Joseph Piacentine was senior at California State University, Chico, majoring in Physics, when attending a summer internship at the Stanford Linear Accelerator Center, where he worked on the detection of galaxy clusters with the use of X-ray telescopes. This work was presented at the AAAS National Conference in 2005. He has since graduated and moved to Pasadena, where he hopes to pursue a Master's Degree in Astronautical Engineering.

Phil Marshall is a postdoctoral research fellow at the Kavli Institute for Particle Astrophysics and Cosmology, Stanford University. Primarily based at SLAC, he works on data analysis techniques in optical, radio and X-ray astronomy, focusing on clusters of galaxies and gravitational lenses and insisting on probabilistic methodology throughout. Phil obtained a Masters degree in Physics at the University of Cambridge, and then remained at the Cavendish Laboratory for his PhD in astrophysics (entitled "Bayesian Analysis of Clusters of Galaxies"). Since October 2003 he has continued this work, and branched out into both investigating the available strong lensing information in large-scale optical surveys such as SNAP and LSST, and the analysis of X-ray spectro-imaging data.

A NEW METHOD FOR THE DETECTION OF GALAXY CLUSTERS IN X-RAY SURVEYS

J.M. Piacentine, P.J. Marshall, J.R. Peterson, and K.E. Andersson

ABSTRACT

For many years the power of counting clusters of galaxies as a function of their mass has been recognized as a powerful cosmological probe; however, we are only now beginning to acquire data from dedicated surveys with sufcient sky coverage and sensitivity to measure the cluster population out to distances where the dark energy came to dominate the Universe's evolution. One such survey uses the XMM X-ray telescope to scan a large area of sky, detecting the X-ray photons from the hot plasma that lies in the deep potential wells of massive clusters of galaxies. These clusters appear as extended (not point-like) objects, each providing just a few hundred photons in a typical observation. The detection of extended sources in such a low signal-to-noise situation is an important problem in astrophysics: we attempt to solve it by using as much prior information as possible, translating our experience with wellmeasured clusters to define a "template" cluster that can be varied and matched to the features seen in the XMM images. In this work we adapt an existing Monte Carlo analysis code for this problem. Two detection templates were dened and their suitability explored using simulated data; the method was then applied to a publically avalable XMM observation of a "blank" field. Presented are the encouraging results of this series of experiments, suggesting that this approach continue to be developed for future cluster-identication endeavours.

INTRODUCTION

Galaxy clusters are of interest because they are the largest gravitationally bound structures in the Universe; their composition is believed to be dominated by dark matter. The number density of clusters as a function of mass and redshift is a sensitive probe of cosmology: the detection of clusters is important to the study of the evolution of large scale structure.

To date, X-ray observations have provided the highest precision measurements of galaxy clusters. The heated gas trapped within the gravitational potential well of the dark matter emits strongly in the X-ray band: much has been learned about the physics of galaxy clusters through pointed X-ray observations. However, to discover new clusters via their X-ray emission a "blank sky" survey is needed. For surveys where the detection of X-ray photons occurs less than 1 per second, it would be useful to be able to construct an approach to distinguish galaxy clusters from other astronomical objects using as much information as possible. The angular position on the sky and the energy spectrum of the X-ray emission are observable quantities: we will use both.

Objects may be identified based on their angular size; it is important to understand the parameters that may affect this observed spatial extent. The smearing of a point source to appear wider is due to the telescope's point spread function (PSF): a point source is understood to have the spatial extent of the PSF. Any source observed to be larger is by definition extended (Figure 1). The physical sizes of the two most abundant X-ray sources, active galactic nuclei (AGN) and galaxy clusters, are such that, when placed at cosmological distances (100's of Mpc) the former appear point-like, and the latter extended. Typical cluster sizes are ~100 Kpc at the cluster, and appear with arcminute angular diameters. However, some clusters are smaller than this and the X-ray emission is more concentrated than the physical extent of the cluster gas so that there is some possibility of confusion, especially if the PSF is large.

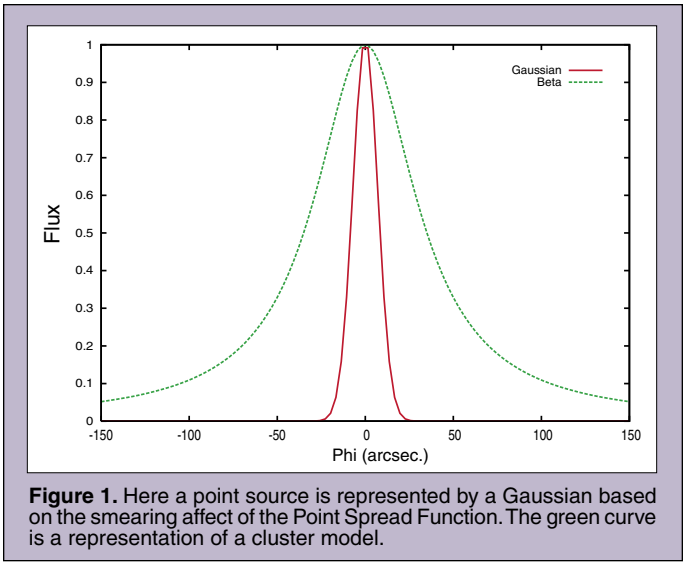

Figure 1. Here a point source is represented by a Gaussian based on the smearing affect of the Point Spread Function. The green curve is a representation of a cluster model.

The AGN emission comes from a comparatively very small region around the black hole at the center of the galaxy and always appear point-like. It is important to keep in mind that it is possible for two AGN to be aligned and may appear as a single extended source.

The classification of an object can also be determined by spectral analysis. AGN are known to have power law spectra, whereas cluster spectra go as $\propto \exp(-h\nu/kT)$ [1], with distinct line features produced by ions in the intracluster gas. To some extent it is possible to mimic a cluster spectrum with a AGN spectrum, and vice versa: one of the aims of this work is to investigate this degeneracy.

We use data from the XMM Newton observatory; this telescope currently provides the largest collection area and field of view, making it the present instrument of choice for blank sky surveys. Specifically we focus on spectroimages taken with XMM-Newton's European Photon Imaging Camera (EPIC). EPIC has three cameras, two Metal Oxide Semiconductor (MOS) charge-coupled devices (CCDs) and one PN CCD. As photons strike the CCDs, the location of the event and the photon energy is recorded. The analyzed observation data is simply a list of photon CCD coordinates and energies.

We are interested in detecting and classifying objects in previously unobserved areas of sky: the XMM-Newton Large Scale Structure survey is a large program designed with this end in mind [2]. The observations cover an 8x8 deg^2 section of sky centered at 2h 18m 00s, -7° 00' 00" , and are comprised of 24 x 24 obervations each with a field of view of some 30 arcminutes in diameter.

The PSF of all X-ray telescopes is energy dependent: the chromatic aberrations in the optics are significant. To properly correct for this the data must be simulated using Monte Carlo methods; the analysis code for doing this is called X-ray Monte Carlo (XMC) [3]. While used extensively in the analysis of single cluster observations, here we apply this "forward folding" approach to the analysis of survey data containing both AGN and clusters, and use both the spectral and spatial information to distinguish the objects. We do this by defining a flexible template for each object (or part-object), and then fitting the parameters of these templates to the data.

In the first section we discuss the experiments performed on simulated data to test our ability to locate and identify objects based on their spatial and spectral properties; in the second section we present results from an analysis of an XMM blank field. We discuss our findings and draw brief conclusions in the final section.

METHODS

In this section we give a brief description of the analysis code used, and then outline a number of experiments designed to explore the suitability of the template-fitting approach.

A. Overview of XMC

Given noise-free data, the inferences drawn using XMC would extend to all parameters of all the astronomical objects in the eld. How well this task can be accomplished depends on the quality of the dataset; X-ray data are typically quite noisy, with very few photons per bin in energy and detector position.

It is first necessary to provide XMC with input, either a previously created simulation or data from an observation. A model is then created to test against the data. The model's parameters will be set with initial trial values and a range that the parameters are allowed to explore. XMC generates a set of simulated data, which are compared with the real data using a two-sample multinomial likelihood function, adjusts the simulated data and repeats. For each iteration it returns a set of all parameters being explored and the mist statistic(χ^2). The lower the value of this mist, the closer the mock data is to that of the real data. The parameter space is explored using the Markov Chain Monte Carlo technique [4], the end result of which is an ensemble of parameter sets that define models which all fit the real data acceptably well. For each model a mock sky can be plotted; by averaging these maps, deconvolved images can be reconstructed. This can be done separately for the AGN and cluster parts of the model: combining these images in different color channels (for instance, plotting AGN photons in green and cluster photons in red) allows the user, at a glance, to get an idea of the nature of objects analyzed.

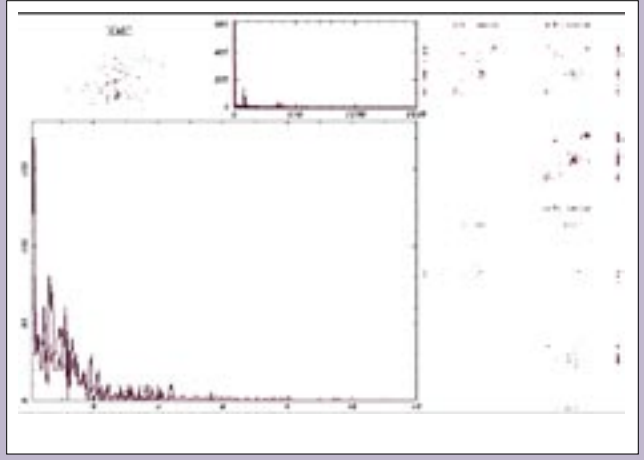

Figure 2. The main window of the XMC windows shows the spectra (real and fitted). The upper right shows the locations of the input data and the bottom right XMC's inferred location.

Figure 2 shows the real time progress window displayed by XMC, allowing a user to monitor the evolution of the spatial and spectral analysis. This display shows the user the location of the input data as well as the location of the tted objects. The main windows displays the spectral analysis.

B. Experiment 1

First a simulation was dened using a pointlike AGNspectrum object centered in the field of view. This simulated data was then analysed assuming the same model. All parameters of the model were xed at the true values save one, so that the object and the model were identical except for the object location. This one coordinate was then allowed to vary, to see if XMC could locate an object in a single spatial direction.

The simulation object's location parameter, Phi, was set at a value of -60. The variation in mist as the model Phi value was changed is shown in Figure 3. As shown in the plot the lowest misfitt value corresponds with the correct parameter value.When applying this same technique to the Psi coordinate, an error was discovered and corrected in the XMC code when the program failed to locate the object. It is interesting to note the slope of the curve in this plot. As XMC's estimation of the parameter value approaches that of the simulation, the slope sharpens drastically, while the slope levels out the further away one gets from the true value. From this we can better understand the burn in time required for XMC. For the first few hundred iterations the program simply "guesses" parameter values in the specied allowed range and records a misfit value until it has enough data points to recognize sloping curves, making it possible to hone in on the parameter value. However, to some extent XMC is always guessing; the χ^2 simply helps XMC judge when and how much to adjust a parameter.

C. Experiment 2

The purpose of this second experiment is to determine whether or not XMC can recover the nature of a simulated object at fixed position. The spectral signatures of AGN and clusters are fairly well-defined; we use the community standard analysis code XSPEC to calculate predicted spectra given model parameters. We also need to define a spatial model for the emission from each object, and here we investigate two approaches. Both entail the superposition of two spatial models, one for each spectral type. The centroids of the two components are tied together, and the fraction of flux associated with each is allowed to vary: this parameter encodes the "clusterness" of the object. We refer to these two-component models as "blobs".

In the first approach we envisage fitting one two-component model to each object in the field: in this case it makes sense to make the spatial emission distribution as close to what is already known about clusters and AGN as seen in X-ray observations. Hence, we assign the AGN flux component a delta-function spatial distribution (making it a true point source). For the cluster we use the surface brightness given by Sarazin [1],

$$I(r) \propto \frac{1}{\left[1+\left(\frac{r}{r_c}\right)^2\right]^{3\beta-\frac{1}{2}}} \qquad (1)$$

where r is the radius of the cluster and r_c is the core radius and is the parameter allowed to vary within XMC. The parameter β has been calculated for many pointed observations and for the simplicity of the simulation was set to 0.66 in accordance to these findings.

In the second approach we admit the possibility of more complex objects, and attempt to build these from superpositions of many models, each of some general spatial model. We choose a Gaussian distribution,

$$I(r) \propto \exp\left(\frac{-r^2}{2\sigma^2}\right) \qquad (2)$$

where σ is the parameter allowed to vary within XMC. For a truly point-like source the delta function model is at first sight the better option, but allowing in the spatial Gaussian model to become arbitrarily small achieves the same aim. When combining many blobs to represent an object, do the Gaussians do a better job of dealing with the smearing effect of the PSF on a point source?

Experiment 2 consisted of two tests. In the first, a point-like AGN object was simulated, centered on the exposure map. This simulated data was then fitted using XMC with a blob of the first kind (beta model plus delta function). The core radius of the beta model component and the spectral parameters of both were allowed to vary. In the second test, the simulation was identical, but the blob used in the analysis was of the second kind (two Gaussian spatial models, with equal extent, one with AGN spectrum and and one with cluster spectrum). In both tests, XMC was used to explore the spatial and spectral parameters of the model and return the fraction of photons that were "AGN-like" and "cluster-like".

As can be seen in Figure 4, XMC has correctly inferred that the simulation object was an AGN. The spatial extent corresponds

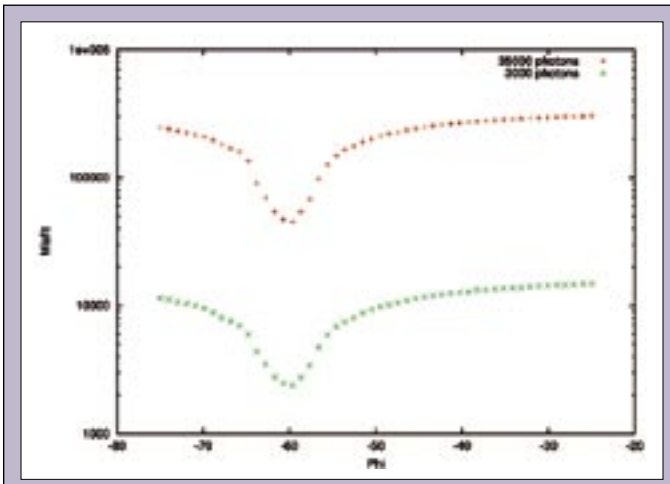

Figure 3. Plot of misfit χ^2 vs. Phi. XMC located the AGN simulation with 3000 photons and again with 35000 photons.

closely to the XMM PSF, with a diameter of about 6 arcseconds. The blobs of this run consisted of a point source model and Beta model. The number of photons XMC determined to be of AGN origin is approximately 1000 times greater than cluster photons. This information can be used to give a rough estimate on the probability of being an AGN. The results of the blobs with two spatial Gaussians tied together were similar. However, the ratio of AGN to cluster flux was not as extreme. This can be understood in terms of the different spatial models used: if the data are well represented by a beta model plus delta function, then we would expect this model to perform better in the classification of the object from noisy simulated data. We shall see in section III that real astronomical objects are less clear-cut cases.

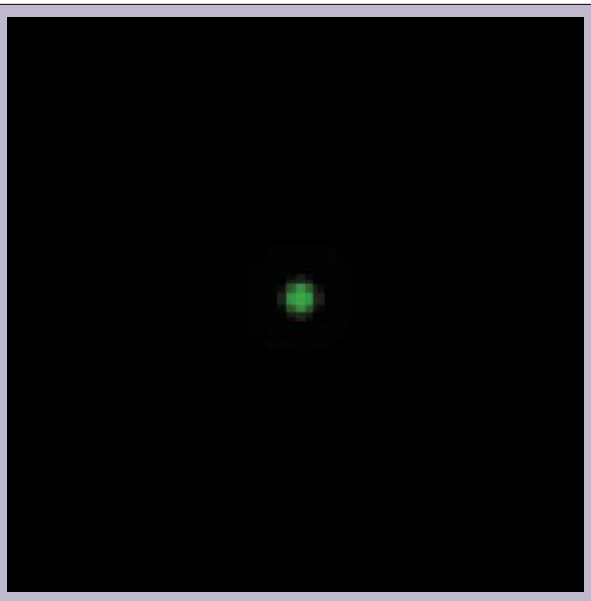

Figure 4. The simulation was set as an AGN located at the center of the field. The model was allowed to vary between AGN and cluster parameters. XMC determined that the simulation photons consisted that of only AGN to one part in 1000.

D. Experiment 3

A final simulation was created using a mixture of AGN and cluster objects (three of each) placed in various locations on the sky. Again two tests were run. The analysis blobs were the same as the ones used in experiment 2. Six blobs were used to match the six simulation objects. The objective of this experiment was to ascertain the ability of XMC to track the location, spatial and spectral parameters of each object simultaneously. Knowing that XMC can locate and distinguish simulation objects, it is necessary to evaluate how well XMC can distinguish objects in close proximity of one another. In the upper right corner of this simulation two AGN were placed near one another, and both were located within the radius of a cluster object, so that XMC must distinguish between three tightly packed objects. In the center of the field was placed a cluster. A cluster was placed in the bottom left and an AGN in the top left.

The results (Figure 5) show our ability to identify the two AGN and cluster in the upper right corner. However, there seems to be

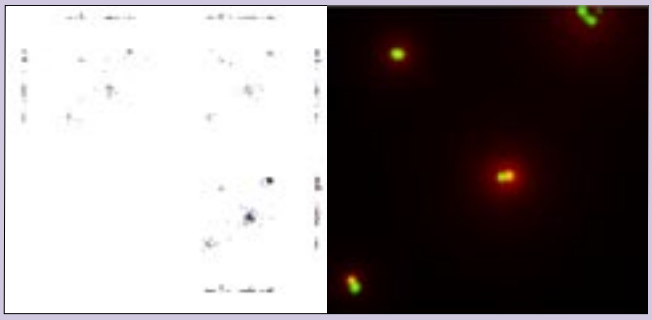

Figure 5. The left picture shows the locations and sizes of the objects before analysis. The more spread out objects are the clusters. In the right picture the 3 cluster & 3 AGN objects can be seen. The central and bottom left corner objects are clusters, in the top right two AGN and top left another AGN. XMC located every object but the two clusters still have significant AGN flux.

some ambiguity with XMC's classification of the other objects. The two remaining cluster objects appear to have just as much AGN flux as cluster flux.

The upper right conglomeration of three objects has some cluster (red) but is dominated by AGN (green). This indicates that XMC can distinguish objects near or overlapping one another. However, the purely cluster object has a significant amount of AGN flux to it. This effect is analyzed in the final section.

XMM DATA

We now go on to apply the methodology introduced and demonstrated in the previous section to real XMM data. A dataset from the Large Scale Structure survey was retrieved from the on-line archive, and reduced to a list of photon locations and energies readable by XMC. It is interesting to note some features of the EPIC cameras. We can see (Figure 6) the PSF affects the objects in the exposure map. The central objects all have a width of about 6 arcseconds and up to 20 arcseconds at the edges of the field of view. In the upper left corner there is an object which is much larger than any other object. This could be the PSF greatly smearing the object near the edge of the camera or this could be an extended source. The gaps between the CCDs is also noticeable in the exposure map (only the data from the PN CCDs were used).

For simulations, the number of objects is known and creating the number of model objects to be the same is simple enough. However when analyzing real data, the number of objects will not be known, so it is necessary to take a somewhat different approach. Experience with pointed observations suggests that a combination of Gaussian components (typically ~100) is a good way of characterising a complex image. For this analysis we used blobs of the second kind; to recap, these are two concentric Gaussians of equal size, that differ in their spectral properties. The fraction of the flux assigned to each spectral component is an indication of the "clusterness" of that blob. Bright objects in the data can then be built up from collections of blobs, whose clusterness fraction should change during the XMC analysis to reflect the spectral type of the object.

When dealing with real data, the background emission has to be taken into account. This was modelled as a spatially uniform distribution, composed of a variable fraction of high energy particles

and the remainder detector noise. The flux from this background can be predicted and fitted to the data at the same time as the blobs. Note that the astronomical X-ray background was not included in this way: the X-ray background is made up of clusters and AGN and is in fact what we are trying to detect.

It can be seen that in areas between the CCDs, XMC assumes there is flux, although no data is actually recorded by the telescope in these regions. When the data from the MOS detector along with the PN is used the gaps between the CCDs will be resolved.

Comparing the raw photon counts map (Figure 6) with the output from XMC, we see that the majority of features in the former are detected in the latter. In a survey such as this, it was expected that 70% of the objects would be AGN. As is shown in Figure 7, all the objects have some AGN component associated with them. However, a significant number of objects seem to have a considerable amount of cluster flux as well.

Focusing on the object in the top left hand corner of the field, which seems to be smeared, we see both AGN and cluster flux. Also, the elongation of the reconstructed object is much greater than that of the object in the data. This could possibly be the greater PSF of the telescope at the edges, smearing the object. However, one would expect the reconstructed image to be similar to that of the exposure map. Considering that this object is the brightest in the field of view, it is possible that XMC positioned a majority of the 100 blobs on this one object; with the averaging process comes some variance, which is seen as a smearing in the recontructed image. Further work should include study of this dilution effect, especially in regard to the classification of objects.

DISCUSSION AND CONCLUSIONS

The test runs provided some level of confidence in our ability to locate and distinguish astronomical objects using template-fitting with XMC. The accuracy achieved for simple simulation was encouraging. When complicating a test run by introducing a larger number of blobs, the program became slower, because of the greater number of parameters to be varied.

In experiment 2, there seemed to be some ambiguity of the nature of the object when using the spatial Gaussian blobs. However, some ambiguity is to be expected when allowing the spatial parameters of the Gaussians to vary. By tying the parameter of each Gaussian, the model with AGN spectrum would be allowed to grow to the size of a cluster, whereas the model with cluster spectrum would also be allowed to shrink to the size of the telescope's PSF. The extra degree of freedom allowed provides us with useful information on the ability of the program to determine astronomical objects.

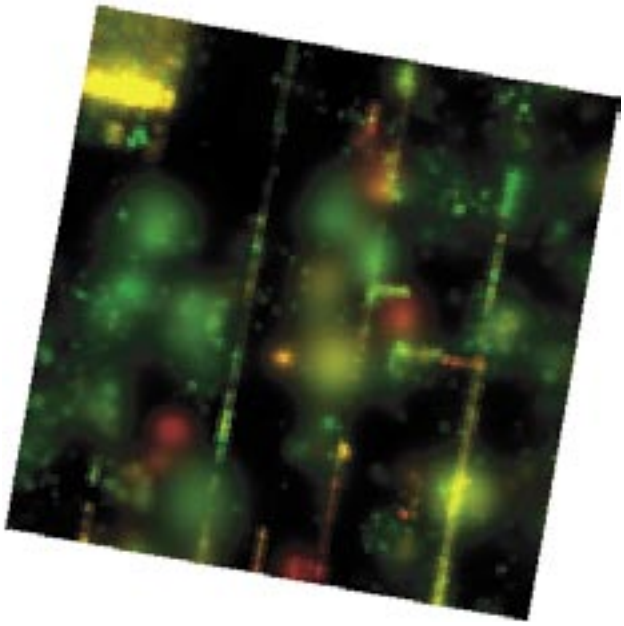

Figure 7. The results of the LSS-20 analysis with green representing AGN flux and red representing that from clusters.

The six blobs in experiment 3 successfully matched the location of the simulated objects. However, the nature of the objects was not determined accurately. Whilst this may be due to the relatively low number of iterations XMC was allowed to make, there is also some considerable degeneracy between the different spectral and spatial signatures of the two classes of objects considered. The spectra of the mock data are suffieciently similar to allow for the fitting of cluster spectra to a power-law within the uncertanties of the data, so that a power-law spectrum seems to be an acceptable fit to a cluster image, leading to the question of spatial extent. Given the wide range in the variable assigned to the size of a cluster, may indicate that extended source distributions fit poorly to AGN images, leaving the purely green AGN spots in Figure 5. However when analyzing a cluster image, the PSF of a point source may be large enough to fit both photon ditributions to the image.

This was also seen in the analysis of the XMM data: Figure 7 shows many objects with ambiguous classifications for reasons seen in experiment 3. However, this image clearly contains an incredible amount of information. A large number of objects are detected, and the range of colors is impressive: the objects are not all yellow!

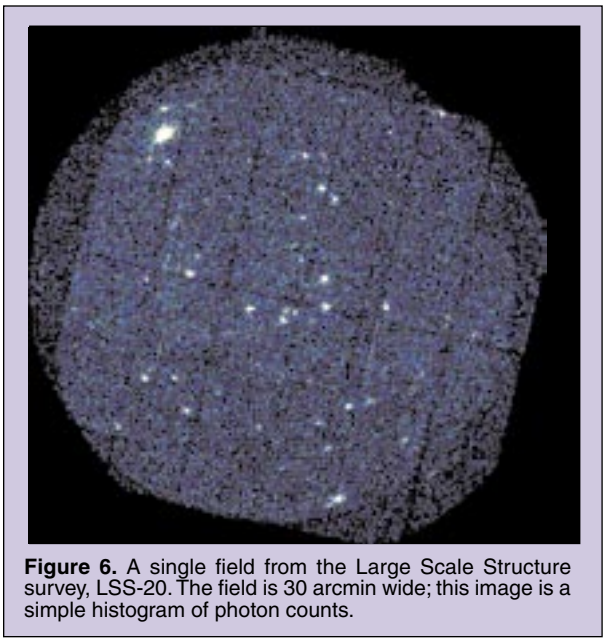

Figure 6. A single field from the Large Scale Structure survey, LSS-20. The field is 30 arcmin wide; this image is a simple histogram of photon counts.

The results of the LSS data suggest that using multiple blobs to reconstruct a single object is a viable procedure for distinguishing objects. Averaging the ensemble of models generated by XMC to make a deconvolved map demonstrates the effectiveness of the program's ability to locate objects. The color representation of AGN and cluster flux has allowed rough estimates of the object classification probability to be made.

The methodology demonstrated in this work, whilst not without teething problems, seems to us to deserve further effort in its development. The major benefits it brings include full propagation of errors from noisy data to reconstructed image, correct treatment of a PSF that varies both spectrally and across the image, the input of sensible prior knowledge (via the choice of blob surface brightness distributions and spectra), and the separation and visualisation of the two putative image components without the need for isolation of individual objects (which often lie in close proximity to each other).

ACKNOWLEDGMENTS

We thank Alex Refregier for bringing our attention to the problem of LSS analysis. J. Piacentine was supported by funding from the Department of Energy, Office of Science during a summer internship at Stanford Linear Accelerator Center. This work was supported in part by the U.S. Department of Energy under contract number DE-AC02-76SF00515.

REFERENCES

[1] C. L. Sarazin, X-ray emission from Clusters of Galaxies, Cambridge University Press (2004)

[2] M. Pierre et al, J. Cosmol. Astropart. Phys., 9, 11 (2004)

[3] J.R. Peterson, J.G. Jernigan, and S.M. Kahn, Multivariate Monte Carlo Methods for the Reflection Grating Spectrometers on XMM-Newton, ApJ, 615, 545 (2004)

[4] W.R. Gilks, S. Richardson, and D.J. Spiegelhalter, Markov-Chain Monte-Carlo In Practice, Cambridge: Chapman and Hall (1996)