



Joint Deconvolution of Astronomical Images in the Presence of Poisson Noise

Axel Donath¹, Aneta Siemiginowska¹, Vinay L. Kashyap¹, David A. van Dyk², and Douglas Burke¹

¹Center for Astrophysics | Harvard & Smithsonian, Cambridge, MA 02138, USA; axel.donath@cfa.harvard.edu

²Department of Mathematics, Imperial College London, London, SW7 2AZ, UK

Received 2024 March 15; revised 2024 July 24; accepted 2024 July 25; published 2024 September 30

Abstract

We present a new framework for joint likelihood deconvolution (*Jolideco*) of a set of astronomical observations of the same sky region in the presence of Poisson noise. The observations may be obtained from different instruments with different resolution, and different point-spread functions (PSFs). *Jolideco* reconstructs a single flux image by optimizing the posterior distribution based on the joint Poisson likelihood of all observations under one of several prior distributions. Most notably, we employ a patch-based image prior that is parameterized via a Gaussian mixture model, which we train on high-signal-to-noise astronomical images, including data from the JWST and the GLEAM radio survey. This prior favors correlation structures among the reconstructed pixel intensities that are characteristic of those observed in the training images. It is, however, not informative for the mean or scale of the reconstruction. By applying the method to simulated data, we show that the combination of multiple observations and the patch-based prior leads to much improved reconstruction quality in many different source scenarios and signal-to-noise regimes. We demonstrate that with the patch prior *Jolideco* yields superior reconstruction quality relative to alternative standard methods such as the Richardson–Lucy method. We illustrate the results of *Jolideco* applied to example data from the Chandra X-ray Observatory and the Fermi Gamma-ray Space Telescope. By comparing the measured width of a counts-based and the corresponding *Jolideco* flux profile of an X-ray filament in SNR 1E 0102.2–7219, we find the deconvolved width of $0.''58 \pm 0.''02$ to be consistent with the theoretical expectation derived from the known width of the PSF.

Unified Astronomy Thesaurus concepts: [Astronomy image processing \(2306\)](#); [Bayes' Theorem \(1924\)](#); [Astrostatistics tools \(1887\)](#); [Posterior distribution \(1926\)](#); [Gaussian mixture model \(1937\)](#); [Open source software \(1866\)](#)

1. Introduction

The resolution of any physical or astronomical imaging process is inherently limited by the instrument or telescope used. Furthermore, the quality of the resulting image is affected by background, instrumental measurement noise, and nonuniform exposure. This is especially true in the low-signal-to-noise regime, where images are affected by Poisson noise. This includes images taken by X-ray and gamma-ray telescopes but also medical images obtained with methods such as positron emission tomography or single-photon emission computed tomography. This low-count imaging process is illustrated in Figure 1.

To maximize the scientific use of available data it is desirable to correct for the instrumental imprint on the data as well as reduce its noise level. The process of improving the resolution of an image in post-processing is often called “deblurring” or “deconvolution,” as it aims to invert the effect of the data being degraded by a forward convolution through the point-spread function (PSF) of the instrument. In general, this class of problems is categorized as “inverse problems.” If there is prior information available on the shape of the PSF, we typically refer to the process more precisely as “unblind deconvolution.” In the case of astronomical observations the PSF can, for example, be estimated empirically from observation of known point sources, such as distant stars, or a model of

the PSF can be obtained from simulation of the telescope or instrument.

1.1. Related Work

In the literature there have been many efforts to solve the unblind deconvolution problem for images affected by Poisson noise. Most methods rely fundamentally on modeling image generation as a linear process. Given a known PSF a prediction of the discrete measured image can be obtained via “forward-folding” and be compared against an observation assuming Poisson statistics in each pixel. However, the methods differ in their additional prior or regularization assumptions as well as their approach to optimization or sampling of the underlying likelihood function or posterior distribution. In the following section, we give a short overview over these methods.

The standard baseline method for image deconvolution in the presence of Poisson noise was proposed independently by Richardson (1972) and Lucy (1974; hereafter the RL method). The method involves estimating the reconstructed image by an iterative procedure that converges to a maximum likelihood solution of the problem.

However, in practice the RL method has multiple, partly related limitations. First, when the number of iterations is too high, the method shows a tendency to decompose regions of constant brightness into a set of bright emission peaks (“speckles”). This problem was described and studied, for example, by Reeves (1995) and Fish et al. (1995). This means that in practice the RL method is not iterated until it converges, leading to the second limitation: there is no objective criterion for stopping the optimization process. Usually, the iteration is



Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](#). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

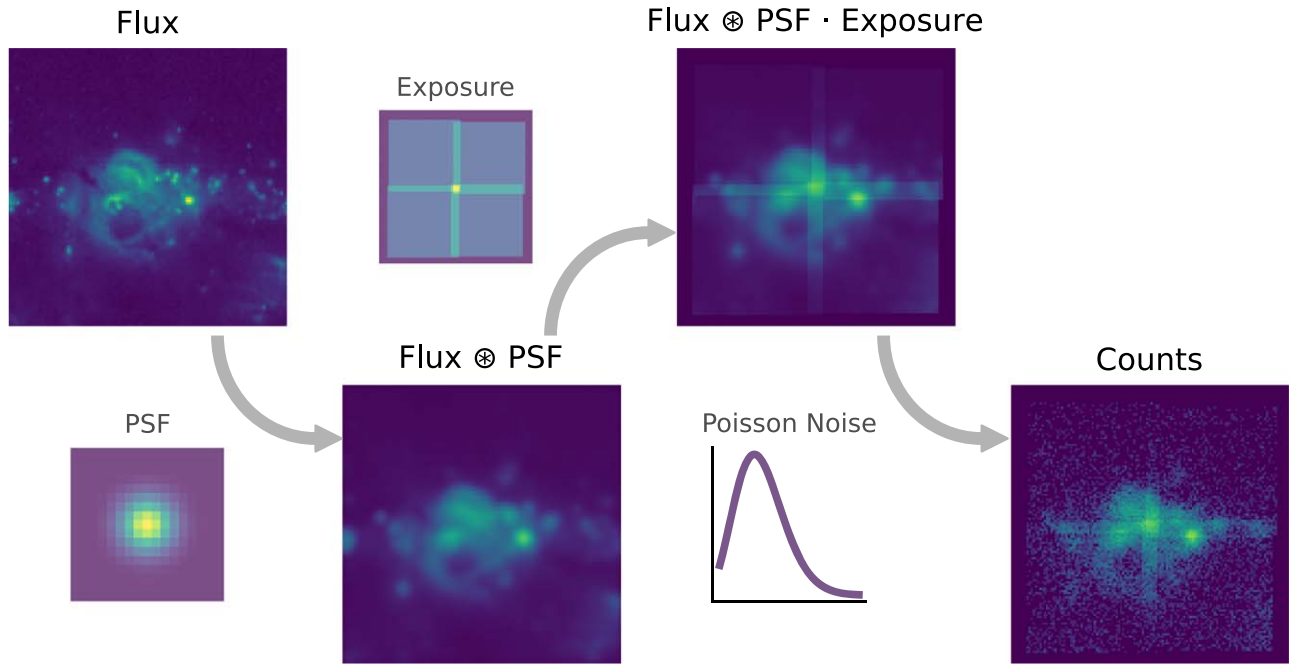


Figure 1. Illustration of the forward model of the low-count imaging process. The flux represents the true unknown-flux distribution in the sky. The flux is convolved with an image of the point-spread function (PSF), which represents an estimate of the resolution of the imaging instrument. The exposure consists of the superposition of multiple independent observations with different observation times and pointing positions. The exposure-weighted flux image is degraded by Poisson noise. 📡

stopped when the reconstructed image seems plausible to the eye. Lastly, the RL method only provided a “point estimate” and no associated estimate of the uncertainties of the reconstruction. Uncertainties are desirable especially in the Poisson domain, for example to measure the significance of specific features found in the reconstructed image.

To address the issue with the RL method decomposing extended emission structures into groups of point sources, multiple additions to the RL method have been proposed. Most importantly, this led to the introduction of regularized RL methods, such as use of total variation (Dey et al. 2006) or sparse representations based on “wavelets” or “curvelets” (Starck et al. 2003). The challenge is rooted in two fundamentally contradicting goals: a deconvolution process that favors point sources will decompose extended structures present in the image, while introducing local correlations between neighboring pixels will lead to less sharp reconstruction of point sources.

One possible approach to overcome this challenge is proposed by Esch et al. (2004) and Connors et al. (2011) with the EMC2 and LIRA methods. Their key idea is to introduce a prior assumption that favors smoothness on multiple spatial scales of the image. Instead of trying to find a single point estimate for the reconstructed image they also sampled from the posterior distribution, which allowed them to quantify the associated uncertainties. However, the prior distribution is based on the generic assumption of smoothness and does not encode any “real world” or physical information on the structures present in the image.

An interesting modification of the standard RL method is proposed by Ingaramo et al. (2014). They show that RL could be effective in creating a merged image from multiple observations. They also observed that RL would take the best visual properties out of each image and merged them to result in a reconstruction closer to the “ground truth.”

Selig & Enßlin (2015) and Pumpe et al. (2018) approach the problem of deconvolution of Poisson noise data implicitly using the framework of information field theory (IFT). Based on physical prior assumptions on the spatial distribution of point sources and local correlation structures of diffuse emission, they are able to deconvolve Fermi Large Area Telescope (LAT) observations into a mixture of point-like and diffuse emission components. While their method works well for large-scale diffuse emission, such as the Galactic diffuse gamma-ray emission, it shows artifacts around the Galactic plane when applied to gamma-ray data from the Fermi-LAT (Selig et al. 2015).

Westerkamp et al. (2024) applied the same IFT framework to a Chandra X-ray observation of the supernova remnant (SNR) SN1006. The authors demonstrated that it is possible to separate the point-like emission of the extra-galactic background sources from the diffuse foreground emission of the SNR. The method delivered good reconstruction quality on synthetic data. When applied to observational data the authors found plausible reconstructions that include small artifacts, which they attributed to a mismatch between the modeled and actual PSF.

Other interesting approaches to the problem of deconvolution can of course be found in the field of computer vision. For example, Zoran & Weiss (2011) build on the very similar assumption of local correlation for natural images. However, instead of assuming a fixed local correlation structure, such as smoothness defined by a Gaussian kernel with a given bandwidth determining the correlation scale, they proposed to learn the local correlation structure on millions of patches extracted from a set of reference images. For this, they propose to split training images from representative data sets into small patches and learn a Gaussian mixture model (GMM) for the distribution of extracted patches. To use the GMM in reconstruction they introduce the concept of expected patch likelihood. They show that the patch-based GMM prior leads to

much improved image reconstructions for multiple inverse problems such as denoising, deblurring, and inpainting. Bouman et al. (2016) later adapted the patch prior successfully for the reconstruction of submillimeter radio data from the Event Horizon Telescope.

1.2. Some Remarks on Deep-learning Methods

In the past decade machine learning (ML), especially deep convolutional neural networks (CNN), have become the standard for deblurring and super-resolution of natural images (Zhang et al. 2022). However, their success relies on the availability of massive sets of training data with typically low noise and corresponding ground truth to perform supervised learning. For a variety of reasons, which we outline in the following, this approach is not ideal for astronomical X-ray and gamma-ray data.

First, astronomical X-ray and gamma-ray data are rare, expensive to obtain, and there is no ground truth, because we can only observe the sky with our given instruments. To circumvent this issue one could rely on transfer learning and simulated data. For example, radio or optical images could be forward-folded with a known PSF and degraded with Poisson noise to generate training data. Astronomical images, however, have a large dynamic range and show a high diversity in morphological structure, which would require a large amount of training data.

Nonetheless, this approach has been successfully applied for specific classes of sources, such as images of galaxies in the Sloan Digital Sky Survey (Schawinski et al. 2017) or as a post-processing step of results obtained with regularized RL (Akhaury et al. 2022). Recently, Sweere et al. (2022) used images from large-scale structure simulations with promising results.

In addition to the challenge of obtaining training data, it is known that the PSF of instruments such as Chandra is highly variable. It varies with observation conditions and depends, for example, on energy as well as offset angle from the pointing direction. To cover the whole parameter space of the PSF one would need to train a network for each of the PSF models.

Some of the challenges of deblurring with CNNs are due to the convolutional nature of the networks. In a single-layer, linear approximation, the network would need to learn the inverse filter of the PSF, corresponding to the well-known “Wiener filter” solution of the problem. This requires either large convolution kernels or deep network architectures, which in turn require large amounts of training data. This challenge was described early on by Xu et al. (2014), who introduced a baseline architecture for a deep CNN for image deblurring.

One last important challenge for deep-learning-based methods is obtaining uncertainty estimates for the reconstructed image from CNNs. In the limit of Poisson noise it is desirable to trace the full posterior distribution, which is not possible for most network architectures, with the exception of, for example, normalizing flows (NFs). This problem could be addressed with Monte Carlo (MC)-based methods, such as bootstrapping, which further increases the demand on training data and computational resources. For these reasons, we take a statistical modeling approach to the deconvolution problem in the presence of Poisson noise. We do, however, include elements from ML to represent the prior information on images.

2. Method

2.1. Poisson Joint Likelihood

Our goal is to recover an image $\mathbf{x}$ from a set of multiple low-count observations. Most generally, we assume our total data set consists of J individual observations of the same region of the sky, which we jointly model. The assumption is that the underlying, unknown “true emission” image that we are interested in does not change over the timescale of the observations. Under this assumption the data sets can be, for example, the following:

1. Different observations with one instrument or telescope at different times and observation conditions. For example, multiple observations of Chandra of the same astrophysical object with different offset angles and exposure times.
2. Observations from different telescopes which operate in the same wavelength range. For example, a Chandra and XMM observation of the same region in the sky.
3. A single observation with one telescope with different data quality categories and different associated instrument response functions, such as event classes for Fermi-LAT.

Any arbitrary combination of the possibilities is also possible. In principle, we can combine observations in different energy ranges. In this case, however, the assumption of the morphology not changing with energy would apply, which is rarely fulfilled for astronomical objects.

For each individual observation j the predicted counts can be modeled by forward-folding the unknown-flux image $\mathbf{x}$ through the individual instrument response:

$$\lambda_j(\mathbf{x}) = \mathbf{E}_j \cdot (\text{PSF}_j \otimes \mathbf{x}) + \mathbf{B}_j, \quad (1)$$

where the expected counts λ_j are given by the convolution of the unknown-flux image $\mathbf{x}$ with the observation-specific point-spread function PSF_j and an observation-specific image of the exposure $\mathbf{E}_j$, and background emission $\mathbf{B}_j$ can be optionally taken into account. A visual representation of the model of the imaging process for a single observation is shown in Figure 1.

Given a single observation j of an unknown-flux image $\mathbf{x}$ and assuming the noise in each pixel i in the recorded counts image $\mathbf{D}_j$ follow independent Poisson distributions with expectation λ_j , the likelihood P of obtaining the measured image with N pixels is given by

$$P(\mathbf{D}_j | \lambda_j) = \prod_{i=1}^N \frac{e^{-\lambda_{j,i}} \lambda_{j,i}^{D_{j,i}}}{D_{j,i}!}, \quad (2)$$

where λ_j depends on $\mathbf{x}$ as defined by Equation (1). By taking the logarithm, multiplying by -1 , and dropping the constant terms one can transform the product into a sum over pixels, which is also often called the “Cash” (Cash 1979) fit statistics:

$$\mathcal{C}(\mathbf{D}_j | \lambda_j(\mathbf{x})) = \sum_{i=1}^{N_j} (\lambda_{j,i} - D_{j,i} \log \lambda_{j,i}). \quad (3)$$

By summing over all observations, we obtain the (negative) joint log-likelihood of measuring a set of counts images $\mathbf{D}$

given an unknown-flux image $\mathbf{x}$:

$$\mathcal{L}(\mathbf{D}|\mathbf{x}) = \sum_{j=1}^J \mathcal{C}(\mathbf{D}_j|\lambda_j(\mathbf{x})). \quad (4)$$

In principle, we could obtain the maximum likelihood estimate (MLE), $\hat{\mathbf{x}}$, of $\mathbf{x}$ by directly optimizing Equation (4). Indeed, this is the objective of the RL algorithm. However, in general this represents an “ill-posed” inverse problem. Estimates for $\mathbf{x}$ which have similar likelihood values $\mathcal{L}(\hat{\mathbf{x}}_i)$ may look very different from each other and may also look different compared to a given ground truth, if available. Both the convolution operation and the noise lead to a loss of information in the data which cannot easily be recovered. The likelihood surface shows many local, close-by maxima. Depending on the starting value used for $\mathbf{x}$ and the optimization algorithm, the computed MLE is likely to be a local maxima. This is fundamentally the reason for RL decomposing estimates into point sources, described in Section 1.1.

2.2. A Posteriori Distribution

When reconstructing images in the context of inverse problems we are operating in a high-dimensional parameter space, because each pixel in the image represents a parameter in the optimization process. High-resolution X-ray-images are typically composed of up to $\mathcal{O}(10^6)$ pixels. By using prior information, we can introduce correlations among the parameters to reduce the “effective dimensionality” of the problem and guide the optimization process toward unique and more stable solutions.

Using Bayes’ rule, we can express the a posteriori likelihood under a given prior distribution, $P(\mathbf{x})$:

$$P(\mathbf{x}|\mathbf{D}) = P(\mathbf{x}) \frac{P(\mathbf{D}|\mathbf{x})}{P(\mathbf{D})}. \quad (5)$$

The maximum a posteriori (MAP) estimate of $\mathbf{x}$ is the value that maximizes Equation (5). Taking the logarithm of Equation (5), replacing the definition of the likelihood $P(\mathbf{D}|\mathbf{x})$ with Equation (4), and dropping the normalization term $P(\mathbf{D})$ which is independent of $\mathbf{x}$, we obtain the following expression for the (negative) log-posterior distribution $\mathcal{L}$:

$$\mathcal{L}(\mathbf{x}|\mathbf{D}) = \sum_{j=1}^J \mathcal{C}(\mathbf{D}_j|\lambda_j(\mathbf{x})) - \beta \cdot \mathcal{P}(\mathbf{x}), \quad (6)$$

where $\mathcal{C}(\mathbf{D}_j|\lambda_j(\mathbf{x}))$ represents the summed log-likelihood for an individual observation j , and $\mathcal{P}(\mathbf{x})$ represents the log of the prior distribution, with the factor β adjusting the weight of the prior relative to the joint log-likelihood term. The parameter β can be viewed as a hyperparameter, chosen by the user to adjust the strength of the prior distribution relative to the likelihood. The choice of β is optional, and for all experiments presented in this paper we set $\beta = 1$.

2.3. Patch-based Image Prior

In general, it is difficult to capture global image characteristics in a prior distribution taking into account all pixel-to-pixel correlations. The size of the corresponding correlation matrix would grow with the square of the number of pixels and thus quickly become computationally untraceable. However,

astronomical images typically only exhibit local correlations among the pixels. This is because the spatial scale of astrophysical processes in an image is always limited by the finite interaction time and the fact that the interaction cannot propagate faster than the speed of light. This means on smaller scales astronomical images often contain basic structures such as edges, corners, filaments, or periodic patterns formed by local astrophysical processes. For this reason, we adopt the patch-based image prior distribution introduced by Zoran & Weiss (2011).

A patch is defined as a small, typically square region of size $N_p \times N_p$ pixels, representing a local neighborhood in a given image. Mathematically, we can formulate the extraction of such patches as

$$\mathbf{x}_n = \mathbf{P}_n \mathbf{x}, \quad (7)$$

where $\mathbf{P}_n$ represents a neighborhood matrix to extract the n th patch from the image $\mathbf{x}$, and where $\mathbf{x}_n$ is the vector of length N_p^2 representing the “flattened” image patch. By obtaining a large number of patches we can model the distribution of pixel intensities within the patches using a GMM (McLachlan & Peel 2004). The GMM is a universal function approximator and thus can model arbitrary distributions. For now, we assume that we have a pretrained GMM which encodes the patch distribution of some reference image we have chosen. During the actual image reconstruction we aim to identify the most likely GMM component per patch, treating this as a discrete nuisance parameter that is optimized along with the image $\mathbf{x}$.

A convenient property of the GMM is that its probability density can be evaluated in closed form. For a given single patch $\mathbf{x}_n$,

$$P_{\text{GMM}}(\mathbf{x}_n) = \sum_{k=1}^K \pi_k \mathcal{N}(\mathbf{x}_n | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k^2), \quad (8)$$

where $\mathcal{N}$ represents a multivariate Gaussian normal distribution, $\boldsymbol{\mu}_k$ is the vector of Gaussian means (with dimension equal to the number of pixels in each patch), $\boldsymbol{\Sigma}_k$ is the variance-covariance matrix of the Gaussian distribution (which tabulates the variances and pairwise covariances of the patch pixels), and π_k is the weight of component k of the GMM.

If we knew the relevant component, k_n , for patch $\mathbf{x}_n$ we could derive the conditional log-prior distribution directly from Equation (8):

$$\mathcal{P}_{\text{GMM}}(\mathbf{x}_n | k_n) = -\frac{1}{2} [d \log(2\pi) + \log(|\boldsymbol{\Sigma}_{k_n}|) + (\mathbf{x}_n - \boldsymbol{\mu}_{k_n})^T \boldsymbol{\Sigma}_{k_n}^{-1} (\mathbf{x}_n - \boldsymbol{\mu}_{k_n})]. \quad (9)$$

If we split the image $\mathbf{x}$ into multiple nonoverlapping patches, for each patch $\mathbf{x}_n$ we can find the GMM component $\hat{k}_n$, which maximizes the conditional log-prior distribution given in Equation (9):

$$\hat{k}_n = \arg \max_{k \in \{1, \dots, K\}} \mathcal{P}_{\text{GMM}}(\mathbf{x}_n | k). \quad (10)$$

As the number of components in the GMM is typically limited to $\mathcal{O}(10^2)$ we can optimize the conditional log-prior distribution by simply evaluating it for each of the GMM components and choose the one with the maximum value. Finally, to evaluate the (optimized) log-prior of the full image

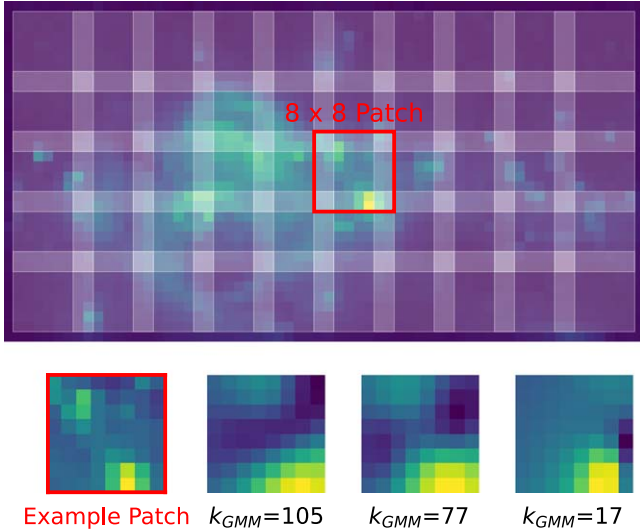


Figure 2. Grid of overlapping quadratic patches of size 8×8 pixels on top of an example image of size 128×68 pixels. The overlap size was chosen to be 2 pixels, which results in 50 patches in total. The patches are shown translucent, such that the horizontal and vertical “stripes” result from two neighboring overlapping patches, while the small, even lighter squares result from four patches overlapping at their corners. The bottom row shows the red example patch $\mathbf{x}_n$ on the left. The three other patches show this example patch represented by a linear combination of the six largest eigenvectors of the covariance matrices of the three most likely GMM components $k = 105, 77$, and 17 . This example uses the GMM learned from the JWST Cassiopeia A image. 

we sum the maximum values of the individual patches:

$$\mathcal{P}(\mathbf{x}, \hat{\mathbf{k}}) = \sum_{n=1}^N \mathcal{P}_{\text{GMM}}(\mathbf{P}_n \mathbf{x} | \hat{k}_n) + \sum_{n=1}^N \log \pi_{\hat{k}_n}, \quad (11)$$

where $\mathbf{P}_n$ is again the matrix that extracts the n th patch from the image $\mathbf{x}$ and $\pi_{\hat{k}_n}$ is the weight of GMM component $\hat{k}_n$, with $\hat{k}_n$ being the index of the GMM component that maximizes the log-prior distribution for patch n .

The bottom row of Figure 2 shows an example of a patch along with its representation in the “eigenbasis” of the covariance matrices of its three most likely GMM components. Visually, the representation of the patch in the eigenbasis of the covariance results in a denoised version of the patch.

Zoran & Weiss (2011) found that the reconstruction quality can be improved by working with a grid of overlapping patches. This is illustrated in Figure 2. However, using a grid of overlapping patches makes the interpretation of the prior distribution more difficult, because pixels at the corners and edges of patches are counted multiple times. For this reason, we introduce a weighting mask when evaluating the GMM prior distribution. This mask weighs the pixels of each patch by their overlap with neighboring patches. Overlapping edges are assigned a weight of a half while overlapping corners are assigned a weight of a quarter when computing the log-prior distribution for a patch as given by Equation (9).

Apart from this, the process for choosing the most likely GMM prior component remains unmodified. However, we note that we neglected the prior dependence between neighboring patches.³ As the choice of the GMM component per patch is

discrete, a simultaneous gradient-based optimization of all patches is not easily possible, because we cannot compute gradients across GMM components. This would require a continuous representation or interpolation between GMM components.

2.3.1. Image Dynamic Range and Patch Normalization

Astronomical images typically show a much higher dynamic range compared to natural images. To normalize the patches, Zoran & Weiss (2011) proposed to subtract the patch mean before learning the GMM. This leads to the GMM modeling the same morphological features at different levels of contrast as different components. While this proved to be beneficial in the context of reconstructing natural images, it does not scale to images with high dynamic range, because every morphological feature would need to be present at many levels of intensity. While this could be handled by introducing significantly more components to the GMM, it seems unfeasible in practice due to the limited availability of training data and computational resources. For this reason, we instead propose a modified normalization for the patches:

$$\widetilde{\mathbf{P}}_n \mathbf{x} = \frac{\mathbf{x}_n - \bar{\mathbf{x}}_n}{\bar{\mathbf{x}}_n}. \quad (12)$$

By dividing by the mean of the patch, $\bar{\mathbf{x}}_n$, the structure in the patch becomes independent of the total brightness the patch contains. This results in patches being grouped only by their pixel-to-pixel contrast and not by their overall brightness. This corresponds to the assumption that the structure in the patch is a priori independent of the total flux contained in the patch, which we consider a valid assumption for astronomical images. This also generalizes to the prior distribution of the whole image described by Equation (11). The value of the prior distribution is independent of the total flux of the image. The prior distribution simply regularizes the structure in the image, while the overall intensities are typically well constrained by data as quantified by the Poisson likelihood term.

2.3.2. Cycle Spinning

The evaluation of the patch prior divides the reconstructed image into a fixed grid of overlapping patches. To avoid artifacts in the reconstruction due to the fixed position of the grid, we use the method of “cycle spinning” (Coifman & Donoho 1995). Before evaluating the patch prior at each iteration of our optimizer, we shift the image $\mathbf{x}$ by a random number of pixels in both the x - and y -directions, where the shift is less than or equal to half the size of a patch. The same method has been used, for example, in Esch et al. (2004).

A variation of cycle spinning specifically for the patch prior is proposed in Parameswaran et al. (2018). Instead of a regular grid for the patches, they propose to work with a randomized grid of patches, where the center of each patch is randomly shifted while ensuring that patches overlap by at least 1 pixel. In doing so, they find that a smaller total number of patches can be used, while retaining the same reconstruction quality. This improves the computational performance of the method. We also implemented the evaluation of the patch prior on a randomized patch grid for comparison and found similar results as Parameswaran et al. (2018): The reconstruction was slightly faster with the same reconstruction quality. We found,

³ That is, we continue to evaluate the log-prior distribution of the full image by summing over the overlapping patches as in Equation (11), even though overlapping patches include common pixels and are thus clearly not a priori independent.

however, (i) there are practical issues with the reproducibility of results when using this method on a GPU, and (ii) the method requires significant code complexity to handle the edge of images correctly. Given that the performance improvement was small, we thus decided to use the simpler regular grid and cycle spinning, as described in the previous paragraph.

2.4. Systematic Errors of Predicted Counts

Each observation $\mathbf{D}_j$ is subject to systematic errors. For example, in the pointing accuracy of the telescope when repeatedly pointed to the same object, or when using different types of telescopes for observation, the reconstructed position of an astronomical object may vary among the observations. The process of correcting for these inaccuracies is called astrometry and is a standard procedure in, for example, Chandra data analysis. A second example of systematic errors is the variation of background emission between multiple observations.

To account for both systematic effects, we extend the model in Equation (1) for the predicted counts by a background normalization factor α_j and a linear shift in position $\phi_j(\mathbf{x}|\delta_{x,j}, \delta_{y,j})$, which is specific for each observation. This yields the following modified model:

$$\lambda_j(\mathbf{x}, \delta_{x,j}, \delta_{y,j}, \alpha_j) = \mathbf{E}_j \cdot (\text{PSF}_j \otimes \phi_j(\mathbf{x}|\delta_{x,j}, \delta_{y,j})) + \alpha_j \cdot \mathbf{B}_j. \quad (13)$$

In practice, one observation is used as a reference and its corresponding systematic shifts in position $\delta_{x,j}, \delta_{y,j}$ are frozen during optimization. An example of the effect of the positional shift and background normalization can be seen in Figure 11.

Modifying the (negative) log-posterior distribution given in Equation (6) for the normalization of the $\mathbf{x}_n$ in Equation (12) and the adjustment for systematic effects in Equation (13) yields

$$\mathcal{L}(\mathbf{x}, \mathbf{k}, \boldsymbol{\delta}, \boldsymbol{\alpha}|\mathbf{D}) = \sum_{j=1}^J \mathcal{C}(\mathbf{D}_j | \lambda_j(\mathbf{x}, \delta_{x,j}, \delta_{y,j}, \alpha_j)) - \beta \left\{ \sum_{n=1}^N \mathcal{P}_{\text{GMM}}(\tilde{\mathbf{P}}_n \mathbf{x} | k_n) + \sum_{n=1}^N \log \pi_{k_n} \right\}, \quad (14)$$

where $\mathbf{k} = (k_1, \dots, k_N)$, $\boldsymbol{\delta} = (\delta_{x,1}, \delta_{y,1}, \dots, \delta_{x,J}, \delta_{y,J})$, and $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_J)$. For fixed $\mathbf{x}$, Equation (14) is optimized as a function of $\mathbf{k}$ by Equation (10), but with $\mathbf{x}_n$ replaced by $\tilde{\mathbf{P}}_n \mathbf{x}$. The joint prior distribution of $(\mathbf{x}, \mathbf{k}, \boldsymbol{\delta}, \boldsymbol{\alpha})$ is given by the term in Equation (14) within curly brackets. Note that this places uniform prior distributions on $\boldsymbol{\delta}$ and $\boldsymbol{\alpha}$ and formally sets the prior on $\mathbf{P}_n \mathbf{x}$ to the GMM evaluated at $\tilde{\mathbf{P}}_n \mathbf{x}$. Since this does not depend on $\tilde{\mathbf{x}}_n$, we implicitly assume a uniform prior on each $\tilde{\mathbf{x}}_n$. In practice, we do not expect the choice of prior distribution on the scale of each patch to influence the final fit, but some users may prefer to assume a flat prior on the log of each $\tilde{\mathbf{x}}_n$ by adding $\sum_{n=1}^N \tilde{\mathbf{x}}_n$ within the curly brackets on the right-hand side of Equation (14).

2.5. Estimation of Statistical Errors

Estimating the global covariance matrix for an estimated image is computationally infeasible, because the size of the matrix grows with the squared number of pixels. However, there are other ways to estimate the statistical uncertainties associated with the flux image. We propose two methods:

1. Bootstrapping observations: given that multiple observations of the same region of the sky are available, *Jolideco* can be run repeatedly on a random subset of images. This results in a set of reconstructed images representing a distribution from which we can estimate the errors on $\mathbf{x}$. Depending on the size and number of the images, the associated computational effort can be very large as it requires running *Jolideco* many times to sufficiently describe uncertainty in the fitted image.
2. Bootstrapping events: if only one or a small number of observations are available, we propose bootstrapping of the events used to compute the counts image(s). This corresponds to splitting a single observation into multiple observations by splitting the single exposure into multiple time intervals. This way, one can apply the first proposed method.

How well these bootstrap error estimates approximate the true posterior uncertainty and/or the frequency coverage of the derived bootstrap intervals is a topic of future study. We also experimented with computing the diagonal of the Hessian of the posterior distribution evaluated at the MAP estimate. However, we found this numerically challenging because of the random cycle spinning.

3. Implementation

3.1. Jolideco Framework

The goal of the reconstruction process is to find the mode of the posterior distribution defined by Equation (14). Given that there is a parameter for each pixel in the image, this is a high-dimensional optimization problem. Differentiable programming frameworks such as PyTorch (Paszke et al. 2019) allow for solving high-dimensional optimization problems using back propagation. They also offer a variety of optimization methods, such as stochastic gradient descent (SGD). The latter is also applicable in our case, as we work with multiple observations at the same time, where a subset of the observations represents a “mini batch” in terms of ML.

We implemented the *Jolideco* method as an independent Python package, based on PyTorch as the computational backend. We defined a modular, object-oriented code structure, to allow parts of the algorithm to be flexible and interchangeable, such as the choice of the patch normalization, the choice of GMM, optimization methods, and serialization formats for the reconstruction results, as well as the corresponding diagnostic information, such as the trace of the posterior and prior likelihood values. This will simplify future improvement and extension of the package as well as the adoption of *Jolideco* by the astronomical community. The package is available on the platform GitHub (Donath 2024).⁴

We use Numpy (Harris et al. 2020) for handling data arrays and Matplotlib (Hunter 2007) for plotting. To learn the GMM from patches, we use the standard implementation provided in the Scikit-Learn (Pedregosa et al. 2011) package. To handle the serialization of images to the FITS data format and to handle World Coordinate Systems (WCSs), we use Astropy (Astropy Collaboration 2018).

⁴ <https://github.com/jolideco/jolideco>

3.2. Learning the GMMs

The patch-based image prior described in Section 2.3 requires a pretrained GMM. We train the GMM by choosing a set of reference images, splitting them into patches of size 8×8 pixels, and learning a GMM with $K = 128$ components from the whole distribution of patches. In general, the strength of the GMM prior is its high “expressiveness”. This means the prior distribution contains a large variety of structural shapes and has freedom to adapt to the data in the optimization process by choosing the most likely GMM component for each patch in the image. Zoran & Weiss (2012) found empirically that the reconstruction quality depends on the number of GMM components. A higher number of components clearly leads to improved reconstruction quality, as the GMM was more expressive. This also aligns with the intuitive argument that GMMs are universal function approximators; with a sufficiently large number of components they can approximate any distribution. Zoran & Weiss (2012) found the reconstruction quality begins to saturate with about 32 components, at least for natural images. We decided to adopt a more conservative number with 128 components.

With radio data, Bouman et al. (2016) found their results to be mostly independent of the choice of data used for learning the GMM prior. In the low-signal-to-noise regime, however, we typically expect a higher dependence of the results on the choice of prior, simply because the prior distribution is more influential when the data are less informative. Thus, there is the potential for the choice of GMM training data to be influential on results, and the prior should be adapted to the particular analysis scenario in order to avoid bias. For this reason and for purposes of testing, we learned and provide a selection of GMMs to be used with the patch prior for download in a GitHub repository; illustrative examples of the training images are also provided in this repository.⁵

3.2.1. Zoran & Weiss

As a baseline reference we use the original trained GMM provided by Zoran & Weiss (2011). It was trained from the Berkeley image database (Martin et al. 2001), which contains natural images and images of everyday scenes and objects.

The training images were split into patches of size 8×8 pixels, resulting in a training set of 2×10^6 patches which were used to fit a GMM with 200 components. The training patches were mean subtracted before the GMM was fit. The model is provided by the authors as part of the source code for download.⁶ More details of how the model was trained are given in Zoran & Weiss (2011).

3.2.2. GLEAM Radio Data

As an alternative we consider a GMM prior learned from astronomical radio data. For this, we relied on high-signal-to-noise radio data from the GLEAM survey (Hurley-Walker et al. 2022). We downloaded the data using the Glean VO Client.⁷ We restricted the choice of data to the Galactic plane between $b = \pm 10$ deg and $\ell = \pm 180$ deg to avoid a bias toward extra-galactic point sources and to represent structures of Galactic sources with rich morphology, such as SNRs, pulsar winds, etc. The data set includes Galactic as well as extra-galactic point

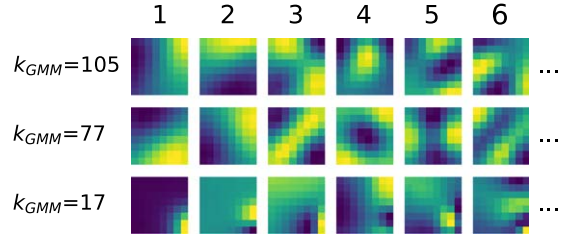


Figure 3. Example eigenimages of the covariance matrices of the GMM learned from the JWST Cas A image. The figure shows the first six eigenvectors, reshaped as images, of the GMM components {105, 77, 17} (same as Figure 2), in descending order of the value of their corresponding eigenvalue. The color scale varies between the images, so their contrast is different. 

sources in the background. In total, we extracted $\approx 6 \times 10^5$ patches and trained a GMM with 128 components. We used the patch normalization described in Section 2.3.1.

3.2.3. JWST Cassiopeia A

As a specific prior for the analysis of SNRs we learned a GMM prior from data of the JWST. The JWST is a large, infrared space telescope launched in late 2021. It delivers image data at very high spatial resolution with an excellent signal-to-noise ratio. Both of these properties make JWST data an excellent choice for extracting patches for training a GMM. To limit the data-reduction effort we relied on a promotional image of the SNR Cassiopeia A (Cas A).⁸ We first converted the image to grayscale using a perceived brightness model from the Scikit-Image package (van der Walt et al. 2014). Then we split the image into ≈ 2 million 8×8 patches and learned a GMM with 128 mixture components. Again, we relied on the patch normalization described in Section 2.3.1. Figure 3 shows the principal eigenimages of the covariance matrices of three example components of the GMM. The patches show a large variety in morphological shapes, including edges, gradients, bright compact spots, and more combinations thereof.

3.2.4. Chandra SNRs

To demonstrate the flexibility in the selection of data to train the GMM, we also chose higher-signal-to-noise images from the Chandra catalog of SNRs. We learned a GMM prior from images of the bright Chandra SNRs Cas A and G292.0+01.8. In total, we extracted ≈ 0.5 million 8×8 patches from each of the images and learned a GMM with 128 components.

3.3. Optimization Strategy

Maximizing the a posteriori likelihood defined by Equation (14) with respect to the model parameters $\mathbf{x}$ is a high-dimensional optimization problem. We found the convergence greatly improved when the optimizer operates on $\log \mathbf{x}$. This is to be expected, as the brightness in pixels of an astronomical image can span multiple orders of magnitude. We used the ADAM algorithm (Kingma & Ba 2014), a SGD method, without decay and a fixed learning rate of $\text{lr} = 10^{-3}$. In practice, we found the convergence of the method to be very stable. However, we applied the following strategy to speed up the optimization process:

⁵ <https://github.com/jolideco/jolideco-gmm-prior-library>

⁶ <https://people.csail.mit.edu/danielzoran/eplcode.zip>

⁷ <https://github.com/ICRAR/gleanvo-client>

⁸ We downloaded the full-resolution (4k) PNG image from <https://webbtelescope.org/contents/media/images/2023/121/01GWQBBY77MHGFV3M3N63KDCEJ>.

1. We start from a random image where each pixel was drawn from a normal distribution with the mean corresponding to the total mean of the flux in the image and a standard deviation of 10% of the mean. We first estimate $\mathbf{x}$ using a uniform prior and early stopping after 50 iterations. This corresponds approximately to a reconstruction obtained with the standard RL algorithm. As the uniform prior is much faster to evaluate ($\mathcal{O}(100)$ times faster), this step quickly produces a much better starting value for the optimization with the patch prior described in Step 3.
2. We freeze the image parameters and optimize the parameters associated with the systematic error described in Equation (13). Again, this procedure quickly delivers better starting values for the optimization in Step 3. Computation is fast because the number of parameters is very small (and scales linearly with the number of data sets).
3. Finally, we simultaneously optimize the image $\mathbf{x}$, the GMM component indicators for each patch k_n , and the parameters associated with systematic error until convergence. Algorithmically, this is accomplished by iteratively updating (a) $\mathbf{k}$ for fixed $\mathbf{x}$, α , and δ (via Equation (10) with $\mathbf{x}_n$ replaced by $\mathbf{P}_n \mathbf{x}$), and (b) $(\mathbf{x}, \alpha, \delta)$ for fixed $\mathbf{k}$ via SGD. Convergence is verified visually by inspecting the series of values of the log-posterior distribution evaluated at the iterates or the model parameters. We adjust the number of epochs (number of passes through the whole set of observations) such that the curve had flattened out for the last few hundred iterations. This is also the point in the iteration where the noise in the loss, introduced by SGD and the cycle spinning, clearly dominates any remaining decrease in the mean loss.

4. Experiments

4.1. Test Datasets

We have designed multiple test scenarios to challenge the several deconvolution algorithms listed in Section 4.2 and to evaluate and compare their performance. The test scenarios consist of mixtures of point and extended sources, with typical morphological characteristics of astrophysical sources. We arranged the sources in different spatial patterns and with different spacings between them. We include combinations of point and extended sources, sometimes superposed, in four arrangements (labeled A–D in Figure 4). Each arrangement is considered at several signal-to-noise ratios (obtained with a constant background and numbered in order of increasing signal). The resulting scenarios are denoted by concatenating the two labels, e.g., A3 or C1; see Appendix A for details. The goal of the experiments is not to set up or devise a “perfect” deconvolution algorithm, but to help assess their relative strengths and limitations, and assess the situations where they fail.

In addition, we adopt two distinct instrument scenarios: one which mimics Chandra observations with a sharp PSF (a Gaussian with $\sigma = 2$ pixels) and low background (at 0.01, 0.1, and 1 cts pixel^{−1}) but lower effective area (nominally with an exposure of 1 unit), and the other which mimics XMM-Newton observations with a broader PSF ($\sigma = 3$ pixels), higher background (at 0.1, 1, and 10 cts pixels^{−1}), and higher effective area (imitated with an exposure of 5 units).

The ground truth and corresponding expected counts for several exemplar combinations of source and instrument scenarios are shown in Figure 4.

4.2. Methods

We compared the performance of Jolideco with different prior assumptions and against the LIRA method introduced by Esch et al. (2004). More specifically, we compared the following methods:

- M1. Jolideco (uniform, $n = 10$): we used Jolideco with a uniform prior on $\mathbf{x}$ and ran the optimizer for 10 iterations. We consider this as a baseline method reference similar to RL. While the optimization procedure is different, we expect similar behavior for the local correlations of pixels. This ad hoc approach of early stopping is often used in practice to regularize the RL output.
- M2. Jolideco (uniform, $n = 1000$): we used Jolideco with a uniform prior on $\mathbf{x}$, but running the optimizer for 1000 iterations. This is the same as above but without early stopping. This gives rise to the issue of RL decomposing extended emission into point sources.
- M3. Pylira: we used the LIRA method (Esch et al. 2004) via the Pylira Python wrapper (Donath et al. 2022). We relied on the default prior definition and kept the background norm fixed. In total, we ran the Markov Chain Monte Carlo (MCMC) sampler for 5000 iterations and discarded the first 500 iterations as a “burn-in” period to allow the MCMC chain to converge. We computed the mean of the posterior distribution ignoring the burn-in period.
- M4. Jolideco (patch, Zoran–Weiss): we used Jolideco with the patch prior provided by Zoran & Weiss (2011). To be consistent with how the GMM was learned, we only subtracted the mean of the patch and skipped the division by the mean we introduced in Equation (12). To make the prior applicable to astronomical images, we rescaled the image to the range $0 > \mathbf{x}_i > 1$ by dividing them by the maximum value defined by the ground truth.⁹
- M5. Jolideco (patch, GLEAM v0.1): we used Jolideco with a patch prior learned on the GLEAM data, described in Section 3.2.2 and the patch normalization described by Equation (12). We used the optimization strategy described in Section 3.3 with 2000 epochs.

We ran the comparison on all combinations of methods, source scenarios, background levels, and instrument scenarios. Here we only discuss the subset of the results that we consider the most relevant. The complete set of results is available on GitHub.¹⁰

4.3. Results

4.3.1. Comparing Methods

First, we compare the performance of the five methods and their differing prior assumptions on the different source scenarios. Figure 5 shows the results for a selected subset of scenarios A1, B3, C3, and D4. For quantitative comparison we list in Table 1 the corresponding values of the structural similarity index (SSI, Wang et al. 2004) and the normalized root mean squared error (NRMSE).¹¹

⁹ This is not completely fair, as it requires knowledge of the brightest expected pixel in $\mathbf{x}$. In practice, this could be estimated from the brightest pixel in the data and the shape of the PSF.

¹⁰ <http://jolideco.github.io/jolideco-comparison>

¹¹ We use the definition of NRMSE = $\frac{\|\mathbf{x}_{GT} - \mathbf{x}\|}{\|\mathbf{x}_{GT}\|}$, where $\|\cdot\|$ is the Frobenius norm.

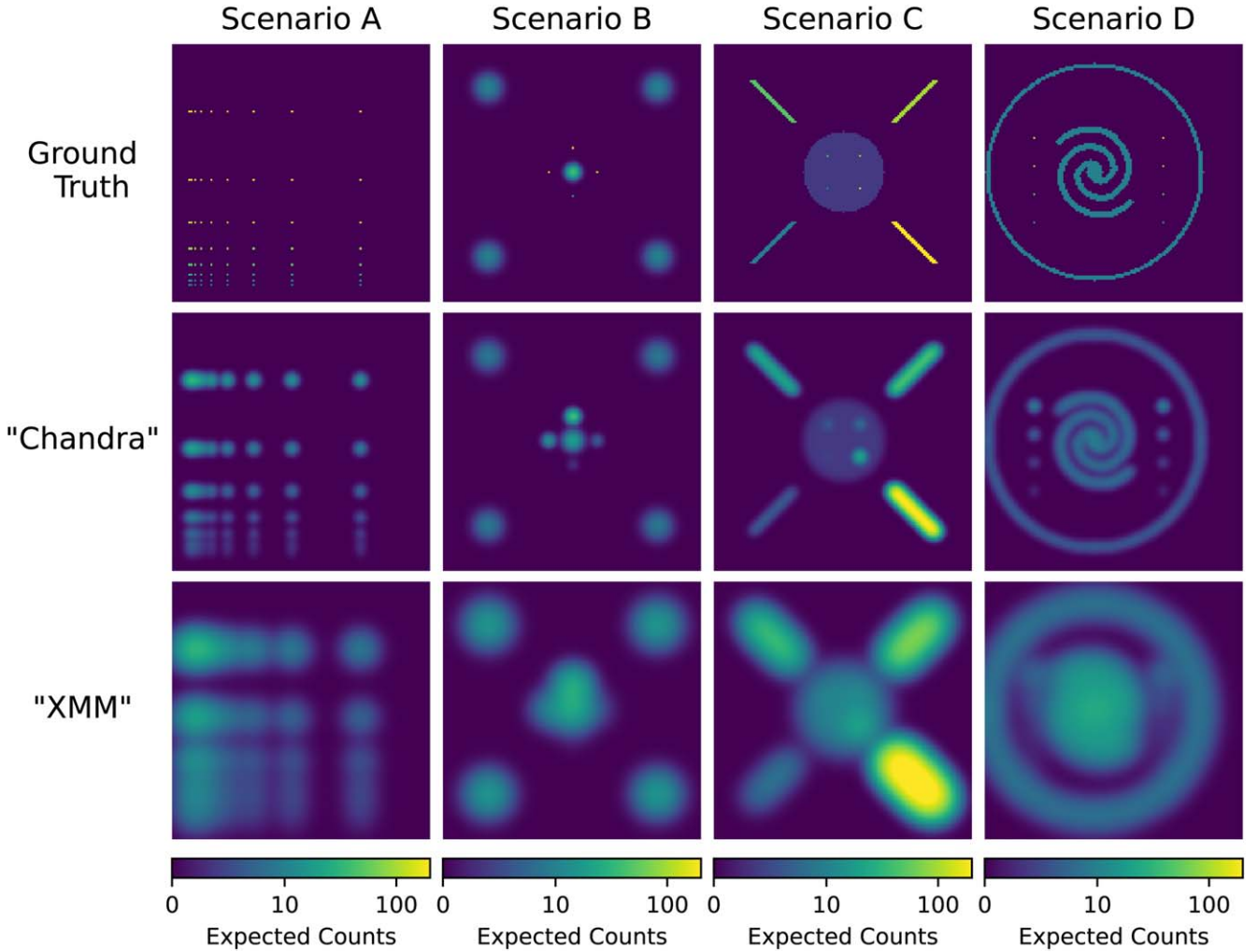


Figure 4. Illustration of the expected counts for the different source scenarios we used to generate test data sets. The source patterns shown are for the (a) “Points,” (b) “Asterism,” (c) “Shield,” and (d) “Spiral” cases (see Section A. For the Chandra instrument scenario we convolve the ground truth image with a narrow Gaussian-shaped PSF of width $\sigma = 2$ pixel. For the XMM scenario we used a Gaussian-shaped PSF of width $\sigma = 6$ pixels and an effective area that is 5 times larger than that of Chandra’s. All images are of size 128×128 pixels.

For scenario A1 all methods provide acceptable to excellent results for the isolated point sources in the upper right of the image. This demonstrates all methods take into account the PSF and background models correctly. The exception is method M2 (Jolideco, uniform prior, stopped early) because its optimization process has not finished. M3, which optimizes for 1000 iterations, clearly improves the result and shows almost perfect point sources corresponding to the size of approximately 1 pixel. The Jolideco method with the Zoran–Weiss patch prior (M4) results in less sharp images, because the prior has been learned on natural images of landscapes and everyday objects and scenes, which typically do not include point sources. The Jolideco method with the prior learned from GLEAM data (M5), which also includes point sources, shows much improved results relative to M4 but is not as sharp as the results obtained with the uniform prior (M3). In the lower left of the image the point sources are fainter and closer together, challenging all methods. The uniform prior optimized “to end” (M3) shows good results, separating the sources well. All other methods, however, cannot fully resolve these weak and closely situated point sources.

For scenario B3 all methods recover the point sources well. There is, however, a large difference in the reconstruction of

the Gaussian-shaped extended sources. As expected, the uniform prior assumption lets the emission decompose into many point sources, neglecting the spatial correlations among them. Pylira shows slightly improved results with the gaps between the bright emission peaks being more filled. However, Pylira lacks a convincing result of reconstructing the source into a single object. The Jolideco method with the patch priors shows the best results both visually and as measured by SSI; see Table 1. With the GLEAM v0.1 patch prior the reconstruction is almost perfect, visually almost indistinguishable from the ground truth. The Zoran & Weiss prior also shows good reconstruction of the extended sources, but less smooth and leaving a more visible extension in the surrounding point sources. The weak extended source with four embedded brighter point sources of scenario C3 is challenging for all the methods. While the methods again reconstruct the point sources well, the visual differences are large for the weak underlying extended emission. The uniform prior with 1000 iterations decomposes the extended source into a grid of almost equidistant point sources resembling only the outer boundary of the original source. Pylira shows a similar tendency but improves the reconstruction of the extended source by leaving more emission in between the bright emission spots. The best

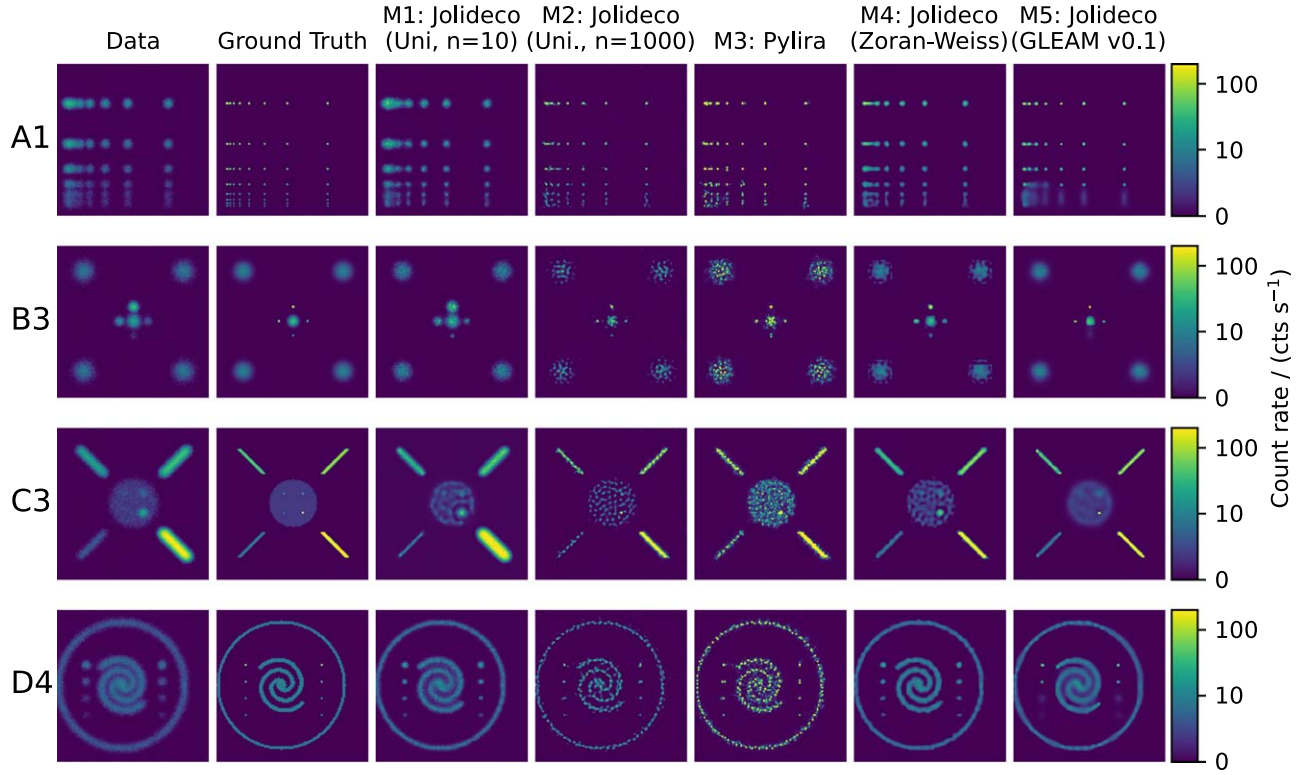


Figure 5. Comparison of the different deconvolution methods and prior assumptions for selected source scenarios, as described in Section 4.1. The simulations depicted used a fixed background level of $\lambda_{\text{Bkg}} = 0.01 \text{ cts pixel}^{-1}$, a uniform exposure, and a Gaussian PSF of width $\sigma_{\text{PSF}} = 2$ pixels, corresponding to the “Chandra” instrument scenario. The leftmost column shows the simulated counts and the second column shows the underlying ground truth. The remaining columns show the reconstructed flux for each method and prior assumptions for *Jolideco*. To enhance weak structures, the images use an arcsinh stretching with a scale parameter of $a = 0.02$. The stretching is the same for all images. The methods are described in detail in Section 4.2. A more detailed representation of the results is available at <https://jolideco.github.io/jolideco-comparison>.

Table 1
Image Metrics for the Different Source Scenarios and Methods Shown in Figure 5

Scenario	Jolideco (Uni, $n = 10$)	Jolideco (Uni., $n = 1000$)	Pylira	Jolideco (Zoran–Weiss)	Jolideco (GLEAM v0.1)
A1	0.91/0.94	0.98/0.52	0.97/0.68	0.93/0.82	0.95/0.73
B3	0.99/0.98	0.99/0.32	0.99/0.43	0.99/0.68	1.00/0.51
C3	0.95/0.76	0.96/0.83	0.95/0.80	0.98/0.36	0.98/0.56
D4	0.94/0.75	0.85/1.21	0.81/1.65	0.96/0.68	0.94/0.60

Note. The first number in each cell shows the structural similarity index (SSI; the closer to unity, the better) and the second number shows the normalized root mean square (NRMSE; the lower the better).

results are achieved by the patch priors, which reconstruct a plausible extended source with approximately uniform brightness. The jet-like features are decomposed into a “trail” of point sources by the uniform prior and Pylira method. The patch prior achieves a much more plausible reconstruction with almost uniform brightness for the linear structures. We also note that the patch prior does not show a preferred direction or bias in reconstruction, which could be possibly encoded in the training data.

The last scenario, D4, shows a similar result as the previous test: The uniform prior and Pylira decompose the spiral and ring-like structure of the image into points. The patch priors reconstruct a visually more plausible flux image, much closer to the ground truth.

In summary, *Jolideco* with the patch priors shows excellent performance on a variety of morphological shapes. We find the prior learned from the GLEAM data to be a very

good compromise between the ability to separate nearby point sources and the ability to reconstruct the spatial correlation of extended sources.

4.3.2. Comparing Signal-to-Noise Scenarios

As a second experiment we explore the dependency of the results on the background level and signal strength. Figure 6 shows the results for the *Jolideco* method with the GLEAM prior in scenarios D1–D5. The background level varies between $\lambda_{\text{Bkg}} = 0.01, 0.1$ and 1 cts pixel^{-1} . As expected, the reconstruction quality improves with higher signal strength and lower background level. Visually, the method converges to the ground truth for very high signal-to-noise ratio. Indeed, with very high signal strength there is no dependency of the result on the background level, particularly in scenario D5. On the other hand, the method struggles to reconstruct a useful signal in the case of very high background (scenarios D1–D3).

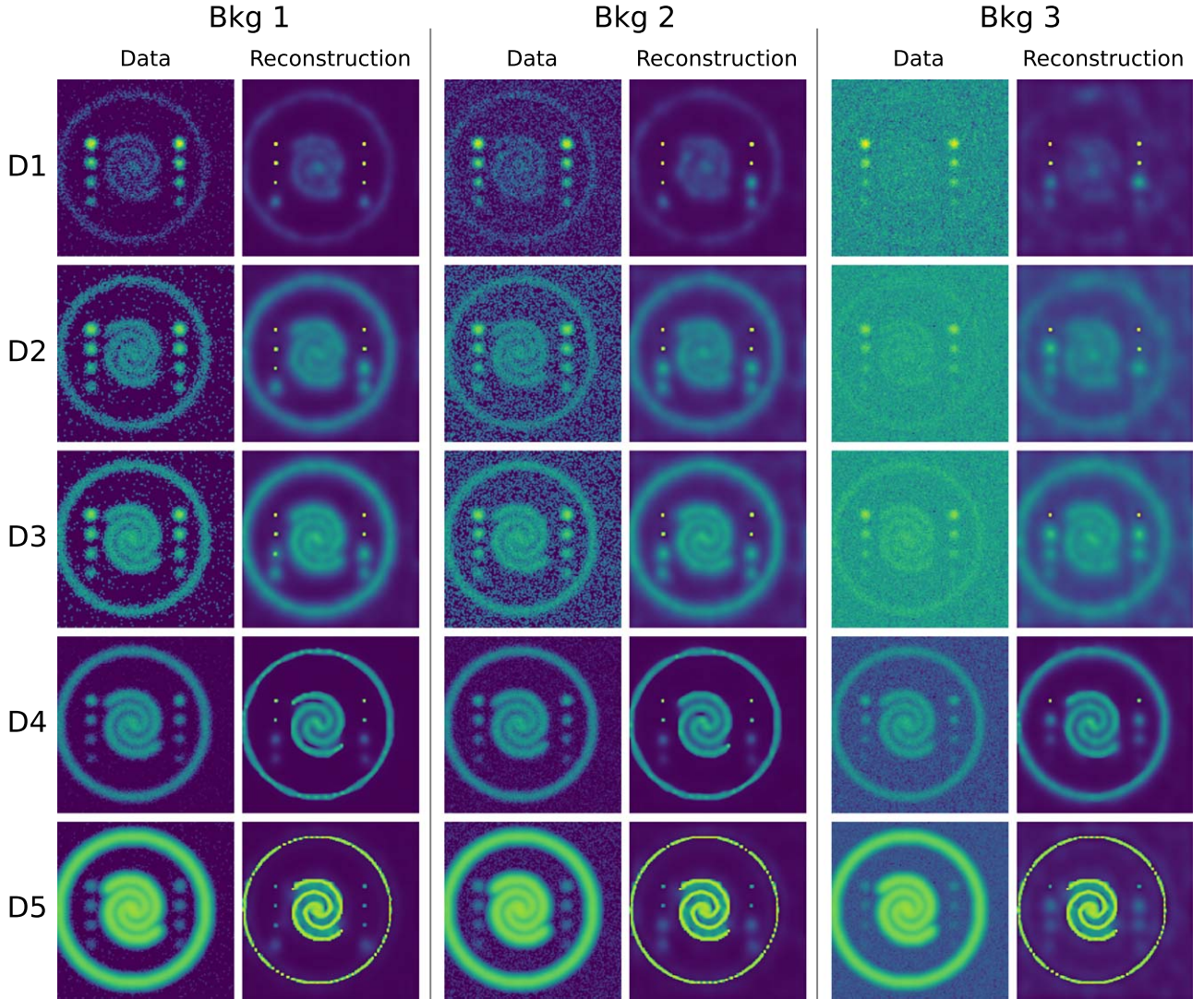


Figure 6. Comparison of *Jolideco* with the GLEAM v0.1 prior for different background levels and signal strength, as represented by scenarios D1–D5. The images are grouped by background level, with each pair of columns showing the counts data in the left column and the reconstruction in the right column. The rows show the scenarios D1–D5 as described in Section 4.1. Each pair of columns corresponds to a background level, 0.001 cts pixel⁻¹, 0.01 cts pixel⁻¹, and 0.1 cts pixel⁻¹. The complete results for the simulations and additional information are available at <https://jolideco.github.io/jolideco-comparison>.⁵

4.3.3. Comparing Instrument Scenarios

Finally, we compare *Jolideco* results for the two instrument scenarios. Figure 7 shows the data as well as flux images reconstructed using *Jolideco* and the GLEAM v0.1 patch prior. We consider source scenarios D1–D5 and the “Chandra” and “XMM” instrument scenarios as well as using a combination of their data to reconstruct a single image.

Using the data from the “Chandra” scenario alone results in a good reconstruction in all variations of scenario D. As expected, the reconstruction quality improves for higher signal-to-noise ratio.

Using only the data from the “XMM” scenario results in lower-quality reconstructions, due to the lower angular resolution of the data. However, we find that the spiral arm structure is reconstructed at all levels of signal strength, even when barely visible by eye. The ring structure is sharpened considerably in all cases as well. The nearby point sources on the left and right of the spiral arm structure cannot be separated into individual point sources, but remain blurred along the

direction of their neighbors. This is a similar situation as in the lower left of the image in the A1 point-source scenario, when sources are close by.

Combining the data results in the best reconstruction quality. The addition of the lower-resolution “XMM” data does not negatively affect the quality of the overall reconstruction. When comparing scenario D5 we see that the “Chandra” reconstruction shows a very good reconstruction of the spiral structure but shows speckles in the surrounding ring. The latter suggests that the GMM prior is not strong enough to maintain the pixel-to-pixel correlations in the case of high signal strength. This observation is confirmed in the metrics shown in Table 2, where scenario D5 shows lower SSI values than D4 for the same background level. The strength of the prior could be increased with the β parameter, but for this study we use the default settings and do not tweak individual parameters. The speckly behavior is not observed in the “XMM” case, where the reconstruction of the ring is smoother. In this case the correlation between pixels that are further apart is stronger because of the larger correlation PSF. However, the

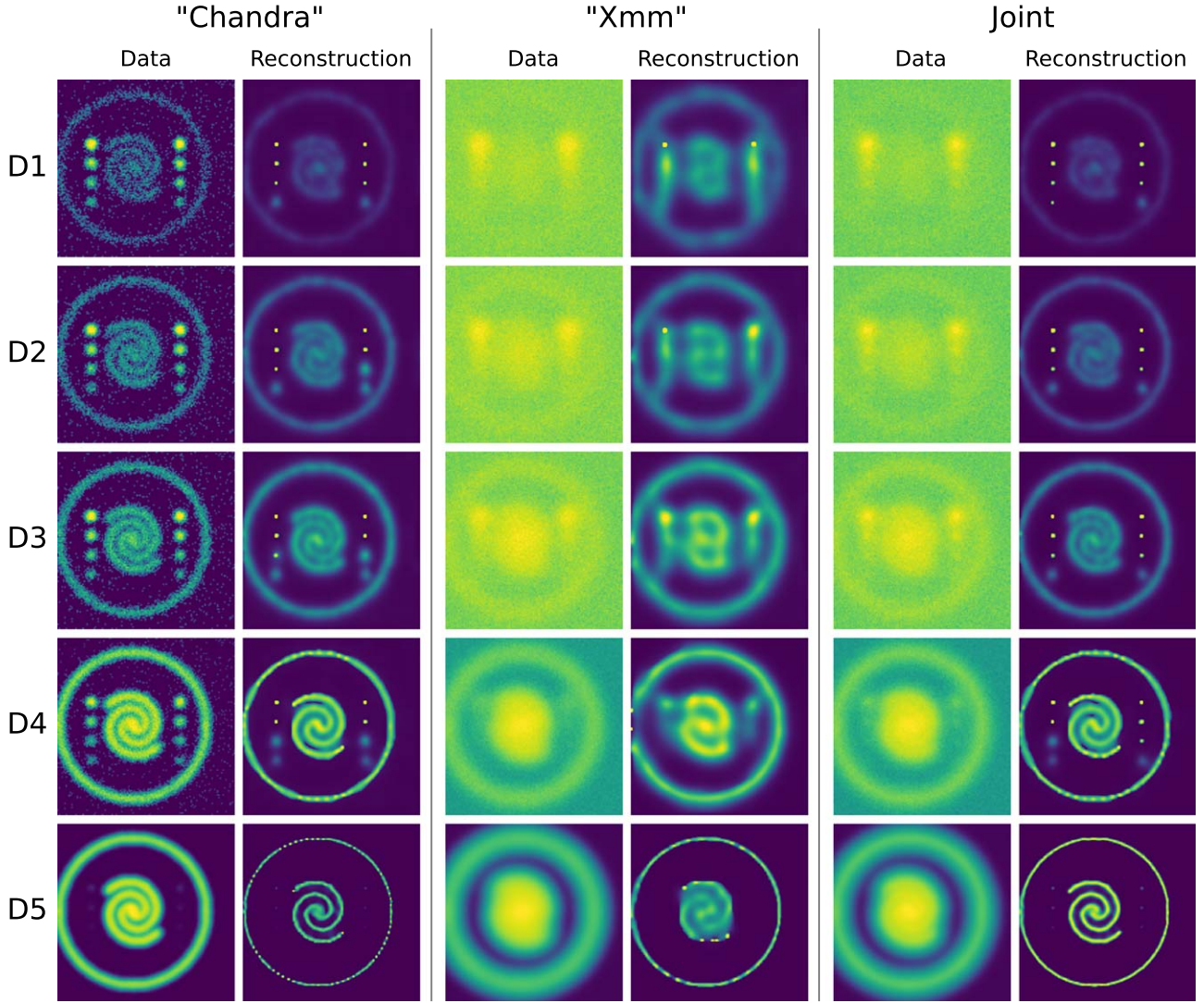


Figure 7. Comparison of Jolideco with the GLEAM v0.1 prior for the different instrument scenarios. The images are grouped by instrument scenario, with each pair of columns showing the counts data in the left column and the reconstruction in the right column. The rows show scenarios D1–D5 as described in Section 4.1. All results are available at <https://jolideco.github.io/jolideco-comparison>.

Table 2

Image Metrics for the Different Signal-to-noise Ratios Shown in Figure 6

Scenario	Bkg 1	Bkg 2	Bkg 3
D1	0.99/0.70	0.99/0.73	0.99/0.71
D2	0.99/0.64	0.98/0.68	0.98/0.76
D3	0.98/0.68	0.98/0.69	0.97/0.72
D4	0.94/0.60	0.94/0.61	0.91/0.69
D5	0.88/0.79	0.89/0.74	0.89/0.73

Note. The first number in each cell shows the structural similarity index (SSI; the closer to unity, the better) and the second number shows the normalized root mean square (NRMSE; the lower the better).

reconstruction of the spiral structure in the center shows visual artifacts, with the individual arms not completely separated. The joint reconstruction shows the positive qualities of both individual reconstructions. This confirms the results from Ingaramo et al. (2014), where the authors found that combining multiple measurements for deconvolution with RL to be a “general tool for combining images with complementary strengths.”

5. Application Examples

5.1. Deep Chandra Observation

5.1.1. Image Example

We apply Jolideco to a set of short-exposure Chandra observations of the SNR 1E 0102.2–7219 and reconstruct a single long-exposure image using the patch prior learned from JWST data. In total, we chose 25 observations taken with the ACIS-S instrument and no grating. We chose the data taken after 2006 when the contamination layer results in lower pileup. The list of selected observations including their exposure duration is shown in Table 3 of Appendix B.

We reduced the data using CIAO v4.15 (Fruscione et al. 2006) and a custom Snakemake workflow (see also Section 6 on reproducibility). For the counts image we selected an energy range from 0.5 keV to 7 keV and a pixel size of 0.25 of the native pixel size of the Chandra ACIS detector. This corresponds to an absolute pixel size of $\approx 0''.125 \text{ pixel}^{-1}$ or a sampling of ≈ 2 pixels for the 38% containment radius of Chandra’s PSF. Chandra’s instrument response varies with the observation offset and energy, so we simulated the PSF for each observation independently using

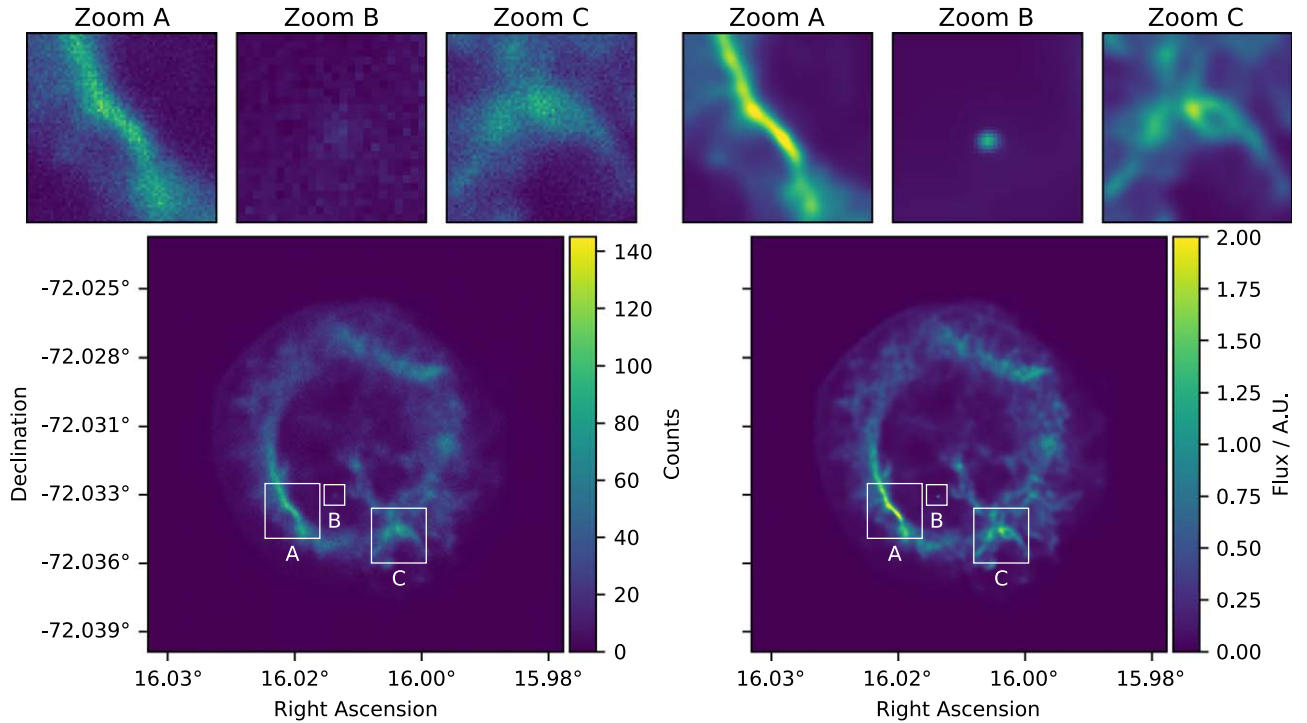


Figure 8. *Jolideco* result with the JWST patch prior for Chandra observations of SNR 1E 0102.2–7219. The bottom-left image shows the summed counts from 25 observations. The bottom-right image shows the the *Jolideco* deconvolution. The flux is given as a counts rate relative to ObsID 1308. The panels Zoom A, Zoom B, and Zoom C show a zoomed-in version of three example regions in the SNR, marked with the white rectangles in the large images. The color map and scale used in the zoom insets is the same as used for the larger images.

the *marx* tool (Davis et al. 2012) and the *simulate_psf* script.¹² As a reference spectrum for the PSF simulation we assumed the model derived by Plucinsky et al. (2017). We converted the spectral script from XSPEC to Sherpa using the helper script *convert_xspec_script*.

We assume a spatially constant background of 1×10^{-5} cts pixel⁻¹ but allow the normalization to vary in the reconstruction process. We assume the exposure to be spatially constant and express its absolute value for each observation as the ratio of observation time with respect to a reference observation, namely ObsID 8365. For *Jolideco* we use the GMM prior learned on the JWST Cas A image. As there are many counts in the image we additionally use an oversampling factor of 2, meaning the image x is computed on a finer grid and downsampled after convolution with the PSF to the same grid as the counts data. The patch prior is also evaluated on the finer grid.

We allow the background level and position calibration (see Equation (13)) to vary among the observations. As a reference for the positional shift we also use ObsID 8365. Figure 8 shows the final result. *Jolideco* clearly both improves the angular resolution and reduces the noise of the counts image.

Zoom A of Figure 8 details an example of a filament structure. *Jolideco* maintains the extended structure along the filament, while sharpening in the perpendicular direction. Looking at the upper-right corner one can also see that weak background emission is suppressed, improving the contrast of the image.

Zoom B shows the region around the putative central compact object of the SNR (Vogt et al. 2018; Hebbar et al. 2020, but

Long et al. 2020). While the emission is barely visible in the counts image, it is clearly evident above the background in the *Jolideco* reconstruction. Visually, the reconstruction does not result in a point source, but leaves a visible extension of a few pixels.

However, at this level of analysis and signal-to-noise, both scenarios remain plausible: a real intrinsic extension of the feature, or a real point source with a residual extension in its reconstruction, owing to the low signal-to-noise. We observed the second possibility in the simulated point-source data set discussed in Section 4.1.¹³

To draw stronger scientific conclusions a dedicated analysis of the region, including a simulation of the false-detection ratio of point sources, would be needed. We estimate the position of the feature by finding the brightest pixel in the map, obtaining $\alpha_{J2000} = 01^{\text{h}}04^{\text{m}}02.^{\text{s}}75$ and $\delta_{J2000} = -72^{\circ}02'00''.19$, which is consistent with the value reported by Long et al. (2020). This demonstrates an absolute feature position that was reconstructed with *Jolideco* which is consistent with the existing literature.

Zoom C shows a circular filament structure of the SNR. In the counts data the structure is barely visible, while the *Jolideco* reconstruction shows a distinct circular structure with a locus of bright emission. This demonstrates the ability of *Jolideco* to reconstruct sub-PSF-scale structures with astrophysical relevance.

Checking the residual images per observation (not shown here), we notice that the assumption of time independence of the source and the instrument response is not completely

¹² See Section 4.2.3 of the Chandra X-ray Center & Chandra Project Science's *The Chandra Proposers' Observatory Guide*, available online: http://cxc.cfa.harvard.edu/proposer/POG/html/chap4.html#th_sEc4.2.3.

¹³ We note that, as illustrated in Section 4.2 for scenario A, the remaining extension of the point source after deconvolution may be smaller if the source is brighter.

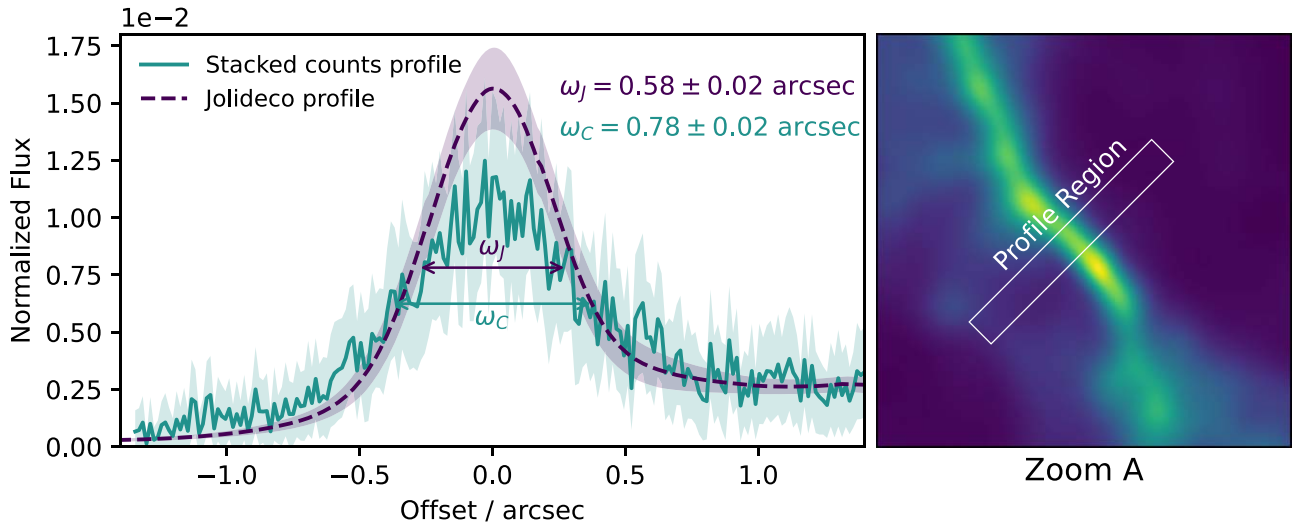


Figure 9. The left panel shows the deconvolved flux and counts profile of the region illustrated in the right panel. The thick solid and dashed lines represent the mean across the 100 bootstrap-resampled data sets. The counts profile for each bootstrapped data set is summed (stacked) across the 25 observations per data set, while the Jolideco profile is computed from a single joint reconstruction based on all 25 observations. The variability across the 100 bootstrapped data sets is represented by the 3σ error bands, shown transparent in the background for both profiles. Both profiles are normalized such that they integrate to unity and are aligned such that the position of their peak coincides.

fulfilled in this case. The detection efficiency of Chandra changed and the substructure in the remnant expanded over the time of the observations. Both effects are previously known and have been reported, for example, in Xi et al. (2019). Despite the clear improvement in spatial resolution, systematic issues such as the proper motion of structures and calibration uncertainties might limit the quality of the result. The full analysis example, with results and residuals is available on GitHub.¹⁴

5.1.2. Flux Profile

To quantify the improvement in angular resolution of the reconstructed image over the raw data we compute a flux profile on the Jolideco (with the JWST patch prior) reconstruction of 1E 0102. We chose a region perpendicular to the filament shown in Zoom A of Figure 8 of size 128×128 pixels. We compute errors via the bootstrap, with resampled data obtained by sampling the event lists with replacement for each individual observation. For each resampled data set (composed of 25 observations) we run Jolideco and compute a binned flux profile on the reconstructed image. In total, we sample 100 different realizations, where each reconstruction runs for about 10 minutes on a standard M1 CPU. In this way, we obtain a set of profiles which we use to compute the mean and error per bin of the profile. For comparison, we also compute the profiles and associated errors directly on the bootstrapped counts summed across the 25 observations (without processing or fitting). The result is shown in Figure 9.

For each bootstrapped data set, we then compute the FWHM of the profiles by fitting a Gaussian shape with a linear gradient as a baseline, for both the counts profile ($\omega_C = 0.''78 \pm 0.''02$) and the Jolideco-deconvolved intensities ($\omega_J = 0.''58 \pm 0.''02$), where the uncertainty is represented by the standard deviations in the bootstrap sample. Note that a significant narrowing of the cross-sectional profile is seen. The width of the filament,

however, is definitively $>0''$, suggesting that it is resolved in the Jolideco reconstruction.¹⁵

5.2. Fermi-LAT Event Types

As a second application example, we apply Jolideco to data from the Fermi-LAT (Atwood et al. 2009). The LAT is a pair-conversion gamma-ray telescope on board the Fermi satellite. It operates in the energy range between 30 MeV and 300 GeV. For this example we chose a Galactic source with complex morphology: the SNR Vela Junior. We use 14 yr (from 2008 August 4 to 2023 August 4) of “P8R3_SOURCE” photons with reconstructed energy in the range 10 GeV–2 TeV and applied a zenith angle cut of 105° . This is consistent with the data selection range in Ackermann et al. (2017). To preserve as much information as possible on the angular resolution of the instrument, we split the data into the four event types “PSF0,” “PSF1,” “PSF2,” and “PSF3,” as defined by the Fermi-LAT “Pass8” data release, where PSF0 represents the subset of data with the worst and PSF3 the best angular resolution.¹⁶ Instead of different observations, the joint likelihood is computed across those different event types.

We chose a pixel size of 0.02 deg, which corresponds to approximately one-fifth of the size of the PSF in the best event class. For the first part of the data reduction we use the standard FermiTools v2.2.0 (Fermi Science Support Development Team 2019).

¹⁵ The projected radial profiles of the on-axis Chandra PSF (see the Chandra/MARX ray-trace assessment at <https://space.mit.edu/cxc/marx/tests/PSF.html>) show that the 39% enclosed counts fraction radius is $\approx 0.''26$. Approximating the PSF as a two-dimensional Gaussian, this corresponds to a 1σ radius and a FWHM $\omega_{\text{PSF}} \approx 0.''62$. Since the observed profile is expected to result from a convolution of the PSF with the true profile, we expect

$$\omega_C^2 = \omega_{\text{PSF}}^2 + \omega_J^2.$$

A true filament width of $\gtrsim 0.''5$ is thus necessary to match the measured ω_C . That the estimated ω_J is larger by $\approx 0.''1$ can be attributed to the presence of several systematic confounding factors, such as the PSF being heavier tailed than a Gaussian, and the intrinsic profile of the filament not being Gaussian.

¹⁶ https://fermi.gsfc.nasa.gov/ssc/data/analysis/documentation/Pass8_usage.html

¹⁴ <http://github.com/jolideco/jolideco-chandra-examples>

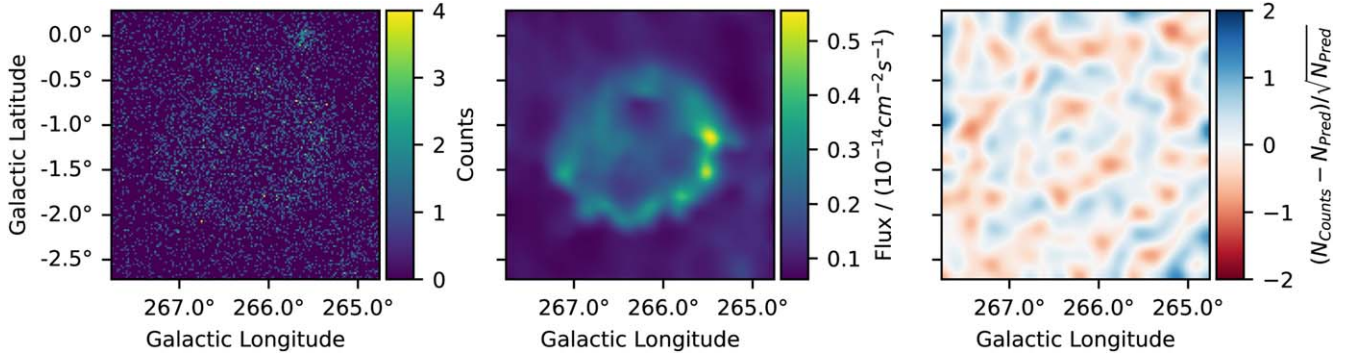


Figure 10. Illustration of the `Jolideco` result for the Fermi-LAT data of the SNR RX J0852.0-4622 or Vela Junior. The left image shows the counts above 10 GeV. The different event classes are stacked into a single image with bin size 0.02° . The image in the center shows the flux reconstructed by the `Jolideco` method. The image on the right shows the standardized residuals as computed by Equation (15) and smoothed with a Gaussian kernel of width 5 pixels $\approx 0.1^\circ$, similar to the size of the PSF. More information on this analysis example can be found on <https://github.com/jolideco/jolideco-fermi-examples>.

The Fermitools produce energy-dependent counts, exposure, and PSF cubes. To further reduce the data to images we use the Gammapy package v1.0 (Acero et al. 2023; Donath et al. 2023). To compute an image of the expected background emission we rely on the latest Galactic diffuse emission model `gll_iem_v07.fits`, and also include nearby sources from the 3FHL catalog (Ajello et al. 2017). We first fit the norm and spectral indices of all the 3FHL sources and the background model to the data using Gammapy, excluding a circular region of radius 1.02 deg centered on Vela Junior. Finally, we reduce this model to an image by summing over the energy dimension. To compute the exposure and PSF image we assume a power law with index $\Gamma = 1.77$ (Ajello et al. 2017) as a spectral model for Vela Junior and integrate over the energy dimension of the cube weighting according to this spectral model.

We chose the GMM patch prior learned from Chandra SNRs described in Section 3.2.4. Because of the limited statistics we do not use oversampling but keep the pixel size of the counts image. Figure 10 shows the data, the `Jolideco` results, and corresponding residuals. The Fermi-LAT data are sparse, with limited signal counts. `Jolideco` clearly improves the noise level as well as angular resolution compared to the data. The calibration parameters of the predicted counts converged to different norms for the background model and positional shifts of order of <1 pixel for the different event types. In this way, `Jolideco` corrects for systematic errors in the reconstruction of the position. The reconstruction includes background emission around Vela Junior, indicating that the Galactic diffuse background model is either underestimated in brightness and/or associated with systematic errors. The background estimate could be improved by choosing a larger region of interest. However, this would also require a larger effort of modeling nearby sources, which we avoided here.

To further check the quality of the reconstructed flux we compute a map of smoothed standardized residuals, defined by

$$\rho(\mathbf{D}, \boldsymbol{\lambda}) \approx 2\sqrt{\pi\sigma^2} \frac{(\mathbf{D} - \boldsymbol{\lambda}) \otimes g_\sigma}{\sqrt{\boldsymbol{\lambda} \otimes g_\sigma}}, \quad (15)$$

where $\mathbf{D}$ is the measured counts image, $\boldsymbol{\lambda}$ is the expected counts image, and g_σ is a Gaussian smoothing kernel (normalized to integrate to unity) with width $\sigma = 5$ pixels. We apply the smoothing to enhance possible structures in the sparse data, while the normalization factor ensures the

smoothed residual pixel values are still distributed following a unit Gaussian distribution.

The result is shown in the right panel of Figure 10. The image shows no systematic effects and a spatially homogeneous distribution of the residual noise. The resulting distribution of residual significances per pixel is centered at $\mu = -0.001$ with a width of $\sigma = 0.991$, indicating a close agreement with a standard normal distribution, as expected for a good model of the emission. In Figure 11, we also show the residuals per PSF class with and without the correction defined by Equation (13). More details of the Fermi-LAT analysis including results and residuals are available on GitHub.¹⁷

6. Paper Reproducibility

In order to facilitate reproducibility we provide the source code for the numerical results and figures presented in this article on GitHub; a copy has been deposited in Zenodo: DOI:10.5281/zenodo.13176177.¹⁸ The results in this paper can be reproduced using the tool `Showyourwork` (Luger 2021) and following the instructions in the `README.md` file in the same repository. All data are available for download from Zenodo or can be reproduced using the custom `Snakemake` workflows and repositories provided on the `Jolideco` GitHub page.¹⁹ For the Fermi-LAT data reduction we used a custom `Snakemake` workflow, also available on GitHub.²⁰ For the Chandra data reduction we used a custom `Snakemake` workflow that is again also available on GitHub.²¹

7. Summary and Outlook

In this article we presented a new method for deconvolution and denoising of images contaminated by Poisson noise. Using multiple observations of the same sky region the method reconstructs a single flux image using the MAP estimate of the joint posterior distribution under a patch-based image prior. We show that the method leads to improved results compared to existing methods in a variety of morphological, signal-to-noise, and instrument scenarios. Improvement is especially pronounced for extended sources of a size similar to the PSF. In

¹⁷ <https://github.com/jolideco/jolideco-fermi-examples>

¹⁸ <https://github.com/jolideco/jolideco-paper>

¹⁹ <https://github.com/jolideco>

²⁰ <https://github.com/adonath/snakemake-workflow-fermi-lat>

²¹ <https://github.com/adonath/snakemake-workflow-chandra>

this case, *Jolideco* can reliably recover features that are well beyond the angular resolution of the instrument. We also show that the approach of jointly analyzing multiple observations leads to improved reconstruction results. Including data with lower angular resolution in the *Jolideco* reconstruction improves the fitted correlation structure among pixels across larger distances, while not negatively affecting the overall sharpness of the reconstruction.

We imagine that *Jolideco* is the first step of a larger project building ML-inspired methods for the reconstruction of astronomical images that are dominated by Poisson noise. While *Jolideco* has already demonstrated excellent reconstruction quality in multiple scenarios, it still has significant potential for future extension and improvements.

The first category of future improvements is related to the prior handling. For example, more general mixture models such as the Student-*t* or a generalized GMM can be used to model the distribution of patches. van den Oord & Schrauwen (2014) shows that these more flexible models can account for the tails of the patch distribution and lead to improved results on natural images. We expect these improvements to transfer to astronomical images.

To allow the patch-based prior distribution to capture correlations at larger scales, Papayan & Elad (2016) proposed to combine patches extracted at multiple scales of the image when training the GMM. This is a promising approach to further improve the reconstruction quality for extended sources with low signal-to-noise ratios.

Working in the context of computer tomographic image reconstruction, Altekriiger et al. (2022) used NFs in their formulation of the patch prior in order to further improve the general expressiveness of the prior. As with a GMM, we can evaluate the NF-based prior distribution and thus construct an explicit posterior distribution for optimization and computation of the MAP estimate. Ulyanov et al. (2017) introduced another promising prior distribution, known as the “deep image prior,” and showed that prior structure for an image can be built into a deep neural network as an “inductive bias,” where the architecture of the network encodes the prior information on the correlation structure among neighboring pixels. The authors showed that the deep image prior successfully imposes self-similarity on structures in the image. As structures in astronomical images often show self-similarity, this is an interesting approach that in principle could be combined with the patch prior.

A second category of improvement to *Jolideco* is generalizing it to jointly model the energy and spatial domains, thus enabling hyperspectral image reconstruction. Instruments such as Fermi-LAT combine instrument responses that vary strongly with energy along with a limited overall energy resolution. This configuration introduces correlations between images in neighboring energy bands. Accounting for energy data in the reconstruction would allow us to handle these correlations and also open up the possibility of merging multiple X-ray data sets in the joint spectral/spatial domain, thus improving both angular and energy resolution.

Jolideco can already handle an independent background component. However, we plan to extend *Jolideco* to enable it to handle multiple source components at the same time. This allows us in principle, as shown by Selig & Enßlin (2015) and Pompe et al. (2018), to separate extended sources from point-like sources, based on different prior assumptions on the spatial correlation and distribution of pixels for each model component.

A last category of improvement to *Jolideco* is computing principled error estimates for the reconstruction. In this paper we use bootstrap methods. Since we can compute the full posterior distribution, however, we could build an approximation to the covariance matrix of the MAP estimate, for example, by using a low-rank or sparse representation. Alternatively, we can use MCMC to obtain a sample from the full posterior distribution, enabling us to explore the distribution and quantify uncertainty both in the parameters and functions of the parameters constructed to represent structures in the image. Estimating errors for the reconstruction would also include propagating systematic uncertainties on the PSF to the reconstructed image and handling a PSF that varies across the field of view, as recently proposed by Sakai et al. (2023).

Finally, we invite the community of X-ray and gamma-ray astronomers to use *Jolideco* to combine data sets and make new scientific discoveries, beyond what is possible with the angular resolution of existing instruments. We also welcome contributions and feedback to the public *Jolideco* code and validation repositories.

Acknowledgments

This work was conducted under the auspices of the CHASC International Astrostatistics Center. CHASC is supported by NSF grants DMS-21-13615, DMS-21-13397, and DMS-21-13605; by the UK Engineering and Physical Sciences Research Council (EP/W015080/1); and by NASA APRA grant No. 80-NSSC21-K0285.

We thank CHASC members for many helpful discussions, especially Xiao-Li Meng and Katy McKeough. D.v.D. was also supported in part by a Marie-Skłodowska-Curie RISE grant (grant No. H2020-MSCA-RISE-2019-873089) provided by the European Commission.

Aneta Siemiginowska, Vinay Kashyap, and Douglas Burke further acknowledge support from NASA contract to the Chandra X-ray Center NAS8-03060.

Some of the computations in this paper were conducted on the Smithsonian High Performance Cluster (SI/HPC), Smithsonian Institution (<https://doi.org/10.25572/SIHPC>).

This research has made use of data obtained from the Chandra Data Archive, contained in DOI:10.25574/cdc.244. This research also made use of software provided by the Chandra X-ray Center (CXC) in the application packages CIAO (Fruscione et al. 2006) and Sherpa (Freeman et al. 2001; Burke et al. 2022).

Appendix A

Details on Source Patterns Used in Numerical Experiments

We have devised a series of elaborate simulation scenarios consisting of point sources and extended sources of different shapes, sizes, orientations, and spacings in order to test deconvolution algorithms for astronomical low-count images. There are four distinct patterns, illustrated in Figure 4, which we denote (1) “Pointilist,” (2) “Asterism,” (3) “Shield,” and (4) “Spiral.” The construction of each pattern is described below. All patterns are placed in a 128×128 pixel image.

A.1. Scenario A1: “Pointilist”

The first scenario consists of a series of point sources arranged in a grid with increasing horizontal and vertical separations. The sources placed along a horizontal row all have the same strength, which increases with y-axis location. The

source strengths are

$$s_A = \{10, 20, 30, 50, 80, 130, 210, 340\} \cdot \tau \text{ counts}$$

for

$$y = [8, 10, 13, 18, 26, 39, 55, 76] \text{ pix},$$

where τ denotes an arbitrary scaling factor, set to 1 for Chandra-like simulations and to 5 for XMM-like simulations. The x -axis locations of the sources for a given y are

$$x = [8, 9, 10, 11, 13, 16, 21, 29, 42, 61, 95] \text{ pix}.$$

The spatial separation and intensity changes are based on the Fibonacci series, which places them on a cadence that is sub-logarithmic but supra-linear, and thus provides a good trade-off between dynamic range and point density. Notice that at the lower-left corner of the image, adjacent sources are placed closer together, either in the adjoining pixel or separated by 1 pixel, with low source strengths, posing an almost impossible challenge to deconvolution algorithms. In contrast, the upper-right corner of the image is sparse, with strong well-separated sources.

A.2. Scenarios B1–B3: “Asterism”

The pattern for these scenarios includes a diffuse Gaussian source ($\sigma = 2$ pixels) at the center of the image with pixel location (64, 64). Four larger extended sources (Gaussians with $\sigma = 4$ pixels) with the same number of counts as the central source are set at each of the corners. Four point sources with intensities decreasing by a factor $1/3$ are arranged around the central diffuse source to the top, left, right, and bottom. The three signal-to-noise ratios (indicated by the scenario number) are determined by the counts in the diffuse sources, all of which are set to one of the values in

$$\{B1, B2, B3\} := \{100, 500, 1000\} \cdot \tau \text{ counts/pix},$$

where $\tau = 1$ for Chandra-like and $\tau = 5$ for XMM-like simulations, as above. The diffuse sources are located at

$$(x, y)_{\text{diffuse}} = \{(22, 22), (22, 106), (106, 106), (106, 22)\} \text{ pix},$$

and the point sources at

$$(x, y)_{\text{point}} = \{(64, 76), (44, 64), (74, 64), (64, 44)\} \text{ pix},$$

with counts

$$\left\{1, \frac{1}{3}, \frac{1}{3^2}, \frac{1}{3^3}\right\} \times$$

those in the diffuse sources, respectively. The pattern of point sources explores the effects of contamination by large extended sources, while the diffuse sources at the corners help to characterize the edge effects in deconvolutions.

A.3. Scenarios C1–C3: “Shield”

These scenarios consists of a flat disk at the image center, with superposed point sources and jet-like linear extensions radiating to the corners. The disk has a radius of 20 pixels. The overall signal-to-noise ratio of each of the three scenarios is determined by the disk’s brightness, which is set to one of the values in

$$\{C1, C2, C3\} := \{0.5, 1, 2\} \cdot \tau \text{ counts pix}^{-1},$$

where again τ represents an arbitrary scaling factor to mimic telescope sensitivity. Four point sources are placed at positions displaced at $\{\pm 8, \pm 8\}$ pixels from image center, with intensities set to

$$\{10, 50, 100, 500\} \cdot \tau \text{ counts}$$

clockwise starting from the lower left, respectively. Linear sources, each with a horizontal width of 3 pixels, are placed along the $\pm 45^\circ$ lines extending from radial distances of 36 pixels to 64 pixels from the image center, with surface brightnesses of

$$\{10, 50, 100, 500\} \cdot \tau \text{ counts pix}^{-1},$$

matching those of the nearest point sources. This pattern simultaneously tests the ability of deconvolution algorithms to distinguish the disk from the background, and the embedded point sources from the disk. The linear radiating sources present a different difficult challenge to algorithms which use symmetrical kernels, which usually break such features into a string of point-like sources. (We emphasize that such features are often seen in real astronomical images as jets or detector features like streaks or bad CCD columns.)

A.4. Scenarios D1–D5: “Spiral”

These scenarios have four major components. First is a double spiral structure, centered at $(-10, 0)$ pixels, covering 500° , and with a maximum radial extent of 24 pixels and a width of 3 pixels. Second is a disk of radius 4, centered at $(-10, 0)$ pixels. Third are two vertical rows of point sources located at x -axis positions of 40 pixels and 88 pixels, with y -axis locations spanning $\{\pm 7, \pm 21\}$ pixels. Fourth is an annulus that encloses all these elements; it is also centered at $(-10, 0)$ pixels, with inner radius 52 pixels and outer radius 54 pixels. The point sources have intensities of $\{25, 50, 100, 200\}$ counts increasing with vertical height. The overall signal-to-noise ratio of the scenarios is determined by the surface brightnesses of the spiral, disk, and annulus, which are set to one of the values in

$$\{D1, D2, D3, D4, D5\} := \{0.5, 1, 2, 10, 1000\} \cdot \tau \text{ counts}.$$

This configuration tests the recovery of closely located diffuse structures (close spiral arms), field-of-view edge effects (annulus that approaches close to the left edge of the image), and the visibility of point sources close to diffuse structures. The case with the surface brightness of $1000 \text{ cts pixel}^{-1}$ tests the asymptotic limit of the high-counts case. Since only the diffuse emission components have the enhanced brightness, this scenario also demonstrates the benefits of a sharp PSF; broad PSFs contaminate the diffuse emission around the point sources, which results in a significantly higher difficulty in recovering faint sources.

Appendix B Additional Tables and Figures

In this section we present an additional figure (Figure 11, illustrating the deconvolution of Fermi-LAT data of RX J0852.0–4622) and table (Table 3, describing the characteristics of the analyzed data set).

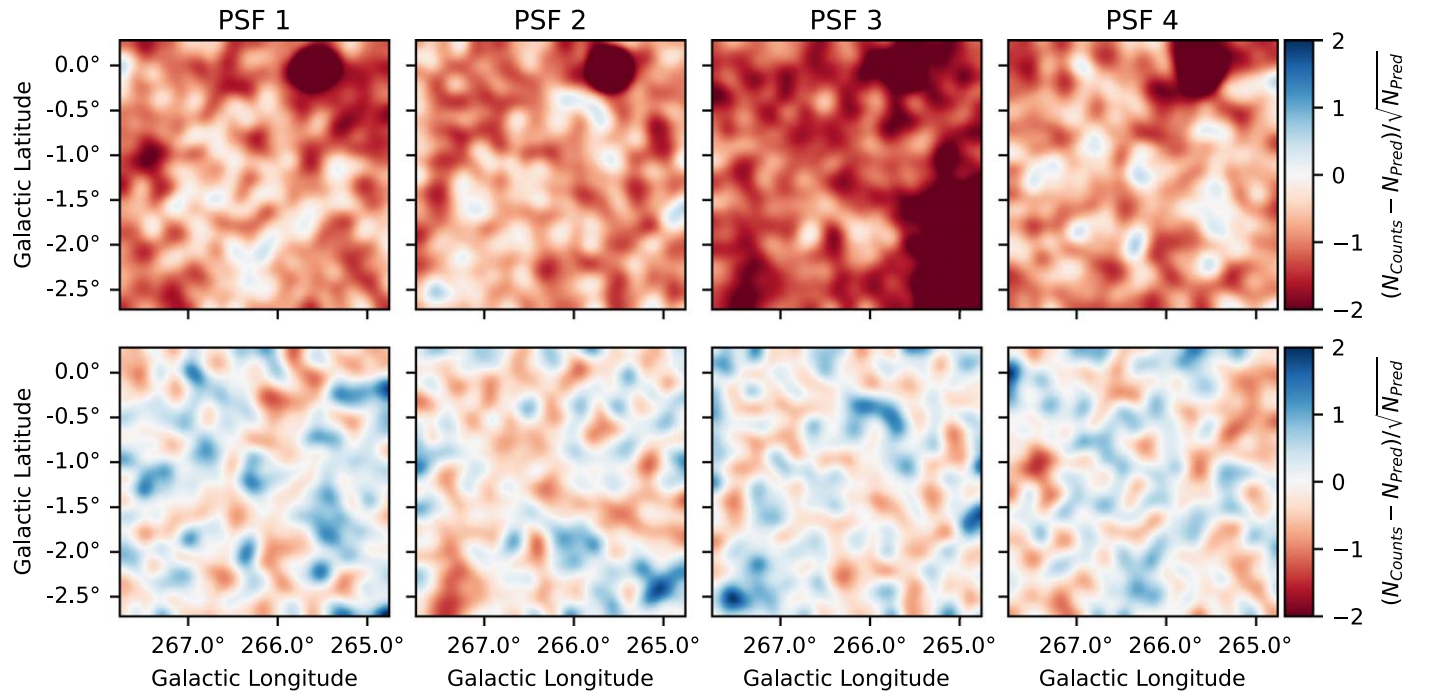


Figure 11. Illustration of the residuals of the Fermi-LAT analysis example of RX J0852.0–4622 from Section 5.2. All images show the standardized residuals as computed by Equation (15) and smoothed with a Gaussian kernel of width 5 pixels $\approx 0.1^\circ$, similar to the size of the PSF. The upper row shows the result without calibration, the lower row shows the result with the calibration applied, according to Equation (13). The columns represent the four different PSF event classes of Fermi-LAT. The color scale is shown in the right panel. More information on this analysis example can be found on <https://github.com/jolideco/jolideco-fermi-examples>.

Table 3

Observation List and Details from the Chandra 1E 0102.2–7219 Example Analysis in Section 5.1

ObsID	Obs Date	Exposure (ks)	Rel. Exposure	Frac. Exposure (%)
8365	02/11/2007	21.0	1.0	6.1
6766	06/06/2006	19.7	0.9	5.8
9694	02/07/2008	19.2	0.9	5.6
15467	01/28/2013	19.1	0.9	5.6
14258	01/12/2012	19.0	0.9	5.6
13093	02/01/2011	19.0	0.9	5.6
11957	12/30/2009	18.4	0.9	5.4
6759	03/21/2006	17.9	0.9	5.2
17380	02/28/2015	17.7	0.8	5.2
26987	03/21/2023	14.4	0.7	4.2
25618	03/25/2022	14.4	0.7	4.2
22805	02/08/2020	14.3	0.7	4.2
21804	03/17/2019	14.3	0.7	4.2
20639	03/12/2018	14.3	0.7	4.2
19850	03/19/2017	14.3	0.7	4.2
18418	03/15/2016	14.3	0.7	4.2
24577	02/10/2021	13.8	0.7	4.0
17688	07/17/2015	9.6	0.5	2.8
16589	03/27/2014	9.6	0.5	2.8
6758	03/19/2006	8.1	0.4	2.4
10656	03/06/2009	7.8	0.4	2.3
6765	03/19/2006	7.6	0.4	2.2
10654	03/01/2009	7.3	0.3	2.1
10655	03/01/2009	6.8	0.3	2.0

Note. The “Rel. Exposure” columns list the exposure for each observation expressed as the ratio with the exposure of the reference observation with ObsID 8365. The “Frac. Exposure” column gives the exposure of each observation as a fraction of the total exposure.

ORCID iDs

Axel Donath <https://orcid.org/0000-0003-4568-7005>
 Aneta Siemiginowska <https://orcid.org/0000-0002-0905-7375>
 Vinay L. Kashyap <https://orcid.org/0000-0002-3869-7996>
 David A. van Dyk <https://orcid.org/0000-0002-0816-331X>
 Douglas Burke <https://orcid.org/0000-0003-4428-7835>

References

- Ajero, F., Aguasca-Cabot, A., Buchner, J., et al. 2023, *Gammapy: Python toolbox for gamma-ray astronomy*, v1.0.1, Zenodo, doi:[10.5281/zenodo.7734804](https://doi.org/10.5281/zenodo.7734804)
- Ackermann, M., Ajello, M., Baldini, L., et al. 2017, *ApJ*, **843**, 139
- Ajello, M., Atwood, W. B., Baldini, L., et al. 2017, *ApJS*, **232**, 18
- Akhaury, U., Starck, J.-L., Jablonka, P., Courbin, F., & Michalewicz, K. 2022, *FrASS*, **9**, 9
- Altekrüger, F., Denker, A., Hagemann, P., et al. 2023, *InvPr*, **39**, 064006
- Astropy Collaboration 2018, *ApJ*, **156**, 123
- Atwood, W. B., Abdo, A. A., Ackermann, M., et al. 2009, *ApJ*, **697**, 1071
- Bouman, K. L., Johnson, M. D., Zoran, D., et al. 2016, *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)* (New York: IEEE)
- Burke, D., Laurino, O., & Wmclaugh, 2022, *sherpa/sherpa: Sherpa*, v4.15.0, Zenodo, doi:[10.5281/zenodo.7186379](https://doi.org/10.5281/zenodo.7186379)
- Cash, W. 1979, *ApJ*, **228**, 939
- Coifman, R. R., & Donoho, D. L. 1995, in *Translation-Invariant De-Noising*, ed. A. Antoniadis & G. Oppenheim (New York: Springer), 125
- Connors, A., Stein, N. M., van Dyk, D., Kashyap, V., & Siemiginowska, A. 2011, in *ASP Conf. Ser. 442, Astronomical Data Analysis Software and Systems XX*, ed. I. N. Evans et al. (San Francisco, CA: ASP), 463
- Davis, J. E., Bautz, M. W., Dewey, D., et al. 2012, *Proc. SPIE*, **8443**, 84431A
- Dey, N., Blanc-Feraud, L., Zimmer, C., et al. 2006, *MRT*, **69**, 260
- Donath, A. 2024, *Jolideco: a Python library for Joint Likelihood deconvolution*, v0.3, Zenodo, doi:[10.5281/zenodo.11267661](https://doi.org/10.5281/zenodo.11267661)
- Donath, A., Siemiginowska, A., Kashyap, V., et al. 2022, in *Proc. of the 21st Python in Science Conf.*, ed. M. Agarwal et al. (Austin, TX: SciPy), 98
- Donath, A., Terrier, R., Remy, Q., et al. 2023, *A&A*, **678**, A157

- Esch, D. N., Connors, A., Karovska, M., & van Dyk, D. A. 2004, *ApJ*, **610**, 1213
- Fermi Science Support Development Team, 2019 FermiTools: Fermi Science Tools, Astrophysics Source Code Library, ascl:1905.011
- Fish, D. A., Brinicombe, A. M., Pike, E. R., & Walker, J. G. 1995, *J. OSAJ*, **12**, 58
- Freeman, P., Doe, S., & Siemiginowska, A. 2001, *Proc. SPIE*, **4477**, 76
- Fruscione, A., McDowell, J. C., Allen, G. E., et al. 2006, *Proc. SPIE*, **6270**, 62701V
- Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, *Natur*, **585**, 357
- Hebbbar, P. R., Heinke, C. O., & Ho, W. C. G. 2020, *MNRAS*, **491**, 1585
- Hunter, J. D. 2007, *CSE*, **9**, 90
- Hurley-Walker, N., Galvin, T. J., Duchesne, S. W., et al. 2022, *PASA*, **39**, e035
- Ingaramo, M., York, A., Hoogendoorn, E., et al. 2014, *ChemPhysChem*, **15**, 794
- Kingma, D. P., & Ba, J. 2014, arXiv:1412.6980
- Long, X., Gaetz, T. J., & Plucinsky, P. P. 2020, *ApJ*, **904**, 70
- Lucy, L. B. 1974, *AJ*, **79**, 745
- Luger, R. 2021, showyourwork, GitHub, <https://github.com/rodluger/showyourwork>
- Martin, D., Fowlkes, C., Tal, D., & Malik, J. 2001, Proc. 8th Int. Conf. Computer Vision, Vol. 2 (New York: IEEE), 416
- McLachlan, G., & Peel, D. 2004, Finite Mixture Models, Wiley Series in Probability and Statistics (New York: Wiley), https://books.google.com/books?id=c2_fAox0DQoC
- Papayan, V., & Elad, M. 2016, *ITIP*, **25**, 249
- Parameswaran, S., Deledalle, C.-A., Denis, L., & Nguyen, T. Q. 2018, *ITIP*, **28**, 687
- Paszke, A., Gross, S., Massa, F., et al. 2019, in Advances in Neural Information Processing Systems 32, ed. H. Wallach et al. (New York: Curran Associates, Inc.), 8024, <http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf>
- Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, *JMLR*, **12**, 2825
- Plucinsky, P. P., Beardmore, A. P., Foster, A., et al. 2017, *A&A*, **597**, A35
- Pumpe, D., Reinecke, M., & Enßlin, T. A. 2018, *A&A*, **619**, A119
- Reeves, S. J. 1995, *IJST*, **6**, 387
- Richardson, W. H. 1972, *JOSA*, **62**, 55
- Sakai, Y., Yamada, S., Sato, T., et al. 2023, *ApJ*, **951**, 59
- Schawinski, K., Zhang, C., Zhang, H., Fowler, L., & Santhanam, G. K. 2017, *MNRAS: Letters*, **467**, L110
- Selig, M., & Enßlin, T. A. 2015, *A&A*, **574**, A74
- Selig, M., Vacca, V., Oppermann, N., & Enßlin, T. A. 2015, *A&A*, **581**, A126
- Starck, J.-L., Nguyen, M. K., & Murtagh, F. 2003, *SigPr*, **83**, 2279
- Sweere, S. F., Valtchanov, I., Lieu, M., et al. 2022, *MNRAS*, **517**, 4054
- Ulyanov, D., Vedaldi, A., & Lempitsky, V. S. 2017, *Int. J. Comput. Vis.*, **128**, 1867
- van den Oord, A., & Schrauwen, B. 2014, *JMLR*, **15**, 2061, <http://jmlr.org/papers/v15/vandenoord14a.html>
- van der Walt, S., Schönberger, J. L., Nunez-Iglesias, J., et al. 2014, *PeerJ*, **2**, e453
- Vogt, F. P. A., Bartlett, E. S., Seitzzahl, I. R., et al. 2018, *NatAs*, **2**, 465
- Wang, Z., Bovik, A., Sheikh, H., & Simoncelli, E. 2004, *ITIP*, **13**, 600
- Westerkamp, M., Eberle, V., Guardiani, M., et al. 2024, *A&A*, **684**, A155
- Xi, L., Gaetz, T. J., Plucinsky, P. P., Hughes, J. P., & Patnaude, D. J. 2019, *ApJ*, **874**, 14
- Xu, L., Ren, J. S., Liu, C., & Jia, J. 2014, in Advances in Neural Information Processing Systems, ed. Z. Ghahramani et al. (New York: Curran Associates, Inc.) <https://proceedings.neurips.cc/paper/2014/file/1c1d4df596d01da60385f0bb17a4a9e0-Paper.pdf>
- Zhang, K., Ren, W., Luo, W., et al. 2022, *Int. J. Comput. Vis.*, **130**, 2103
- Zoran, D., & Weiss, Y. 2011, Int. Conf. Computer Vision (New York: IEEE), 479
- Zoran, D., & Weiss, Y. 2012, Proceedings of the 25th International Conference on Neural Information Processing Systems—Volume 2, NIPS'12 (Red Hook: Curran Associates Inc.), 1736