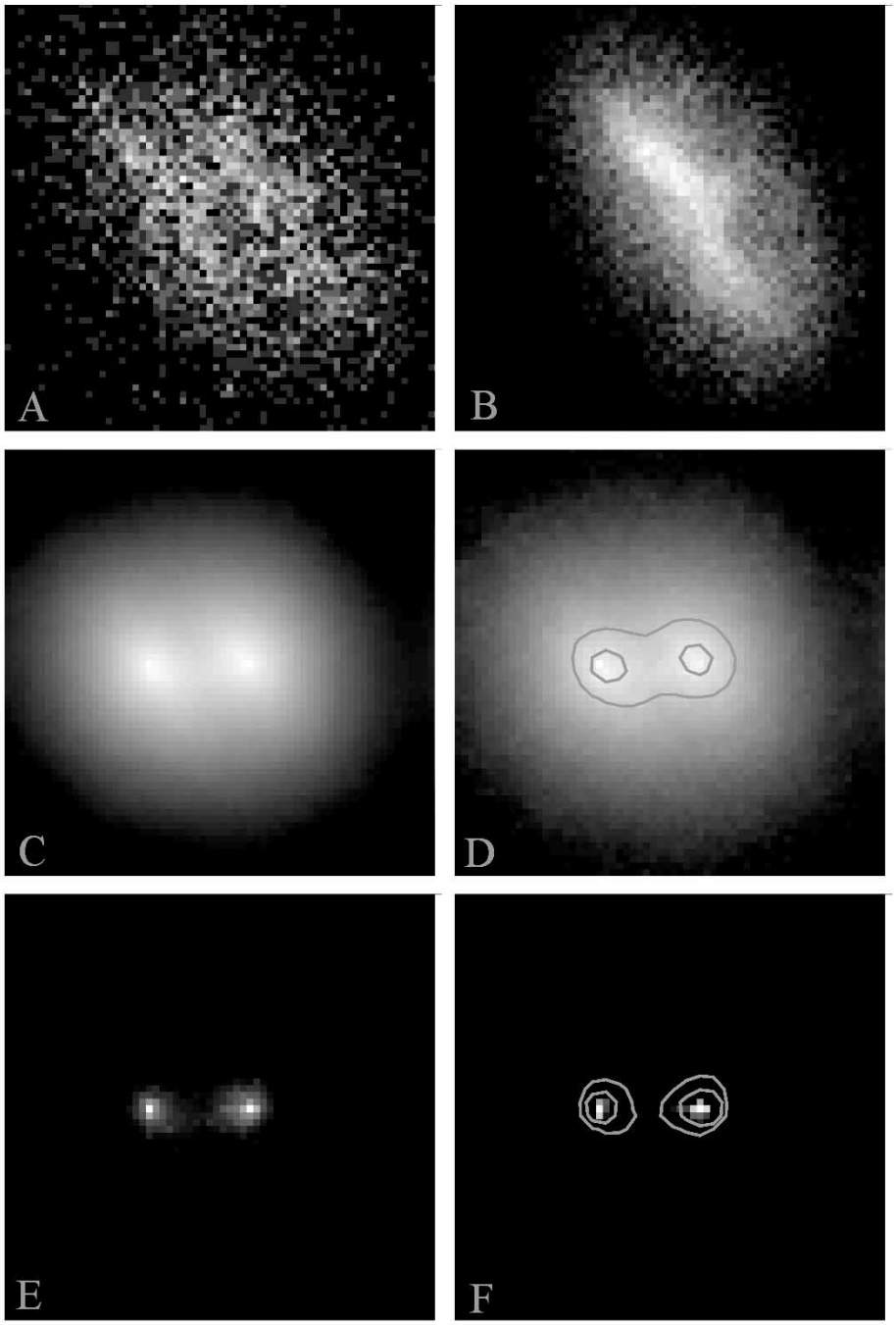

FIG. 7.—Shown is a simulated example. (*a*) The data and (*b*) the PSF are shown with the color map on the square root scale to best illustrate the variability of the counts in these images. EMC2 was run using two priors on the smoothing parameters, $p(\alpha) \propto \exp(-\delta\alpha^3/3)$ with $\delta = 10$ in (*c*) and (*d*) and $\delta = 1000$ in (*e*) and (*f*). The reconstructions under the two prior distributions appear in (*c*) and (*e*). The corresponding significance maps appear in (*d*) and (*f*). The significance maps illustrate the posterior mean divided by the posterior standard deviation. The contour lines in (*d*) and (*f*) correspond to levels 3.0 and 10.0, so we are quite confident that the nature of the source is binary. The EMC2 reconstructions were performed with 500 burn-in iterations and 2000 subsequent iterations used in each of the restorations.

vicinity of the nuclei. These structures are not discernible in the original images or in adaptively smoothed images. EMC2 sharpens these structures while preserving the extended diffuse emission and reveals more details in the fine-scale structure in the vicinity of the nuclei. According to the significance map, the features in the restored image, including the loops and the structures in the vicinity of the nuclear region, are statistically significant. Figure 9 shows two two-color overlays of the optical emission recorded by the *Hubble Space Telescope* (*HST*) and X-ray images from *Chandra*. The first panel shows the raw X-ray emission, and the second shows the EMC2 reconstruction of the X-ray emission. The correspondence between the two wavelengths is much stronger for the EMC2 reconstruction. The X-ray observations provide valuable scientific information on the content of the various components of the source, and by comparing these parts of the image with other frequencies much can be learned about the object.

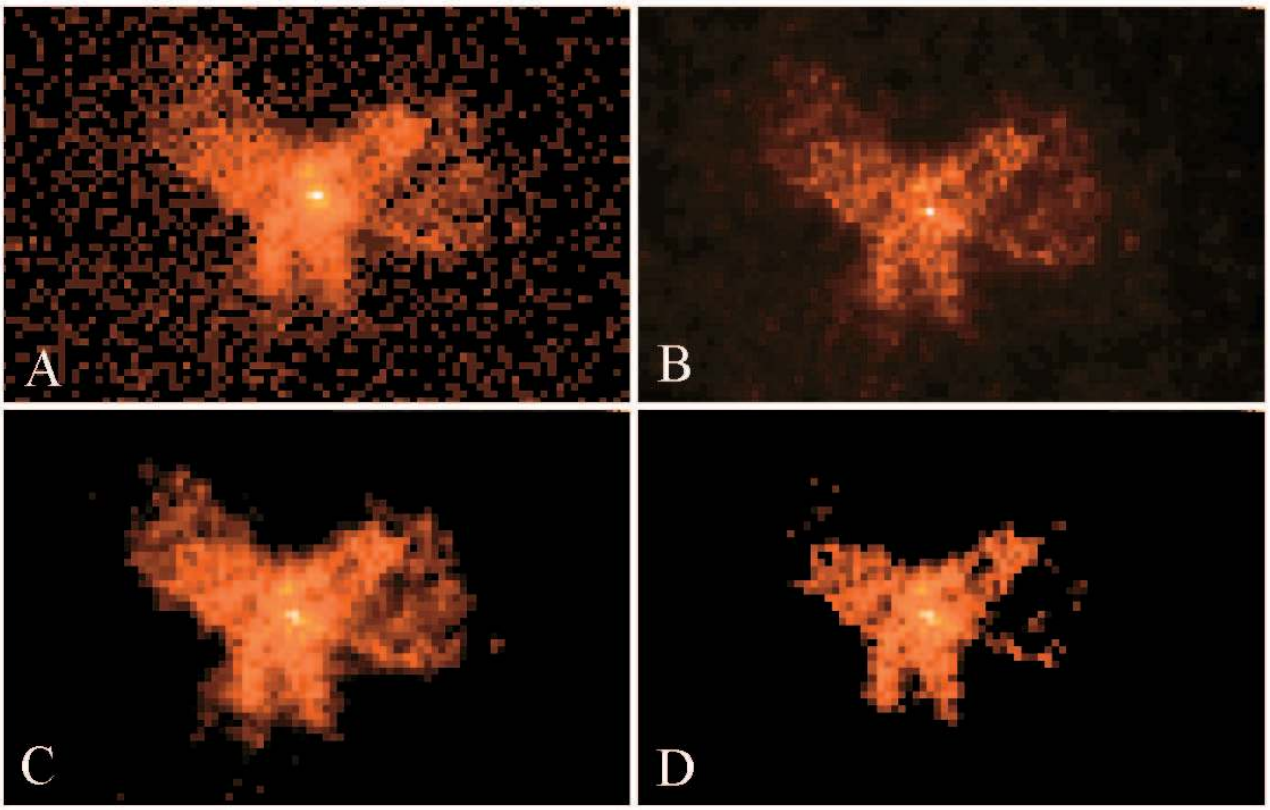

FIG. 8.—EMC2 restoration and significance maps for the *Chandra* image of NGC 6240. (*a*) The original data set as recorded by the telescope. (*b*) The EMC2 reconstruction. (*c*, *d*) The significance map, thresholded at 1 and 3 standard deviations, respectively. The EMC2 reconstruction was performed with 500 burn-in iterations and 2000 subsequent iterations used for the restoration.

For comparison with the EMC2 results we show in Figure 10 the results from Richardson-Lucy deconvolution (20 and 100 iterations) of the NGC 6240 images. The Richardson-Lucy results tend to aggregate intensity into clumps and do not reconstruct well the extended and the low brightness structures.

5. CONCLUSIONS

This work brings several new and important features to a deconvolution analysis, the most important being the quantification of the uncertainty of the outcome. The analysis itself is based on principled methods and is performed entirely in one procedure, thus eliminating any error propagation that may occur by reprocessing data in multiple steps. The regularization technique works well and preserves the integrity of extended sources in the image.

The EMC2 technique can be thought of as an extension of two previous deconvolution techniques, the Richardson-Lucy and the Nowak-Kolaczyk techniques. Our method is related to

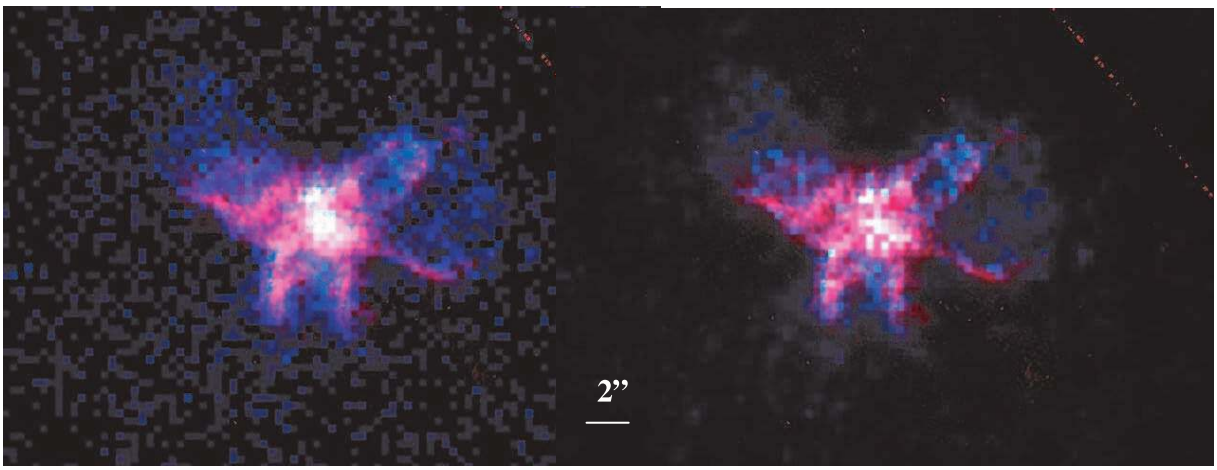

FIG. 9.—*HST* and *Chandra* images of NGC 6240 are compared. The figures shown are overlays of the *HST* optical frequency image with the raw *Chandra* data (*left*) and the EMC2 posterior mean restoration (*right*). Clearly there is more correlation between the wavelengths in the restored image than in the raw data, providing additional confidence in the EMC2 results.

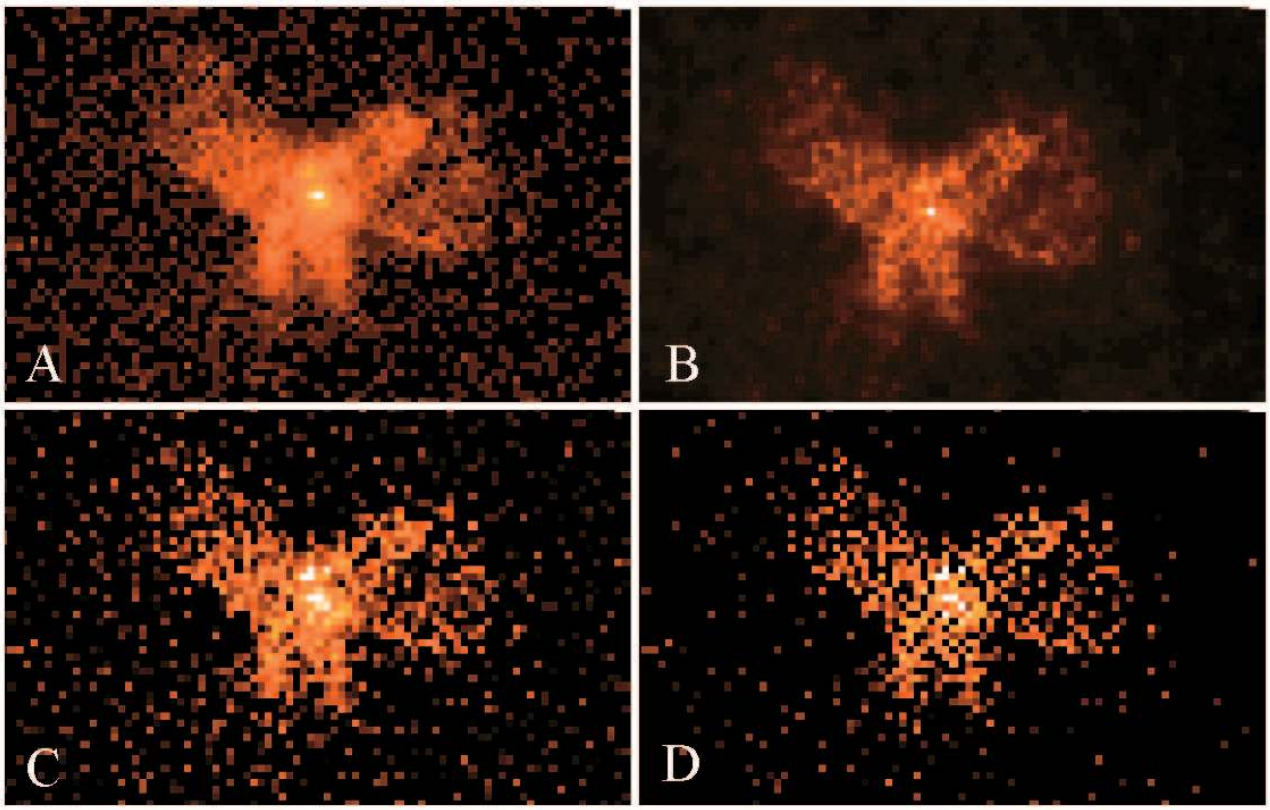

FIG. 10.—EMC2 and Richardson-Lucy reconstructions of the *Chandra* image of NGC 6240 are compared. (*a*) The raw data and (*b*) the posterior mean EMC2 restoration. The lower two panels are the Richardson-Lucy deconvolutions, stopped at (*c*) 20 and (*d*) 100 iterations, respectively. The Richardson-Lucy restorations clearly undersmooth the image, aggregating the intensity at "clumps" in the image.

these techniques in that the likelihood of the statistical model is the same, and the prior is the same as the Nowak-Kolaczyk technique but departs from them in the inclusion of the hyperprior distribution to fit the smoothing parameters and also in the sampling-based fitting method, which makes uncertainty information available to the user. We feel that both of these extensions are useful for investigators, the first greatly improving the usability of the procedure and the second in making scientific inferences based on observed image data.

Of course, the EMC2 procedure adds smoothing to the basic likelihood in the Richardson-Lucy procedure and thus models an image that is smooth. Although the smoothing can be minimized to enable sharper edges in the restoration, it must be remembered that the procedure is a smoothing procedure and should not necessarily be expected to perform well on images with sharp variations between adjacent pixels. For example, when applied to point sources, the procedure does not resolve the point source as well as a Richardson-Lucy restoration. Nevertheless, it does almost as well, and its vastly superior performance for extended emissions is enough to recommend this procedure for almost any restoration in which the user is uncertain as to the nature of the source being observed.

In the current implementation of EMC2, the PSF is a single matrix, representing the scattering of a point source from anywhere on the entire image; however, the procedure described here requires only that we be able to calculate probabilities that an event with any true location is observed in any other location on the detector. This probability can come from one PSF, many PSFs in a mosaic image, or an unknown PSF that can be described probabilistically and simulated from. Thus, the method is extendable to a broad range of deconvolution problems. The current implementation of EMC2 is limited to sources covering a small enough angle so that the PSF is nearly constant across the entire region. Multiple PSFs would greatly increase the utility of the procedure and allow restorations of much larger images.

Deconvolution of larger images also necessitates inclusion of background and instrumental effects into the statistical model, as described in § 3.1, since these effects tend to vary over a larger range of the detector. As discussed, the exposure map is an easy extension of the procedure. However, smoothing similar to the multiscale model or perhaps some other procedure that smooths only on a smaller scale will be necessary to incorporate the instrumental background. This is a suitable area for more research.

Further extension of this technique will include incorporating more prior information. For example, other components of the image could also be specified and included in the analysis. This would amount to adding a step to the MCMC algorithm that splits the true counts according to the current intensities of each of the components for each pixel. Often astronomers have some prior information about the observed source or the observation itself, which would be reasonable to include in an image analysis. For example, the user could include information about point sources in the image at known locations. Stronger prior information from other observations could also be fitted in the statistical model. This information would consist in the model as another additive component of the pixel intensities Λ. The capacity to use previously available information would therefore increase the power of the EMC2 procedure to detect certain types of sources.

Other procedures have been proposed to better represent sharp lines in the image, by introducing partitions that do not split along the pixel grid (see, e.g., Willett & Nowak 2002). Using this type of procedure involves selecting among many different models at the same time as producing an MLE of the parameters of the final model. While this type of procedure may be desirable for producing a more parsimonious representation of the image intensities as well as better representing sharp diagonal features, our procedure has been designed with the principal goal of producing uncertainty information on the image in a square pixel framework. Sharper features and better representations of diagonal features can also be obtained by using a subpixelated ideal image, taking advantage of greater knowledge of the PSF. In principle, it is possible, and perhaps desirable in the future, to include some model selection and more flexible model components such as Willet's wedgelets in our model, but the computational burdens of such an algorithm might be too great to fit in a reasonable amount of time with today's computing resources.

Research was funded in part by NSF grant DMS 01-04129 and in part by NASA contract NAS8-39073 (*Chandra* X-Ray Center), as well as NASA AISRP contract NCC2-1350: Advanced Statistical and Data Analysis Tools for Astrophysics (ASDATA).

We are grateful to an anonymous referee whose comments have served to greatly improve our presentation. David Esch would like to thank professors Xiao-Li Meng, Carl Morris, and Donald Rubin of the Harvard University Department of Statistics, and Eric Kolaczyk of the Boston University Department of Mathematics and Statistics, for gentle encouragement over this long project, which grew out of his doctoral dissertation (Esch 2003). Alanna Connors gratefully acknowledges the support of her AISRP contract (NCC2-1350) with Jeff Scargle and Vinay Kashyap, via Ames and Harvard: Advanced Statistical and Data Analysis Tools for Astrophysics (ASDATA), and would like to thank Wellesley College for hospitality. Margarita Karovska is a member of the *Chandra* X-Ray Center, which is operated by the Smithsonian Astrophysical Observatory under contract to NASA NAS8-39073.

Thanks are also due to the California-Harvard Astrostatistics Collaboration (http://www.ics.uci.edu/~dvd/astrostat.html), including but not limited to Aneta Siemiginowska, Vinay Kashyap, Andreas Zezas, and Peter Freeman of the Harvard-Smithsonian Center for Astrophysics, and Rostislav Protassov, Yaming Yu, Hosung Kang, and Taeyoung Park of the Harvard University Department of Statistics. Code that implements EMC2 is planned as part of a future release of the Chandra Interactive Analysis of Observations (CIAO) software suite, which is available at http://cxc.harvard.edu/ciao.

REFERENCES

Carter, C., Karovska, M., Jerius, D., Glotfelty, K., & Beikman, S. 2003, in ASP Conf. Ser. 295, Astronomical Data Analysis Software and Systems XII, ed. H. E. Payne, R. I. Jedrzejewski, & R. N. Hook (San Francisco: ASP), 477
Coifman, R. R., & Donoho, D. L. 1995, in Wavelets and Statistics, ed. A. Antoniadis & G. Oppenheim (New York: Springer), 125
Dempster, A. P., Laird, N. M., & Rubin, D. B. 1977, J. Royal Stat. Soc. B, 39, 1
Esch, D. 2003, Ph.D. thesis, Harvard Univ.
Frieden, B. R. 1972, J. Opt. Soc. Am., 62, 511
Gelman, A., Meng, X. L., & Stern, H. 1996, Stat. Sinica, 6, 733
Gelman, A., & Rubin, D. B. 1992, Stat. Sci., 7, 457
Gull, S. F., & Skilling, J. 1991, MemSys5: Quantified Maximum Entropy User's Manual (Royston: Maximum Entropy Data Consultants, Ltd.)
Hastings, W. K. 1970, Biometrika, 57, 97
Kaaresen, K. F. 1997, Proc. IEEE Trans. Signal Processing, 45, 1173
Karovska, M., et al. 2001, in ASP Conf. Ser. 238, Astronomical Data Analysis Software and Systems X, ed. F. R. Harnden, Jr., F. A. Primini, & H. E. Payne (San Francisco: ASP), 435
Knödelseder, J., et al., 1996, Proc. SPIE, 2806, 386
Kolaczyk, E. 1999, J. Am. Stat. Assoc., 94, 920
Kolaczyk, E. D., & Dixon, D. D. 2000, ApJ, 534, 490
Komossa, S. 2003, in AIP Conf. Proc. 686, The Astrophysics of Gravitational Wave Sources, ed. J. M. Centrella (New York: AIP), 161
Lira, P., Ward, M. J., Zezas, A., & Murray, S. S. 2002, MNRAS, 333, 709
Lucy, L. B. 1974, AJ, 79, 745
Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., & Teller, E. 1953, J. Chem. Phys., 21, 1087
Morris, C. N. 1983, Ann. Stat., 11, 515
Murtaugh, F., Starck, J.-L., & Bijaoui, A. 1995, A&AS, 112, 179
Nowak, R. D., & Kolaczyk, E. D. 2000, IEEE Trans. Inf. Theor., 46, 1811
Protassov, R., van Dyk, D. A., Connors, A., Kashyap, V., & Siemiginowska, A. 2002, ApJ, 571, 545
Richardson, W. H. 1972, J. Opt. Soc. Am., 62, 55
Skilling, J., & Bryan, R. K. 1984, MNRAS, 211, 111
Starck, J. L., & Murtagh, F. 2002, Astronomical Image and Data Analysis (Berlin: Springer)
Timmermann, K. E., & Nowak, R. D. 1999, IEEE Trans. Inf. Theor., 45, 846
van Dyk, D. A., Connors, A., Kashyap, V., & Siemiginowska, A. 2001, ApJ, 548, 224
———. 2003, in Statistical Challenges in Modern Astronomy III, ed. E. Feigelson & G. Babu (New York: Springer), 41
Van Speybroeck, L. P., Jerius, D., Edgar, R. J., Gaetz, T. J., Zhao, P., & Reid, P. B. 1997, Proc. SPIE, 3113, 89
Weisskopf, M. C., Brinkman, B., Canizares, C., Garmire, G., Murray, S., & Van Speybroeck, L. P. 2002, PASP, 114, 1
Willett, R., & Nowak, R. 2003, IEEE Trans. Medical Imaging, 22, 332
Wise, M., Huenemoerder, D. P., & Davis, J. E. 1997, in ASP Conf. Ser. 125, Astronomical Data Analysis Software and Systems VI, ed. G. Hunt & H. E. Payne (San Francisco: ASP), 477