Deconvolution in High-Energy Astrophysics: Science, Instrumentation, and Methods

David A. van Dyk*, Alanna Connors†, David N. Esch‡, Peter Freeman§,
Hosung Kang¶, Margarita Karovska‖, Vinay Kashyap**,
Aneta Siemiginowska††, Andreas Zezas‡‡,

Abstract.   In recent years, there has been an avalanche of new data in observational high-energy astrophysics. Recently launched or soon-to-be launched space-based telescopes that are designed to detect and map ultra-violet, X-ray, and γ-ray electromagnetic emission are opening a whole new window to study the cosmos. Because the production of high-energy electromagnetic emission requires temperatures of millions of degrees and is an indication of the release of vast quantities of stored energy, these instruments give a completely new perspective on the hot and turbulent regions of the universe. The new instrumentation allows for very high resolution imaging, spectral analysis, and time series analysis; the *Chandra X-ray Observatory*, for example, produces images at least thirty times sharper than any previous X-ray telescope. The complexity of the instruments, of the astronomical sources, and of the scientific questions leads to a subtle inference problem that requires sophisticated statistical tools. For example, data are subject to non-uniform stochastic censoring, heteroscedastic errors in measurement, and background contamination. Astronomical sources exhibit complex and irregular spatial structure. Scientists wish to draw conclusions as to the physical environment and structure of the source, the processes and laws which govern the birth and death of planets, stars, and galaxies, and ultimately the structure and evolution of the universe.

The California-Harvard Astrostatistics Collaboration is a group of astrophysicists and statisticians working together to develop statistical methods, computational techniques, and freely available software to address outstanding inferential problems in high-energy astrophysics. We emphasize fully model-based statistical inference; we explicitly model the complexities of both astronomical sources and the data generation mechanisms inherent in new high-tech instruments, and fully utilize the resulting highly structured models in learning about the underlying astronomical and physical processes. Using these models requires sophisticated scientific computation, advanced methods for statistical inference, and careful model checking procedures.

*Department of Statistics, University of California, Irvine, CA, http://www.ics.uci.edu/~dvd/
†Eureka Scientific, Oakland, CA, http://wwwgro.unh.edu/users/aconnors/
‡Department of Statistics, Harvard University, Boston, MA,
§Harvard-Smithsonian Center for Astrophysics, Boston, MA, pfreeman@cmu.edu
¶Department of Statistics, Harvard University, Boston, MA, http://www.people.fas.harvard.edu/~hkang/
‖Harvard-Smithsonian Center for Astrophysics, Boston, MA, mkarovska@cfa.harvard.edu
**Harvard-Smithsonian Center for Astrophysics, Boston, MA, http://hea-www.harvard.edu/~kashyap/
††Harvard-Smithsonian Center for Astrophysics, Boston, MA, http://hea-www.harvard.edu/QEDT/aneta/HomePage.html
‡‡Harvard-Smithsonian Center for Astrophysics, Boston, MA, azezas@cfa.harvard.edu

Here we discuss the broad scientific goals of observation high-energy astrophysics, the specifics of the data collection mechanism involved with the *Chandra X-ray Observatory*, current statistical methods, and the Bayesian models and methods that we propose. We illustrate our statistical strategy in the context of several applied examples, including the estimation of hardness ratios, spectral analysis, multiscale image analysis, and reconstruction of the distribution of the temperature of hot plasma in a stellar corona. This paper was presented at the *Case Studies in Bayesian Statistics Workshop 7* held at Carnegie Mellon University in September 2003.

Keywords: Background Contamination, Censoring, Chandra X-ray Observatory, Chi Square Fitting, Count Data, Contingency Tables, Deconvolution, Differential Emission Measure, EM-type Algorithms, Frequency Evaluations, Richardson-Lucy, Hardness Ratios, Hubble Space Telescope, Image Analysis, Log-Linear Models, Markov chain Monte Carlo, Measurement Errors, Multiscale Methods, Sampling Distributions, Smoothing, Prior Distribution, Point Spread Function, Posterior Predictive Checks, Power Law, Poisson Models, Spectral Analysis, Timing Analysis

1   Astrostatistics

The disciplines of statistics and astronomy have long and mingled histories. Indeed, it was Babylonian astronomers who appear to have been the first to tackle the fundamental statistical problem of estimating parameters from observational data. Although nothing has survived to indicate what methods these ancient astronomers used, it is clear that they incorporated a compromise between their observations and their need for computation (Neugebauer 1951; Plackett 1958). Thousands of years later, toward the end of the sixteenth century, the Danish astronomer Tycho Brahe used the arithmetic mean to eliminate measurement error in his measurements of the locations of stars and planets (Plackett 1958). The development of least squares regression in the second half of the eighteenth century is also owed largely to the ingenuity of astronomers and the statistical challenge of the astronomical problems of the day (Stigler 1986). When looking over the expanse of history the anomaly appears to be the relatively recent divergence of the two disciplines; the current flurry of collaboration among statisticians and astronomers appears as an overdue renaissance.

As noted by Connors (2003), Bayesian (or more generally, model-based) ideas have helped to fuel this renaissance. In the 1970's, new electronic recording devices triggered new questions on how to handle the 'inverse' problem: the characteristics of the new detectors were known but the sky was not. In statistical terms, understanding the detectors means that given the characteristics of astronomical source, astrophysicists knew what to expect in the data that the detector would produce; ideally this could be formalized into a likelihood function. The 'inverse' problem was to reconstruct the model parameters (i.e., the sky) from the observed data. Albert Bijaoui (1971a; 1971b), used a new electronic camera (instead of the traditional photographic plates) to try to measure the star counts in the dense stellar cluster M13. His analysis was inspired by the physicist Ed Jaynes's insistence that the treatment of probability in

measurements and inference be as rigorous as that in statistical mechanics. His work also influenced radio astronomers using the new 'interferometer' telescopes. From a more traditional statistical point of view, Richardson (1972) and Lucy (1974) were also very influential. They proposed an EM algorithm for computing the maximum likelihood (ML) estimate for Poisson image reconstruction; Lucy went further to suggest that maximum a posteriori (MAP) estimates might be better behaved. (Like many others before them, Richardson and Lucy's EM algorithm predate the general formulation of EM by Dempster, Laird, and Rubin, 1977; see Meng and van Dyk, 1997.) The utility of such model-based methods was demonstrated to astronomers when Richardson and Lucy's algorithm was used to account for the out-of-focus mirror on *The Hubble Space Telescope*. This was a dramatic illustration of the power of careful data analysis to improve scientific inference, even when the usual solution of building more powerful telescopes to gather better data is unavailable.

Indeed, as the *Hubble* incident illustrates, the complexity of the newest generation of telescopes has led to more intricate data collection mechanisms that require sophisticated data analysis technique; these instruments have thus fueled the renaissance of astrostatistics. Take for example, the Space Interferometry Mission[1] (SIM). This instrument is scheduled to be launched in 2009 and is designed to measure the direction to an astronomical source with much higher accuracy than is now available. Among the statistical challenges posed by SIM is the allocation of the observation protocol to optimize information. SIM measurements are expected to be precise enough to detect the stellar wobble caused by an orbiting Earth-like planet. Because these measurements are very time consuming, they must be carefully allocated and precisely timed; Bayesian methods to dynamically update observing protocols to optimize the expected information are being developed by Loredo and Chernoff (2003). A rather different challenge is posed by the Sloan Digital Sky Survey[2] (SDSS), an on going, Earth-based, multi-wavelength survey of 10,000 square degrees of the sky. Upon completion of one fifth of the survey, SDSS data already included 5×10^7 detectable objects. This rich data set allows for careful investigation of the large scale structure in the distribution of galaxies, the structure of our Milky Way Galaxy, and the detection of new types of objects. The data mining, data reduction, classification, and computational challenges in this project are evident and an area of active interdisciplinary work among astronomers, statisticians, and computer scientists (Strauss 2003; Nichol et al. 2003). These are but two examples among the many large-scale missions in astronomy and astrophysics that offer fertile ground for statisticians interested in methodological development; we discuss a third example, *The Chandra X-ray Observatory*[3] (*Chandra*), at length in this article. Each of the many missions pose unique statistical challenges that span the breath of statistical science and offer scientific insight across the breath of astrophysics. Readers interested in learning more about various areas of current work in astrostatistics should consult the three volumes edited by Feigelson and Babu (1992; 1997; 2003) which chronicle the Statistical Challenges in Modern Astronomy Conferences[4]. Another current resource

---

[1]URL: http://sim.jpl.nasa.gov

[2]URL: http://www.sdss.org

[3]URL: http://chandra.harvard.edu

[4]URL: http://www.astro.psu.edu/SCMA

is the upcoming special issue of *Statistical Science* (2004, Number 2) edited by Chris Genovese and Larry Wasserman and devoted to topics in astrostatistics. Just this week SLAC hosted a conference on statistical problems in Particle Physics, Astrophysics, and Cosmology[5] (Lyons et al. 2004).

This article describes the work of the California-Harvard Astrostatistics Collaboration in developing methodology, algorithms, and software for the analysis of high-resolution X-ray data. Current effort is focused on the data obtained with the state-of-the-art X-ray telescope known as *The Chandra X-ray Observatory*, but many of the new techniques apply directly to other current and upcoming high-energy missions in astrophysics such as *X-ray Multi-Mirror-Newton (XMM-Newton)*[6], *Constellation-X*[7], *Micor-Arcsecond X-ray Imaging Mission (MAXIM)*[8], *Generation-X*[9], and *The Gamma Ray Large Area Space Telescope (GLAST)*[10]. In Section 2 we describe the mission and scientific objectives of *Chandra*, the instrument and data collection mechanism that were designed to meet these objectives, and typical data analytic goals and methods. In Section 3, we focus on a large class of such analyses, which involve deconvolution; in particular, we describe the model-based and Bayesian techniques that we have developed to accomplish more reliable results and illustrate these results with several examples of applications to *Chandra* data. Concluding remarks appear in Section 4.

2   The Chandra X-ray Observatory

2.1   Mission and Science

Early X-ray Astronomy. William Herschel first discovered electromagnetic waves outside the visible spectrum in 1799. He and his sister Caroline were building the best telescopes in the world and using them to study the Sun. Herschel made an astonishing discovery: He found that a significant proportion of the Sun's heat is emitted beyond its spectrum's red end, a region that is not visible to the naked eye. Herschel called this infra-red light, 'invisible light'. In contrast to the infra-red, where the wave-nature was clear, the discovery of X-rays and γ-rays came out of the study of ionizing particles[11]. In 1895 when Wilhelm Röntgen first observed a highly penetrating form of radiation that he called X-rays, it was not clear whether they were 'photons' (electromagnetic radiation with no intrinsic mass) or very speedy particles of matter (usually nuclear particles; electrons, positrons; and more exotic elementary particles such as muons and neutrinos). It was not until 1912 that X-ray were conclusively shown to be very short

---

wavelength electromagnetic waves, and thus a higher energy form of light. Indeed, the first astrophysical X-ray and γ-ray detectors were analogous to flying Geiger counters, with metal tubes on the front, and special layers around the sides to warn when charged particles impinged on them.

Since 'ionizing radiation' is absorbed by the Earth's thick atmosphere, 'black light' (ultraviolet radiation) and higher energy radiation such as X-rays and γ-rays from space cannot be viewed from the Earth's surface. Although this is a distinct advantage for life on Earth, it poses a difficulty for X-ray astronomers. X-ray detectors must be placed above the Earth's atmosphere. Initial efforts were hoisted up with balloons and rockets; Friedman (1960) gives a nice review of the first 15 years of these flights. The early instruments mainly detected the Sun and the Earth's own upper atmosphere, which glows when hit by streams of energetic charged particles. The first detection of X-rays outside the solar system came in 1962, when a small rocket carried an X-ray detector into space; it operated for only a few minutes, but was able to detect the first known X-ray emission outside our solar system: both an apparent diffuse glow and an enhanced brightness toward the center of the Galaxy (Giacconi et al. 1962).

In 1970, the first satellite devoted to imaging the sky in X-rays, *Uhuru*, was launched into a low Earth orbit (at about 500 km). ('Uhuru' is the Swahili word for 'freedom'; *Uhuru* was launched from Kenya.) *Uhuru* surveyed the entire sky and discovered many new X-ray objects, such as X-ray binaries, supernova remnants, galaxies, and diffuse emission from clusters of galaxies. Still, *Uhuru* and her later sisters (*Ariel V*, *The Orbiting Solar Observatories (OSO)*, *The High-Energy Astronomy Observatories-1 (HEAO-1)*, etc.) were proportional counters—more sensitive than the earliest flying 'Geiger counters', but much the same idea. This is a reminder of the fundamentally Poisson nature of the signal of interest. High-energy astrophysics "comes in lumps," as Feynman says of photons in his famous introductory physics lecture; Feynman meant that photons are discrete packages of energy, are countable, and are therefore intrinsically Poissonian.

Such detectors register X-rays but provide the location of the photon source only with very large error. The spatial resolution of these early detectors was roughly equivalent to looking through an array of paper-towel tubes (with no imaging optics). Imaging detectors were not available until much later. Even for the Sun, a *very* bright X-ray source, there were no telescopes[12] until Skylab was launched in 1973, and then astronauts were required to operate the telescope; see Noyes (1982). The first fully imaging X-ray telescope designed for the *much fainter* extra-solar X-ray emission was the *Einstein Observatory* (*HEAO-2*). Launched in 1978, it had an angular resolution of a few arc-seconds, a field of view of tens of arcminutes, and a sensitivity that was several hundred times greater than was available with previous missions. *Einstein* provided, for the first time, the ability to distinguish point sources, extended objects, and diffuse emission.

---

[12]Telescopes are more than simple detectors. They include imaging optics that refract or reflect photons so that the objects being viewed are magnified and focused onto the detector. With a conventional Earth-based telescope, the detector is generally either a camera or a human eye.

Many of the discoveries made by the early X-ray missions[13] motivated the development of new technologies for obtaining higher quality data. The concept of *The Chandra X-ray Observatory*, for example, was conceived of at the time of the *Einstein* mission. It was difficult at that time, however, to imagine the scientific leap forward that such a high quality X-ray telescope would provide. (See Tucker and Tucker (2001) for the history of X-ray astronomy leading up to the design, construction, and launch of *Chandra*.)

Current Scientific Objectives. The sky in X-rays looks very different from that in optical. X-rays are the signature of accelerating, energetic charged particles, such as those accelerated in very strong magnetic fields, extreme gravity, explosive nuclear forces, or strong shocks. Thus, X-ray telescopes can be used to study nearby stars (like our Sun) with active coronae, the remnants of exploding stars, areas of star formation, regions near the event horizon of a black hole, very distant but very turbulent galaxies, or even the glowing gas embedding a cosmic cluster of galaxies. X-ray emission from these diverse objects is also diverse. The spectra (i.e., the distribution of the energies of the photons that a source radiates) can vary dramatically; for instance, a typical star like our Sun radiates about a million times more energy in visible light than in X-rays, whereas strong X-ray sources like cataclysmic variables can produce thousands of times more energy in X-rays than in visible light. This is a striking example of how a spectrum, i.e., the distribution of the energy of the photons that the source radiates, can vary dramatically between objects. These spectra give insight into many aspects of cosmic X-ray emitters: their composition, their density, and the temperature/energy distribution of the emitting material; any chaotic or turbulent flows; and the strengths of their magnetic, electrical, or gravitational fields. The spatial distribution of the emission is also key; it reflects physical structures in an extended source, for example, the distribution of point sources embedded in a diffuse galactic emission, jet emission or indication of an outflow of hot matter, the shape of the remnant of a supernova explosion, and structures created by gravitational lenses. Some sources exhibit temporal variability or periodicity that might result from rotation, eclipses, magnetic activity cycles, or turbulent flow of matter into a deep gravity well. Thus, instrumentation that can precisely measure the energy, location, and arrival time of X-ray photons enables astrophysicists to extract clues as to the underlying physics of X-ray sources.

Chandra observations, for example, have helped us to understand black holes in stellar binary systems with matter flowing toward their gravitational potential (known as *accreting black hole X-ray binaries*). Such black holes are visible to *Chandra* because of material nearby that flows into the black hole potential and releases its gravitational energy in the form of X-ray radiation. If there is no matter close to a black hole we would not see any emission; such black holes are truly black.

Chandra also helps us understand the nuclei of *active galaxies*—galaxies in which the central regions dominate the luminosity of the entire galaxy. Quasars, improbably distant and highly luminous X-ray emitters, are located at galactic nuclei and are seen

---

[13]URL: http://heasarc.gsfc.nasa.gov/docs/heasarc/missions/alphabet.html

as point sources in X-rays. Most of their energy is released within a region which is only about 0.005% of the size of the entire Milky Way; this region is too small to be resolved even by optical telescopes. Quasars are thought to be powered by the gravitational energy released by matter accreting onto a supermassive black hole. How the accretion flow proceeds is still an unanswered question. X-ray emission can, however, provide clues as to the accretion process, giving limits to the available fuel, to the temperature and density within the accreting matter, and to the ionization state[14] of the nearby gas. Mapping the distribution of the X-ray emission in the vicinity of a quasar provides information about the environment in which quasars reside. How this environment reacts to the luminous quasar can be studied in X-rays and this study provides insights into the power associated with jets and hot matter flowing out from the quasar (see Figure 1). X-ray jets emanating from many quasars that were discovered by *Chandra* indicate velocities within a few percent of the speed of light and high-energy particles existing at enormous distances from the quasar (hundreds of kiloparsec or millions of light years)[15]. How the jet is created and collimated (i.e., kept very narrow) over such large distances and where the acceleration of the jet particles occurs are questions currently being studied with *Chandra*.

Chandra observations of clusters of galaxies provided the first good quality temperature maps of the emitting hot plasma on very fine (arcsecond) scales. X-ray emission that was previously thought to be smooth can now be seen to consist of structures and different shapes. The shapes and sizes of these structures as well as the borders between them give clues as to the underlying physical processes, such as the heating and cooling of the gas, and in turn the evolution of the entire cluster. The shape of hot structures (*buoyant bubbles*) and their location within the cluster provide information as to the characteristic epoch during which the central galaxy harbored an active nucleus and gives important energy constraints. A *Chandra* image of the Perseus Galaxy Cluster, which illustrates the fine temperature maps that are available and the complex structures in these massive clusters, appears in Figure 2.

Similar studies can be performed in normal galaxies like the Milky Way. In these galaxies we can study the mechanisms which heat the interstellar gas to temperatures of several million degrees and often form galactic scale outflows extending for several kiloparsecs (i.e., several thousand light years) away from the galactic center and enrich the intergalactic space. *Chandra* also allows the detection of large numbers of discrete X-ray sources associated with supernova remnants, and accreting black-hole or neutron star X-ray binaries [16]. Investigation of their temporal and spectral properties can provide insight into the populations of black-holes and neutron stars in galaxies.

---

[14] The ionization state of a gas determines which elements are ionized and how many of their electrons are missing.

[15] URL: http://hea-www.harvard.edu/XJET/index.cgi

[16] Like black holes, *neutron stars* are remnants of collapsed stars. Because the collapsed star was less massive, however, neutron stars do not collapse indefinitely. Degenerate neutron pressure stops the collapse but not before the electrons of the atoms are forced into the atomic nuclei where they combine with protons to form neutrons and thus a neutron star. Neutron star X-ray binaries are stellar binary systems with matter flowing toward the gravitational potential of the neutron star. As in black hole X-ray binaries, the material falling into the neutron star release its gravitational energy in the form of X-ray radiation that may be visible to *Chandra*.
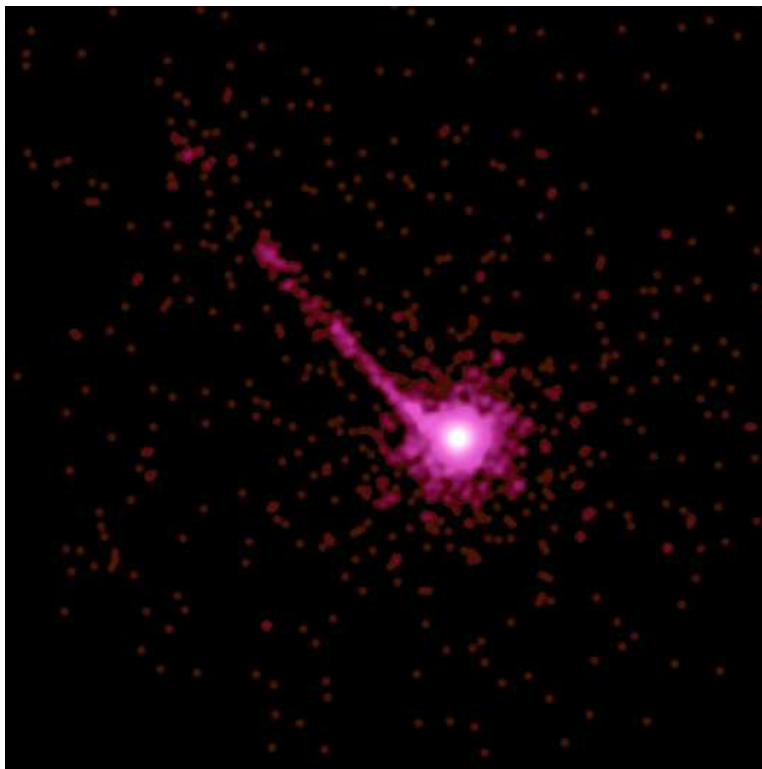
Figure 1: The Distant Quasar PKS 1127-145. The jet of outflowing hot matter toward the upper left of the color image is clearly visible. The jet is believed to be formed by gas swirling around a supermassive black hole. The length of the jet indicates that the explosive activity is long lived; the knots in the jet indicated that the activity is intermittent. (Image Credit: NASA/CXC/A.Siemiginowska(CfA)/J.Bechtold(U.Arizona), Bechtold et al. (2001), Siemiginowska et al. (2002))

With *Chandra*'s high-angular-resolution X-ray data astronomers can obtain temperature and chemical composition maps and study the velocity structures of the expanding gas in supernova remnants. This allows for tracing the history of this gas since the supernova explosion and is critical to our understanding of the evolution of a star and the final moments just before the supernova explosion. It also allows the study of the acceleration of cosmic rays and the production of heavy elements which are important for the creation of life.

Here we have described only a few examples of how the observations made possible with X-ray telescopes inform and develop our understanding of the physical world. *Uhuru* provided the first substantial observational evidence for the existence of black holes and showed that our galaxy is peppered with collapsed stars that radiate most of their energy as X-rays. More generally X-ray astronomy helps to explain the life
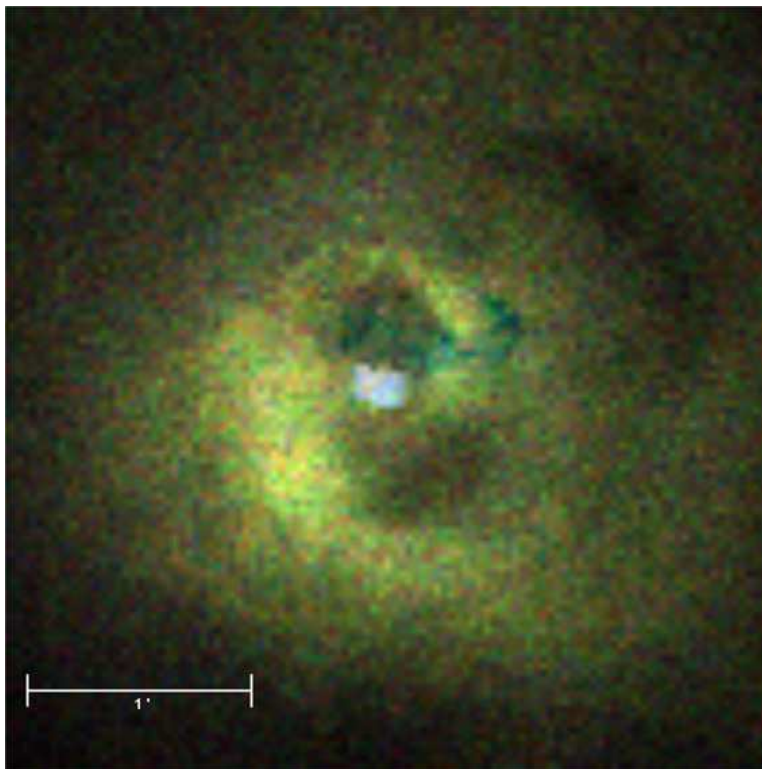
Figure 2: Core of the Perseus Galaxy Cluster. This is a temperate map of the galaxy cluster; red, green, and blue represent low, medium, and high temperatures, respectively. The map is constructed using the energies of the emitted photons. The dark region in the upper right of the image is a smaller galaxy that is falling into the central galaxy; it appears dark because gas in the galaxy absorbs X-rays. The bright blue spot in the center of the image is due to X-ray emission from hot gas falling into a giant black hole at the center of the super galaxy, Perseus A. The dark regions above and below the black hole are thought to be buoyant magnetized bubbles of energetic particles produced by energy released in the vicinity of the black hole. Each of these dark regions is large enough to contain a galaxy half the diameter of our Milky Way galaxy. The white line at the bottom of the image represents one arcminute, a distance of about 100,000 light year. (Credit: NASA/IoA/Fabian et al. (2000).)

cycle of stars, galaxies, and cluster of galaxies. The nature of the evidence that X-ray observations provide the many areas of astrophysical research is beyond the scope of this article. Here we only hope to give the reader a taste of the importance of X-ray astronomy in the exploration of the universe and to motivate the instrumentation on board *Chandra*, the classes of data the instruments provide, and the data analysis methods that we employ.

2.2   Data Collection and Instrumentation

Chandra took its place along side *The Hubble Space Telescope* and *The Compton Gamma-ray Telescope* as part of NASA's Great Observatories when it was launched by the Space Shuttle *Columbia* in July 1999. *Chandra* is by far the most precise X-ray telescope ever produced; it is able to produce images that are over thirty times sharper than those available from previous X-ray telescopes.

Detectors and Gratings. There are two detectors aboard *Chandra*; one is a high spatial resolution microchannel plate detector (the High Resolution Camera, or the HRC), and the other is an imaging spectrometer with higher spectral resolution (the Advanced CCD Imaging Spectrometer, or ACIS). Both instruments are essentially photon counting devices, and register the arrival time, the energy, and the (two-dimensional) direction of arrival of incoming photons. Because of instrumental constraints, each of the four variables is discrete; the high resolution of *Chandra* means that the discretization is much finer than was previously available. For example, spectral data collected with ACIS have over 1024 energy bins, known as *channels*. (ACIS spectra have intrinsic resolutions of $\frac{E}{\Delta E} \approx 30 - 50$.) The ACIS detector is composed of 10 CCDs, each of which has 1024×1024 pixels for spatial data; the pixels are 0.5 arcseconds wide. The timing resolution is generally dictated by the CCD frame readout time, which is of the order of 2-3 seconds. The HRC nominally has more spatial sampling resolution with pixels of size 0.13 arcseconds, and very high temporal resolution ($16\,\mu$s) but has virtually no spectral resolution, ($\frac{E}{\Delta E} \approx 1$). Because the data that is collected with either detector is discrete, it can be compiled into a four-way table of photon counts. Spectral analysis focuses on the one-way marginal table of energy data, spatial analysis or imaging on the two-way marginal table of sky coordinates, and timing analysis on the one-way table of arrival times.

It is possible to use one of two diffraction gratings with either of the two detectors. A diffraction grating is placed in the beam of X-rays and diffracts the photon by an angle that depends on the photon wavelength. (The wavelength of a photon is proportional to the reciprocal of its energy.) One of the two gratings, the High-Energy Transmission Grating Spectrometer (HETGS), is designed for high-energy X-rays; the other, the Low-Energy Transmission Grating Spectrometer (LETGS), is designed for low-energy X-rays. If *Chandra* is focused on a point source, such as a star, and a grating is in place the energy of the photons can be recovered from the locations where they are recorded on the detector. Thus, the gratings greatly increase the spectral resolution of both of the detectors. Because the spectral resolution obtained with gratings is dominated by the size of the image, however, the advantage of the grating for spectral analysis is diminished for more extended sources, such as nebula. Because the gratings also refract about 90% of the photons away from the detector, they are ordinarily only used with bright sources.

Measurement Errors. The data gathered with *Chandra*, although high-resolution, present a number of statistical challenges to the astronomer. *Chandra* focuses X-rays with mir-

rors. Because the mirrors do not focus perfectly, images are blurred. The so-called *point spread function* characterizes the probability distribution of a photon's recorded pixel location relative to its actual sky coordinates. Unfortunately, the shape of the scatter distribution varies across the detectors; it is symmetric and relatively tight in the center and becomes more asymmetric, irregular, and diffuse toward the edge. The scatter distribution can also vary with the energy of the incoming photon. Due to detector response, the energy of a photon is also subject to "blurring"; there is a distribution of potential recorded energies given the actual energy of a particular photon. Generally we refer to the blurring of the sky coordinates and/or energy as the *instrument effect*.

Combining the instrument effects for energy and for sky coordinates results in three dimensional blurring of the photon characteristics. Given the sky coordinates and energy of a photon, there is a distribution of the recorded sky coordinate pixel and recorded energy bin. Thanks to careful calibration, this distribution can be tabulated on a grid of values of the true sky coordinates and true photon energy. Since there are, for example, 4096×4096 pixels on the detector and 1024 energy bins, the resulting blurring six dimensional hyper-matrix can have over 2.9×10^{20} cells. (Here we use a grid of true sky coordinates and photon energies that is as fine as the detector resolution, i.e., $2.9 \times 10^{20} \approx (4096 \times 4096 \times 1024)^2$.) Clearly some simplification is required. For spectral analysis using a small region of the detector, the blurring of energies is more-or-less constant, which results in a reasonably sized (1024×1024) matrix. Thus, utilizing sparse matrix techniques results in efficient computation for marginal spectral analysis. Spatial analysis often involves only a subset of the pixels, reducing the dimension of the problem. One strategy is to assume the blurring matrix is constant across a large number of pixels and energy bins; thus, we might divide the energy bins into 4 groups and the pixels into 16 groups and assume that the instrument effect is constant in each of the resulting 64 cells. This strategy aims at computational feasibility with the hope that the compromise in precision is minor. A careful analysis of this trade off has yet to be tackled.

Stochastic Censoring. Another complication for data analysis involves the absorption of photons and the so-called effective area of the telescope. Depending on the energy of a photon, it has a certain probability of being *absorbed*, for example by the interstellar or intergalactic media between the source and the detector. *Effective area* is a characteristic of the telescope mirrors, but has similar consequences for the data. The mirrors on *Chandra* reflect the X-rays to focus them on the detector. Unfortunately, high-energy photons do not reflect uniformly or simply. Each X-ray has a certain probability of being reflected away from the detector or being absorbed by the telescope mirrors. Since this probability depends on the energy of the photon, the probability that a photon is recorded by the detector depends on its energy. This process results in non-ignorable missing data; both absorption and the effective area of the instrument must be accounted for to avoid bias in fitted spectra and images.

Background Contamination. The data are also degraded by *background* counts—X-ray photons which arrive at the detector but do not correspond to the source of interest.

In spectral analysis, a second data set is collected that is assumed to consist only of background counts. For example, background counts might be collected around, but some distance away from the source in a region of space that contains no apparent X-ray sources. After adjusting for exposure time and the area in which the background counts are collected relative to that in which the source counts are collected, it is standard practice to *directly* subtract the counts observed in the background exposure from those observed in the source exposure; the result is analyzed as if it were a source observation free of background contamination. This procedure is clearly questionable, especially when the number of counts per bin is small. It can lead to the rather embarrassing problem of negative counts and has unpredictable results on statistical inference. A better strategy is to model the counts in the two observations as independent Poisson random variables, one with only a background intensity and the other with intensity equal to the sum of the background and source intensities (Loredo 1992; van Dyk 2003)

Pile-Up. A final degradation of the data is known as *pile-up* and poses a particularly challenging statistical problem. Pile-up occurs in CCD X-ray detectors when two or more photons arrive at the same location on the detector (i.e., in an event detection island, which consists of several pixels) during the same time frame (i.e., time bin). Such coincident events are counted as a single higher energy event. The event is lost altogether if the total energy goes above the on-board discriminators. Thus, for bright sources pile-up can seriously distort the count rate, the spectrum, and the image. Accounting for pile-up is inherently a task of joint spectral-spatial modeling. A diffuse extended source may have no appreciable pile-up because the count rate is low on any one area of the detector. A point source with the same marginal intensity, however, may be subject to severe pile-up because most of the counts land in the same area of the detector. Even given the spatial structure of the source, the degree of pile-up depends on the source intensity. Thus, pile-up can make overall source intensity difficult to measure. Model based methods for handling pile-up are discussed in Kang et al. (2003); see also Davis (2001).

2.3 Data Analytic Goals and Statistical Methods

Broadly speaking, analysis of *Chandra* data falls into three categories: spectral analysis, spatial analysis or imaging, and timing analysis. Each of these categories has specific scientific goals and poses statistical challenges. In this section, we discuss each type of analysis in turn and conclude with a discussion of the possibility of modeling the joint distribution of these variables.

Spectral Analysis. The energy spectrum of an astronomical object can reveal important information as to the composition, temperature, and relative velocity of the object. For example, when an electron jumps down from one energy shell (i.e., quantum state) of an atom to another, the energy of the electron decreases. This energy is radiated away from the atom in the form of a photon with energy equal to the difference of the energies of the two electron shells. Because these energy differences are unique to each element, the

energy spectrum is the cosmic fingerprint of the elements that compose the source. It is from studying stellar spectra that astronomers learn that stars are composed mostly of hydrogen with some helium and traces of heavier elements such as oxygen, nitrogen, and carbon. Thus, spectral analysis is the cornerstone of X-ray astronomy.

Chandra's capacity for high resolution spectra means that it has a much finer discretization of energy than previous instruments. This results in an overall increase in the number of energy channels and leads to lower observed counts in each channel. (*Chandra* also has powerful mirrors that collect many more photons per unit time than previous instruments. This, however, allows astronomers to investigate many dimmer sources, which even with the powerful mirrors of *Chandra* result in few photon counts overall.) With few counts per bin, the Gaussian assumptions that might have been appropriate for data from older instruments are often inappropriate for *Chandra* data. For example, in so-called *minimum χ^2 fitting* (Lampton et al. 1976) one estimates the model parameter, θ, by computing

$$\hat{\theta} = \mathrm{argmin}_\theta \sum_{l=1}^{L} \frac{\{n_l - m_l(\theta)\}^2}{\sigma_l^2(\theta)}, \tag{1}$$

where L is the number of energy channels, n_l is the observed count in energy channel l, $m_l(\theta)$ is the expected count in channel l as a function of the model parameter θ, and $\sigma_l^2(\theta)$ is proportional to the sampling variance of n_l. The model for the expected counts per channel, $m_l(\theta)$ is generally parameterized in terms of quantities of specific scientific interest; this model accounts for the instrument effects and the effective area of the instrument as well as absorption. Because of the Poisson nature of the data, $\sigma_l^2(\theta)$ is often taken to be n_l or $m_l(\theta)$. It is obvious from its functional form that the right-hand side of (1) is an implicit Gaussian assumption. With large photon counts in each energy channel, this assumption is reasonable, and χ^2 fitting is essentially ML estimation. However, the intrinsically low-count data from high-resolution instruments such as *Chandra* are not approximately Gaussian; thus, parameter estimates and error bars computed with χ^2 fitting may not be trustworthy. To avoid this problem, astronomers often group the energy channels until there is a large enough count in each group to justify Gaussian assumptions. Doing so, however, reduces the information in the data and produces a less precise energy spectrum. In order to take advantage of the information that the new class of instruments provides, a method of analysis is needed that does not rely on large-count Gaussian assumptions (Siemiginowska et al. 1997; van Dyk et al. 2001).

Spectral emission lines are local features in the spectrum and represent extra emission of photons in a narrow band of energy. These features are used to model the emission resulting from electrons falling to a lower energy shell in a particular ion. Thus, emission lines are important in the investigation of the composition of a source. The Doppler shift of the location of a known spectral line (such as a particular hydrogen line) can also be used to determine the relative velocity of a source. Thus, determining the precise location of emission lines is a critical task; it is common for astronomers to conduct a formal hypothesis test as to whether a particular emission line should be included in a model. Unfortunately, the likelihood ratio test, or a Gaussian approximation thereof, is routinely used for this purpose. Since the intensity of emission lines are

generally constrained to be positive, the null hypothesis of no emission line is on the boundary of the parameter space. Thus, the standard asymptotic reference distribution of the likelihood ratio test does not apply and more sophisticated methods are needed for testing such hypotheses. See Protassov et al. (2002) for a Bayesian approach based on posterior predictive p-values.

Image Analysis. Unlike spectral analysis, spatial analysis or image analysis does not benefit from models that are parameterized in terms of quantities of specific scientific interest; this is because the spatial structure of the source can be highly irregular. Although some are simple point sources (i.e., delta functions) or collections of point sources, many are composed of diffuse nebula that have no particular predictable structure. Nonetheless the images contain valuable information as to the structure and evolution of X-ray sources.

A standard method of image analysis is Poisson image reconstruction—i.e., ML estimation under an independent Poisson model for the photon counts in each pixel that accounts for both the point spread function and background contamination. In astrophysics this image reconstruction method is known as the Richardson-Lucy Method (Richardson 1972; Lucy 1974). An attractive feature is that no assumptions are made about the underlying structure in the source. The downside of the lack of structural assumption is that the reconstructed image may be of low quality. The model is fit using an EM algorithm that is generally stopped before convergence because at convergence the reconstruction is often very grainy. An alternative strategy, which we explore in Section 3.4, is to quantify the prior belief that the reconstructed image should be smooth into a formal prior distribution. (There are many other less model-based methods such as variants of kernel smoothers that are in common use. Although such ad hoc smoothing routines can produce beautiful images, it is difficult to identify their inherent model assumptions, to quantify their fitting errors, or to access their reliability.)

Timing Analysis. Most astronomers use a few standard descriptive methods to handle *light curves*, i.e., time series data. Fourier transforms and 'folding' or 'binning' data on a known or a supposed period are common methods for periodic (Leahy et al. 1983; Gregory and Loredo 1992) or quasi-periodic sources (van der Klis 1997). Light curves that vary irregularly are often thought to be 'shot noise' (i.e., composed of individual pulses or flares with a sharp rise and slow decay), transitions between two or more quasi-stable states, or indicators of chaotic processes such as a lumpy accretion flow. Auto-correlations or cross-correlations (across multiple energy bands) are popular (with the shot-noise decay rate related to the autocorrelation function), but how best to characterize aperiodic light-curves is a significant open question. A related question is how to quantitatively compare two light-curves from possibly similar sources or the same source measured during different time periods. Over the past decade this has brought interesting non-parametric methods into the field, including chaos and fractal analyses (Perdang 1981; McHardy and Czerny 1987; Lochner et al. 1989; Meredith et al. 1995; Kashyap et al. 2002), wavelet-based methods (Slezak et al. 1990; Kolaczyk and Dixon 2000), newer change-point methods such as Bayesian Blocks

(Scargle 1998), and more general Poisson and multiscale analyses (see the *Time Series Analysis* section of Babu and Feigelson 1997; Young et al. 1995; Kolaczyk and Nowak 1999; Nowak and Kolaczyk 2000). Many of these multiscale methods are now finding their way into imaging analyses (Kolaczyk and Dixon 2000; Nowak and Kolaczyk 2000; Willett et al. 2002, see also Section 3.4).

Joint Analysis. These three types of analysis are almost always embarked upon separately. For example, analysis of how the energy spectrum varies across an extended source are only conducted in an ad hoc fashion, perhaps by dividing the image into two or three regions and modeling the spectra separately in each region. In addition to the scientific questions of understanding how the composition, temperature, and relative velocity vary across the source, there are statistical reasons for joint analysis. For example, if the spectrum varies across a source, the absorption rate and the instrumental effects related to the point-spread function and the detector response may vary as well; this varying censoring rate biases an image analysis that ignores the spectrum. Similar concerns arise for pile-up: The effect of pile-up on a spectrum depends on the source brightness and it is very different for point sources and diffuse extend sources. It becomes even more unpredictable in a varying source, in which the source intensity changes with time, thus modifying the amount of pile-up. Joint analysis of *Chandra* data is an important topic that is as of yet largely unexplored.

3 Bayesian Deconvolution

Because of the nature both of the instrumentation and of the astronomical sources themselves, deconvolution methods play a key role in the analysis of *Chandra* data. The point spread function, for example, means that each detector pixel count in an image is the sum of a number of counts, each of which corresponds to a pixel on the source. Our goal is to sort each of the detector counts into their proper pixel, that is the pixel where they would be recorded if the point spread function were a delta function. Of course, this is a lofty and unattainable goal. Instead, we use Bayesian methods to reconstruct the most probable origin of each photon and to quantify the error in the deconvolved image. Similar concerns arise because of the blurring of the energy in spectral analysis and because of background counts being added to source counts. In this section we outline four examples of Bayesian deconvolution of *Chandra* data. The four cases differ in their complexity and the degree and quality of the prior information that is available. In some cases the prior information takes the form of a neatly parameterized model and in others it is simply a vague notion of smoothness in the reconstructed image. We begin with a simple example, where standard Bayesian techniques lead easily to new methods that offer a dramatic improvement over standard practice.

3.1 A Simple Low-Resolution Example: Hardness Ratios

Scientific Motivation and Models. A hardness ratio is the coarsest description of a spectrum—the ratio of the expected counts in the lower energy end of the spectrum and the expected counts in the higher energy end. The lower energy end of the spectrum is called the *soft* end, while the high-energy end is called the *hard* end; hence the name *hardness ratio*. As we discussed in Section 2.1, the shape of, for example, a stellar spectrum is informative as to the physical processes of the star. Thus, even a coarse measure like the hardness ratio can be used to categorize X-ray sources.

Computing hardness ratios is a deconvolution problem because both the hard and soft counts are contaminated with background. Formally, $X_S = \eta_S + \beta_S$ and $X_H = \eta_H + \beta_H$, where X represents the observed counts, which is a convolution of the source counts, η, and the background counts β; the subscripts S and H represent soft and hard counts, respectively. The source and background counts in both energy bands are unobservable independent Poisson random variables,

$$\eta_S \sim \text{Poisson}(\mu_S) \text{ and } \eta_H \sim \text{Poisson}(\mu_H) \tag{2}$$

and

$$\beta_S \sim \text{Poisson}(\xi_S) \text{ and } \beta_H \sim \text{Poisson}(\xi_H), \tag{3}$$

marginally, $X_S \sim \text{Poisson}(\mu_S + \xi_S)$ and $X_H \sim \text{Poisson}(\mu_H + \xi_H)$. In this notation, the hardness ratio is $\rho = \mu_S/\mu_H$. Background observations are used to help identify the background parameters, ξ_S and ξ_H. These counts are taken, for example, from an annulus around the source of interest and modeled as independent Poisson random variables, $B_S \sim \text{Poisson}(c\xi_S)$ and $B_H \sim \text{Poisson}(c\xi_H)$, where B represents the counts from the background observation and c is a known constant that accounts for differences in the exposure area and the exposure time of the source and background observation.

Inference and Computation. Hardness ratios are typically estimated from data using a simple technique based on the methods of moments. The soft source count, η_S for example, can be estimated by $X_S - B_S/c$ and the hardness ratio by

$$\hat{\rho} = \frac{X_S - B_S/c}{X_H - B_H/c}. \tag{4}$$

The delta method can then be used to compute error bars for the estimated ratio.

Although we might expect this standard method to exhibit reasonable frequentist properties with large counts, hardness ratios are often used to describe weak sources with very low counts. Hardness ratios are attractive summaries of the spectrum for such sources because more sophisticated spectral analysis is not possible. It is not uncommon for either or both of the hard and soft counts to be zero; for example, a catalog of sources from a visible light survey conducted with *Hubble* may be studied with *Chandra*. Some of these sources will not show up at all in the X-ray survey—i.e., both the hard and soft X-ray counts are zero. In this case, it is evident and simulation studies verify (see Figure 3) that the method of moment and the delta method are inadequate. Because

no reliable statistical methods are available, astrophysicists either give up or calculate incorrect (one-sided) intervals.

Fortunately, more sophisticated Bayesian methods are readily available. We wish to summarize the posterior distribution, $p(\rho|X_S, X_H, B_S, B_H)$, perhaps using proper prior information for the model parameters $(\mu_S, \mu_H, \xi_S, \xi_H)$. Since hardness ratios are often computed on a survey of sources, we might formulate the prior distribution hierarchically and model the distribution of the model parameters across sources in the population. For simplicity of presentation, we ignore these issues here and simply use a flat prior distribution on the model parameters; see Park et al. (2004) for a more complete presentation.

To evaluate the posterior distribution of ρ, we begin with the posterior distribution

$$p(\mu_S, \mu_H|X_S, X_H, B_S, B_H) = \left\{ \int p(\mu_S, \xi_S|X_S, B_S)d\xi_S \right\} \left\{ \int p(\mu_H, \xi_H|X_H, B_H)d\xi_H \right\}.$$
(5)

The independence of the hard and soft counts allows us to factor the joint posterior distribution of the model parameters into a soft and a hard component. If we use flat prior distributions, the integrand corresponding to the soft component is proportional to the product of the Poisson likelihoods of $X_S \sim \text{Poisson}(\mu_S + \xi_S)$ and $B_S \sim \text{Poisson}(c\xi_S)$, likewise for the integrand corresponding to the hard component. The two integrals in (5) are available in closed form, which gives us the joint posterior distribution of μ_S and μ_H is closed form. We can then transform via, say, $\rho = \mu_S/\mu_H$ and $Q = \mu_S\mu_H$ and (numerically) marginalize to obtain the posterior distribution of ρ. Computing this distribution on a grid, we can easily compute the MAP estimate and HPD interval. Such summaries of the posterior distribution of $\log(\rho)$ might be more informative from a statistical point of view. If the posterior mean and equal-tail intervals are desired, it is an easy task to construct an MCMC sampler based on the method of data augmentation. If η_S and η_H are treated as missing data, the required complete conditional distributions are all Gamma and Binomial distributions. Thus, a simple sampler is readily available.

We emphasize that this example is not meant to illustrate sophisticated statistical methods but rather to show how simple and straightforward Bayesian techniques can readily solve outstanding statistical problems in astrophysics. In the next several sections, we show how these same ideas can be used to develop new methods for unlocking high-resolution *Chandra* data.

3.2 High Resolution Deconvolution Methods

To deconvolve high-resolution *Chandra* data, we must account not only for background contamination, but also for instrument effect (i.e., blurring) and the effective area of the instrument. The data consist of an $L \times 1$ vector of counts; these may be pixel counts from an image, counts from a set of energy channels, or counts from the set of bins constructed by crossing the image pixels with energy channels. Generally we refer to these as the counts in the *detector bins* or the *detector counts*. This is to distinguish these counts from the idealized counts that we would observe with a perfect instrument—i.e.,
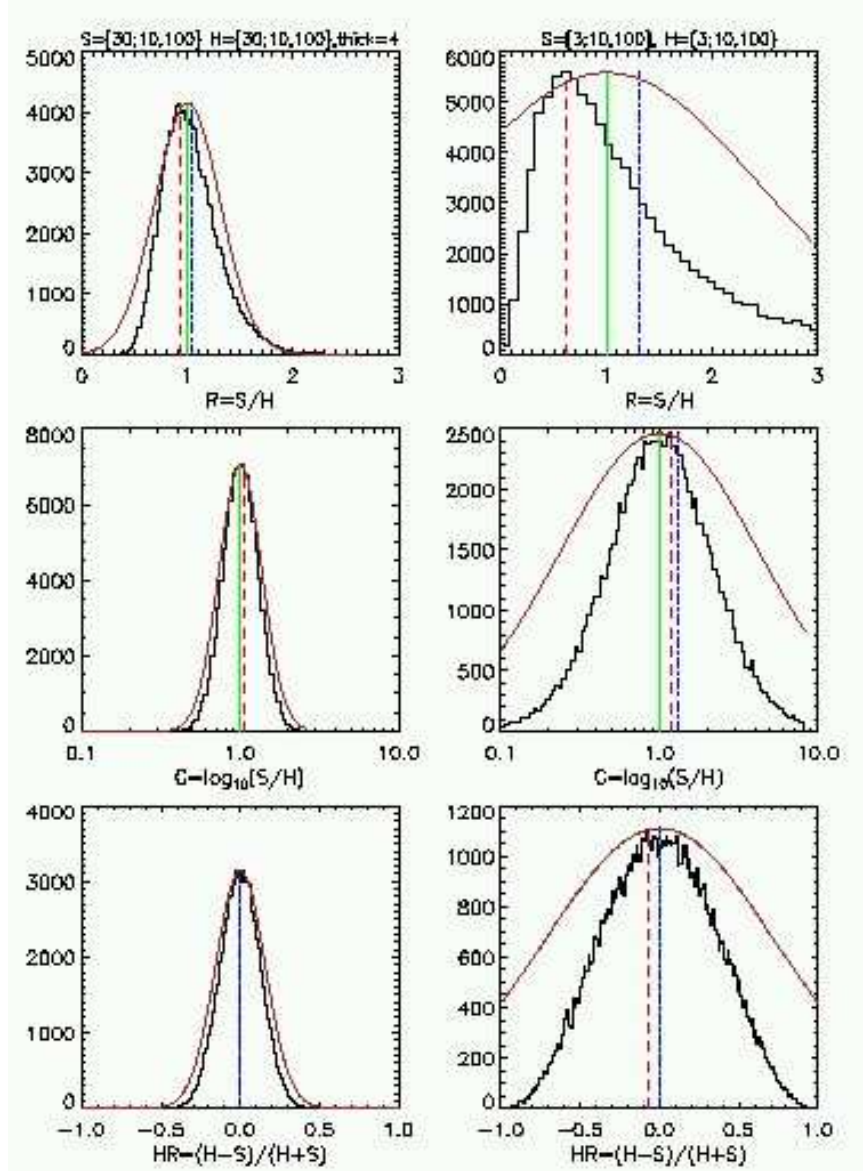
Figure 3: Sampling Distributions of Different Types of Hardness Ratios for High Counts (left column, $\mu_S = \mu_H = 30$, $\xi_S = \xi_H = 0.1$, and $c = 100$) and Low Counts (right column; $\mu_S = \mu_H = 3$, $\xi_S = \xi_H = 0.1$, and $c = 100$). The top two panels correspond to the simple hardness ratio, $\rho = \mu_S/\mu_H$, denoted R=S/H in the figure; the middle two panels correspond to the color, $\log_{10}(\rho)$, denoted C=$\log_{10}$(S/H) in the figure; the bottom two panels correspond to the fractional difference, $(\mu_H - \mu_S)/(\mu_H + \mu_S)$, denoted HR=(H-S)/(H+S) in the figure. The black histogram outline is the sampling distribution of the method of moments estimators derived with a Monte Carlo simulation. The mode of the Monte Carlo distribution is marked by the red dashed line and the arithmetic mean of the sampled values by the blue dot-dashed line. The brown curves are Gaussian distributions centered on the true value, marked by a solid brown line, with standard deviation computed using the delta method and evaluated at the true parameter values. That the Gaussian approximation to the sampling distribution fails in low-count scenarios is evident.

an instrument without blurring, with constant effective area, and without background contamination. We refer to the idealized counts as the counts in the *ideal bins* or the *ideal counts* and emphasize that ideal counts are necessarily missing data. There is no need for the number of ideal bins and detector bins to be the same; indeed, they are generally not the same for spectral analysis. Thus, we suppose the ideal unobserved data are a $J \times 1$ vector of ideal counts.

We quantify the effective area of the detector as a vector of probabilities, one for each ideal bin; the effective area of ideal bin j is the probability that a photon that arrives at the detector corresponding to ideal bin j is recorded. We tabulate the effective area as a diagonal $J \times J$ matrix, $\mathbf{A}$, with diagonal elements equal to these probabilities. If a photon arriving at the detector and corresponding to ideal cell j is recorded, it may be recorded in one of several detector bins. This is because of the instrument effect (i.e., blurring). We tabulate the instrument effect in an $L \times J$ matrix, $\mathbf{P} = (p_{lj})$, where p_{lj} is the probability that a photon corresponding to ideal bin j that is recorded by the detector is counted in detector bin l. Thus, the columns of $\mathbf{P}$ are probability vectors. The expected background count in each detector cell is an $L \times 1$ vector that we denote $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_L)^\top$

We can now formulate the mean structure of our high-resolution deconvolution model as

$$\boldsymbol{\lambda} = \mathbf{P}\mathbf{A}\boldsymbol{\mu} + \boldsymbol{\xi}, \tag{6}$$

where $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_L)^\top$ is the vector of expected detector counts and $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_J)^\top$ is the vector of expected ideal counts. Our goal is to *reconstruct* or *deconvolve* the expected ideal counts, $\boldsymbol{\mu}$, from the observed data: L independent Poisson counts, $X_l \sim \text{Poisson}(\lambda_l)$ for $l = 1, \ldots, L$. In some cases we also have a background observation, $B_l \sim \text{Poisson}(c\xi_l)$, for $l = 1, \ldots, L$ and some known c.

Solving (6) via maximum likelihood is a standard Poisson image reconstruction problem that can be handled with an EM or an EM-type algorithm that treats the ideal counts as missing data (Shepp and Vardi 1982; Vardi et al. 1985; Lange and Carson 1984; Fessler and Hero 1994; Meng and van Dyk 1997). A similar computational strategy can be used to construct MCMC samplers based on the method of data augmentation. Generally, however, we wish to include some form of prior information or prior constraints on $\boldsymbol{\mu}$. Markov random fields are a common and general strategy for incorporating a smoothing prior on $\boldsymbol{\mu}$; the resulting MAP estimate can be computed using similar computational techniques. Here we discuss different strategies by examining three applications of (6) to modeling *Chandra* data. We begin with a highly structured parameterized model for $\boldsymbol{\mu}$ that is used for spectral analysis. We then discuss a multiscale smoothing prior that is used for image analysis, and conclude with hierarchical application of (6) that aims to explore the temperature distribution a stellar corona via ultra-high-resolution spectral analysis.

3.3   Spectral Analysis

In a spectral model, we describe the distribution of the photon energies emanating from a particular source. In the context of (6), we aim to express μ_j as a function of E_j, the energy of ideal energy bin j. In a typical spectral model this function has three basic components: a continuum term, spectral lines, and absorption features. The photons that are emitted from the source are a mixture of photons from the continuum and from several emission lines. Because of absorption, either at the source or between the source and the detector, not all of these photons reach the telescope.

The continuum is formed by thermal (heat) radiation from the hot center of stars to the cold space that surrounds them or by non-thermal processes in relativistic plasmas. The continuum is modeled using a smooth parametric form; several important forms can be expressed as log linear models, linear in a transformation of energy. A power law continuum, for example, can be written αE_j^γ, the log of which is a linear in $\log(E_j)$, where α and γ are unknown model parameters. Emission lines, discussed in Section 2.3, appear as narrow ranges of energy with more counts than would be expected from the continuum. Emission lines are modeled as narrow Gaussian or Lorentzian distributions or as delta functions that are added to the continuum; in statistical terms this can be formulated as a finite mixture model. Finally, because absorption of photons occurs independently across the photons, absorption is represented by a multiplicative factor that represents the complement of the censoring probability. Like the continuum, some absorptions models can be represented by a generalized linear model, in this case with complementary log-log link and a Bernoulli model. Combining the continuum, emission lines, and absorption terms we can express the spectral model as

$$\mu_j = \left\{ \delta_j f(\theta^{\mathrm{C}}, E_j) + \sum_{k=1}^{K} \theta_k^{\mathrm{E}} p_{jk}^{\mathrm{E}} \right\} g(\theta^{\mathrm{A}}, E_j), \tag{7}$$

where δ_j is the width of ideal energy bin j, $f(\theta^{\mathrm{C}}, E_j)$ represents the continuum with parameter θ^{C}, K is the number of emission lines, θ_k^{E} represents the expected ideal count due to emission line k, p_{jk}^{E} is the proportion of emission line k that falls into ideal energy bin j, and $g(\theta^{\mathrm{A}}, E_j)$ represents the absorption model with parameter θ^{A}—i.e., $1 - g(\theta^{\mathrm{A}}, E_j)$ is the probability that a photon with energy E_j is absorbed, and, thus, not observed. We can also parameterize p_{jk}, for example, in terms of the location and width of the Gaussian line profile. Here superscripts C, E, and A represent 'continuum', 'emission lines', and 'absorption features', respectively. This model was introduced and illustrated by van Dyk et al. (2001); Sourlas et al. (2003), Sourlas et al. (2004), and van Dyk and Kang (2004) give further applications. Hans and van Dyk (2003) and van Dyk and Hans (2002) present generalized linear models to account for absorption features.

Prior information for the various parameters in (7) can often be quantified in terms of semi-conjugate prior distributions. Because this spectral model is highly structured, little prior information may be necessary—especially for relatively bright sources. For some parameters, however, such as the locations of the emission lines, prior information is nearly always helpful. Fortunately, such prior information is often scientifically
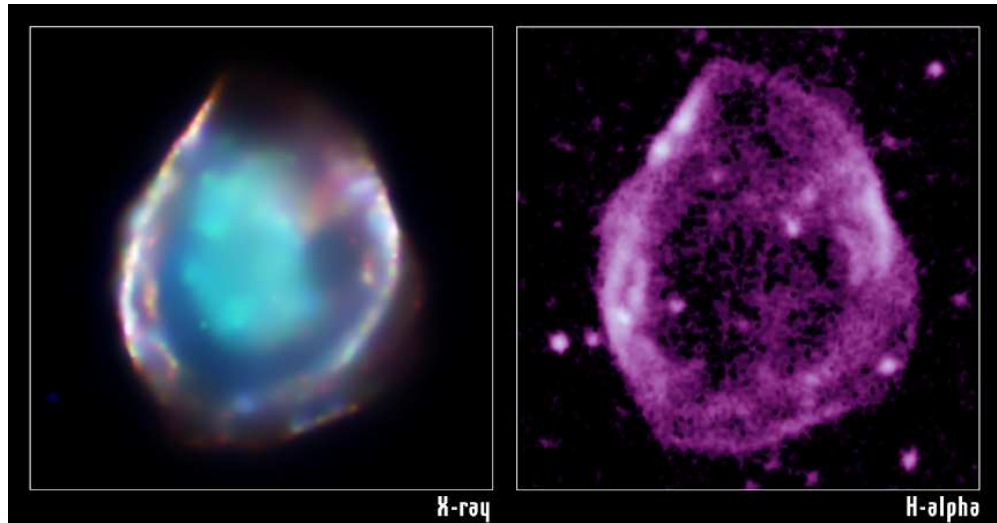
Figure 4: Supernova Remnant DEM L71. The color image on the left is an X-ray image obtained with *Chandra*; that on the right is a *Hubble* optical image. The X-ray image exhibits the smooth irregular structure that is typical of extended sources. (Image Credit: X-ray: NASA/CXC/Rutgers/Hughes et al. (2003) Optical: Rutgers Fabry-Perot.)

forthcoming. For example, there are key spectral lines that might be expected in a particular spectrum. Nonetheless, the predictability of the structure in typical spectra is a powerful resource in this Poisson reconstruction problem.

3.4   Imaging

Irregular Structure. Reconstruction of X-ray images is more challenging because their structure is much less predictable than that of X-ray spectra. We illustrate this by examining several *Chandra* images. Figure 4 shows an X-ray image (left panel) and an optical image (right panel) of the supernova remnant DEM L71. The color scheme in the X-ray image is a representation of the energy of the photons: red corresponds to lower energy X-rays, green to midrange energies, and blue to high energies. The X-ray image reveals a hot inner cloud of glowing iron and silicon surrounded by an outer blast wave; the blast wave is also visible in the optical image[17]. Spectral analysis reveals that this super nova resulted from an exploding white dwarf star. This type of super nova results when a white dwarf pulls too much material off a nearby companion, becomes unstable, and is blown apart in a thermonuclear explosion. The *Chandra* image in Figure 4 illustrates the smooth but irregular features that can appear in extended X-ray images.

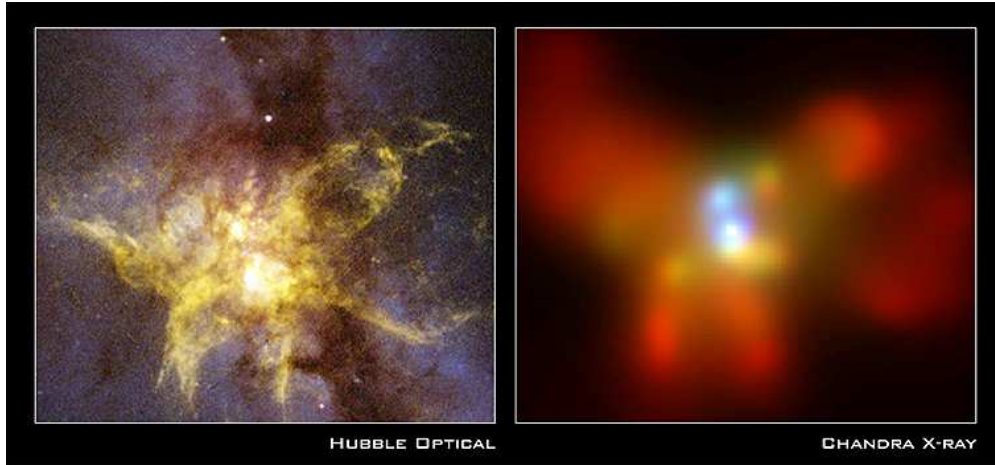[17]URL: http://xrtpub.harvard.edu/photo/2003/deml71/index.html

Figure 5: The Galaxy NGC 6240. The color image on the left is a *Hubble* optical image; that on the right is a *Chandra* X-ray image. In addition to smooth irregular structure, the X-ray image reveals two point sources, which are active giant black holes. The *Chandra* image was compiled using Gaussian smoothing of the raw data. We illustrate Bayesian methods for reconstructing this image. (Image Credit: X-ray: NASA/CXC/MPE/Komossa et al. (2003); Optical: NASA/STScI/R.P.van der Marel & J.Gerssen.)

In addition to smooth extended features, X-ray may include one or more bright point sources. This is illustrated in Figure 5 and Figure 6. Both figures represent galaxies; the same color scheme is used for the X-ray images as in Figure 4. Figure 5 illustrates both optical (left-hand panel) and X-ray (right-hand panel) images of NGC 6240, a galaxy that is the product of the collision of two smaller galaxies. The X-ray image reveals two bright point sources which are giant active black holes. Figure 6 is an X-ray image of M 83, and, is peppered with point sources that are believed to be neutron stars and black holes.

Basic Model. The figures illustrate not only the irregular structure that is typical of extended sources, but also the need to allow for bright point sources in the image. To accomplish this, we parameterize the expected ideal counts as

$$\mu_j = \theta_j^{\text{ES}} + \sum_{k=1}^{K} \theta_k^{\text{PS}} p_{jk}, \tag{8}$$

where θ_j^{ES} is the expected ideal count in pixel j due to the extended source, K is the number of point sources, θ_k^{PS} is the expected ideal count from point source k, and p_{jk} is the expected proportion of the expected ideal count for point source k that falls into ideal bin j. Here the superscripts ES and PS represent 'extended source' and 'point source', respectively.
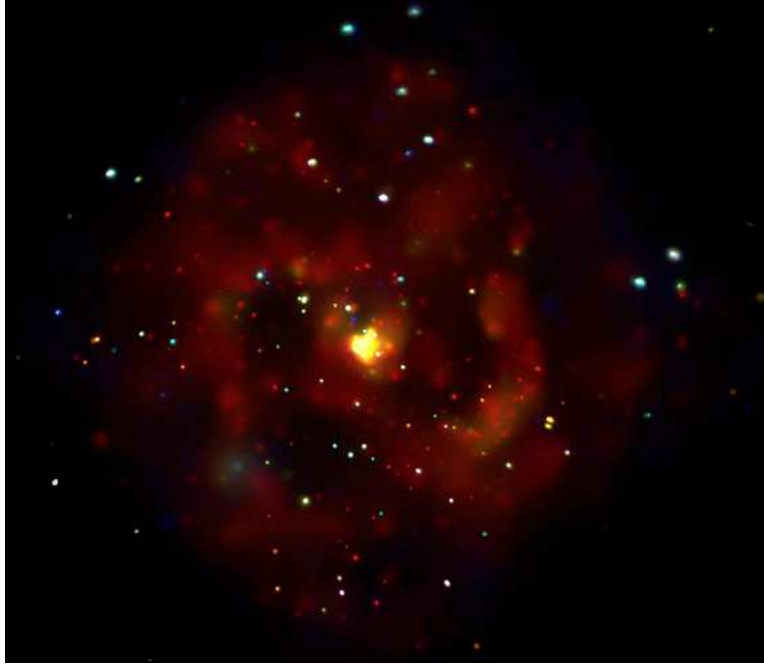
Figure 6: The Galaxy M83. This color image is dominated not by its smooth extended source but by its numerous neutron stars and black holes which appear as point sources. (Image Credit: NASA/CXC/U.Leicester/U.London/R.Soria & K.Wu.)

Multiscale Model and Smoothing Prior Distribution. Model (8) does not yet incorporate smoothness into the extended source. To do this, we use a version of Nowak and Kolaczyk's (2000) multi-scale model (see also Kolaczyk (1999)). As a simple example of how this is done suppose we have a 4×4 grid of ideal pixels, with ideal counts x_{ij} from the extended source in pixel (i,j). We emphasize that these counts are not observed, but we can, nonetheless, hierarchically formulate our model in terms of the ideal counts. Here we use two indexes to identify each of the counts; the first identifies a quadrant of the image, the second a sub-quadrants of each quadrant. Because we are considering a 4×4 grid of pixels, this is enough to identify each pixel; see the bottom diagram in Figure 7. We model the x_{ij} as independent Poisson random variables, $x_{ij} \sim \text{Poisson}(\theta_{ij}^{\text{ES}})$. Let $x_{++} = \sum_{ij} x_{ij}$ be the sum of the pixel counts for the image and $x_{i+} = \sum_{j} x_{ij}$ be the sum of the pixel counts for each of the four quadrants; see Figure 7. We can rewrite the joint distribution of the pixel counts in terms of conditional distributions given the lower-resolution counts (i.e., the quadrants total counts and image total count),

$$p(x_{11}, \ldots, x_{44} | \boldsymbol{\theta}^{\text{ES}}) = \left\{ \prod_i p(x_{i1}, \ldots, x_{i4} | x_{i+}, \boldsymbol{\theta}^{\text{ES}}) \right\} p(x_{1+}, \ldots, x_{4+} | x_{++}, \boldsymbol{\theta}^{\text{ES}}) p(x_{++} | \boldsymbol{\theta}^{\text{ES}}),$$
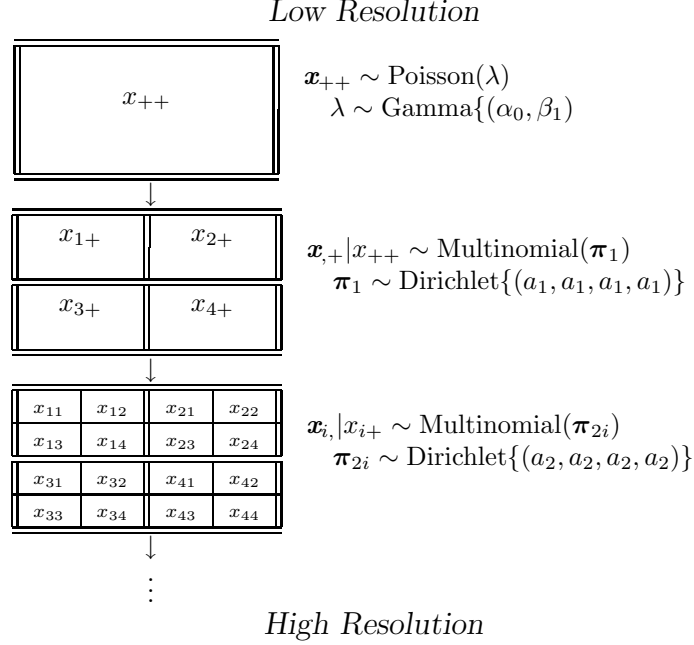
(9)

Low Resolution

Figure 7: A Multi-Scale Parameterization. The top layer represents the Poisson model and prior distribution for the total of the ideal counts. Each subsequent layer splits the cells of its parent layer into four parts. We model these splits using multinomial distributions with Dirichlet prior distributions. Since the hyper-parameters of the Dirichlet distributions are all equal, the prior distributions tend to shrink the multinomial split probabilities toward equal splits, which in turn favor smoother image reconstructions at each level of resolution.

where $\boldsymbol{\theta}^{\mathrm{ES}} = \{\theta_{ij}^{\mathrm{ES}}\}$. The first five terms in (9) are multinomial distributions; the last term is a Poisson distribution. Specifically, we write

$$x_{++} \sim \mathrm{Poisson}(\theta_{++}^{\mathrm{ES}}), \tag{10}$$

$$(x_{1+}, \ldots, x_{4+})|x_{++} \sim \mathrm{multinomial}(\boldsymbol{\pi}_1), \text{ for } i = 1, \ldots, 4, \tag{11}$$

and

$$(x_{i1}, \ldots, x_{i4})|x_{i+} \sim \mathrm{multinomial}(\boldsymbol{\pi}_{2i}), \tag{12}$$

where $\theta_{++}^{\mathrm{ES}} = \sum_{ij} \theta_{ij}^{\mathrm{ES}}$ and the parameters of the multinomial distributions are simple transformations of $\boldsymbol{\theta}^{\mathrm{ES}}$. This is a simple and standard reformulation of the model. The reason we consider this reformulation is that it allows us to easily specify a prior distribution that favors smooth reconstructions. In particular, the prior distribution $\boldsymbol{\pi}_1 \sim$ Dirichlet$\{(a_1, a_1, a_1, a_1)\}$ shrinks the fitted expected quadrant counts toward equality, i.e,. a smooth image at this level of resolution. The larger a_1 is the smoother the reconstruction is at this resolution. Similarly, we specify $\boldsymbol{\pi}_{2i} \sim$ Dirichlet$\{(a_2, a_2, a_2, a_2)\}$,

using the same hyper-parameter for each multinomial at this level of resolution. We, however, use different hyper-parameters at different levels of resolution, allowing for differing amounts of smoothing.

For large images, the idea is completely analogous, but notationally more tedious; with a 256×256 image, there are eight hyper-parameters. Nowak and Kolaczyk (2000) suggest strategies for the user to set the hyper-parameters. We explore another option, by fitting the hyper-parameters given a common hyper-prior distribution, $a_l \sim$ gamma(b_0, b_1), where only b_0 and b_1 are set by the user.[18]

Future Work. There are a number of outstanding statistical issues in this model. Perhaps most important is the character of the point spread function. As we have mentioned, $\mathbf{P}$ is known to vary significantly across the detector. Even at a fixed location on the detector the components of $\mathbf{P}$ are measured with error. A simulation is used to generate $\mathbf{P}$, which introduces both Monte Carlo error and systematic error since, for example, the simulation requires specification of the unknown spectrum of the image.

Other challenges involve components in the structure of the image. As we discuss and illustrate in the example below, astrophysicists often would like to test for structures such as point sources, jets, or loops of hot gas in the image. One possible strategy is to generalize (8) by adding additional structures. For example, a jet often appears as a string of clumps extending from a source, which we might model using a set of bivariate Gaussians with a constraint on their mean structure. Formal tests can be constructed using the posterior predictive distribution of the LRT statistic under the simpler model. We suspect that the power of these tests will depend on the relative smoothness of the extended source and model feature. These issues are discussed below in the context of an example.

The model in (8) is introduced by van Dyk and Hans (2002); the multi-scale prior distribution is introduced by Esch (2003) and Esch et al. (2004). We now turn to an application of the model; other applications and simulations can be found in Esch et al. (2004).

Example: Reconstruction of NGC 6240. The X-ray image of NGC 6240 in Figure 5 is smoothed via a Gaussian kernel, a standard smoothing method in high energy astrophysics. The original data appear in the first frame of Figure 8, where brighter pixels represent higher counts. (This and the other representations of NGC 6240 in this section are plotted on the log scale.) Methods based on kernel smoothers are generally preferred to Richardson-Lucy (i.e., computing the ML estimate under the Poisson image reconstruction model) because they return much smoother and visually appealing images. Richardson-Lucy reconstructions of NGC 6240 appear in the bottom row of Figure 8; the reconstructions result from running the Richardson-Lucy algorithm starting from a smooth image for 20 and 100 iterations, respectively. These reconstructions are much

---

[18]The actual prior distribution is somewhat more complicated; it involves a mixture of priors that vary the origin of the multiscale representation. See the rejoinder and Esch et al. (2004) for details.

grainer than the image produced by Gaussian kernel smoothing, and become grainer as the algorithm proceeds.

The final frame of Figure 8 is the posterior mean of the image, $\boldsymbol{\mu}$, under the multi-scale prior distribution. We use a gamma prior distribution with shape parameter 2 and scale parameter equal to 0.01 for the multi-scale hyperparameters $(a_1, \ldots, a_8)$; there are eight hyperparameters for this 256×256 image because $256 = 2^8$. To reduce sensitivity of the final result to the choice of coordinates in the multi-scale prior distribution, at each iteration we randomly select a pixel that is treated as the origin in the multi-scale specification. The image shown here was generated from a single Markov chain of 3000 total iterations, discarding the first 1000. We know, by subsequently examining Gelman and Rubin's (1992) statistics from two independent chains, that convergence was sufficiently attained (all $\hat{R}$ statistics for image intensities and smoothing parameters less than 1.1) in less than 300 iterations; hence we are quite comfortable in claiming convergence for the single chain.

The resulting reconstruction is clearly much smoother than the Richardson-Lucy reconstructions and does not require the user to decide when to stop the algorithm to avoid an over-fitted reconstruction. Relative to the image fit with Gaussian kernel smoothing, our reconstruction preserved much more structure in the image. Notice, for example, the loop of hot gas in the upper right quadrant of our reconstruction. This loop also appears in the optical image produced by *Hubble*, but has been completely smoothed out by the Gaussian kernel. To sharpen the comparison with the *Hubble* image, Figure 9, overlays the optical image (in red) first with the original data and second with our reconstruction (both in blue). Our reconstruction of the X-ray image matches up very well with the optical image. Although optical and X-ray images often highlight different structures in the source and therefore need not match up, the fact that they do match in this case is a strong validation of our reconstruction.

One of the primary benefits of using model-based methods for image reconstruction is that they not only provide a reconstructed image but also measures of error on the reconstruction. In the Bayesian setting, we not only have a posterior mean, but also the variance or quantiles of the posterior distribution. Figure 10 is an attempt to summarize the high dimensional posterior distribution of $\boldsymbol{\mu}$. The first frame in the second row of Figure 10 is our *significance map*, where the intensities in the image are generated by dividing the posterior mean by the posterior standard deviation; the colormap in the image is black for all intensities less than three. Thus, only the brightest pixels as measured by the pixel posterior standard deviations appear. The second frame in the second row is the same except that a threshold of just one pixel posterior standard deviations is used; thus, more pixels are lit up. These *significance maps* reflect an effort to determine which structures in the reconstructed image represent physical structures in NGC 6240. The loop of hot gas in the upper right quadrant of the image, for example, appears in both the one standard deviation and the three standard deviation significance maps. The larger loop to the left of the image only appears dimly in the one standard deviation map. Thus, we are more confident that the former loop represents a physical structure in NGC 6240 than we are of the latter loop.

We can also summarize the variability in the posterior disturbing for $\boldsymbol{\mu}$ by constructing a data cube, where the first two-dimensional data array is the first simulation of $\boldsymbol{\mu}$ from the MCMC sampler and the subsequent two-dimensional arrays correspond to the subsequent simulations of $\boldsymbol{\mu}$. We can then view the data-cube as a movie that proceeds through the sequence of draws generated by the MCMC sampler. Significant structures in the image appear consistently through the film, while less significant structures appear to flicker on and off.

Formally testing for the presence of features in an image poses a significant challenge. The significance maps are only a crude approximation; they rely on the marginal posterior standard deviation of each pixel intensity. The use of the standard deviation is difficult to justify if the marginal distributions are not roughly symmetric, which they may not be in low intensity regions of the image. Physical structures such as loops or jets are represented by complex combinations of pixel intensities. Since these intensities are likely to be correlated in the posterior distribution, independent evaluation of the marginal distributions of the pixel intensities may not suffice. Standard statistical techniques such as posterior predictive p-values (Gelman et al. 1996) do not appear to solve the problem because the model for the extended source is so flexible. The multiscale model allows for loops and jets and it is not clear how to constrain the model to eliminate such structures in a "null" model. Clearly the posterior distribution of $\boldsymbol{\mu}$ is the right summary of the information in the data, but works needs to be done to parse this summary into useful scientific information.

3.5   Reconstruction of a Differential Emission Measure

Exploring the Physical Enviornment of a Stellar Corona. As a final example, we show how we can use (6) to explore the distribution of the temperature of the matter in a stellar *corona*; this distribution is known as the *differential emission measure* (DEM). The corona is the outermost layer of a stellar atmosphere and contains very low-density (about 10^9 particles/cm^{-3}) and very hot ($> 10^6$ K) heavily ionized gas. Figure 11 illustrates the solar corona by imaging the Sun in three wavelengths of light[19]. The first panel is an optical image taken on March 29, 2001 of a portion of the Sun and illustrates the largest sunspot group to appear in a decade; at its peak, this group was over ten times the size of Earth. The second and third panels illustrate an extreme ultraviolet image and an X-ray image of the same region of the Sun, respectively. Although in visible light the sunspots appear as dark areas against the bright surface of the Sun, they light up in the extreme ultraviolet. The X-ray image shows large loops of glowing plasma arching above the sunspot group. The reason that the images look so different is that they are actually revealing different layers of the Sun's atmosphere. The visible photons originate from the photosphere, the lowest and coolest layer at about 5000 degrees Kelvin, the extreme ultraviolet image reveals the chromosphere/transition region which is above the photosphere and hotter at 10–100 thousand degrees Kelvin. Finally, the X-rays originate from the solar corona that is even higher and is even hotter—at least a million degrees Kelvin.

---

[19]URL: http://antwrp.gsfc.nasa.gov/apod/ap010419.html

The X-ray image in Figure 11 illustrates the complex structure in the intensity of the X-ray emission across the solar corona. The structure of the emission is a tracer of temperature and density in the corona. (The visible image also reveals temperature structure; the sunspots are much cooler than their surroundings. X-ray images, however, are not useful for viewing the temperature structure of the relatively cool photosphere.) Figure 12 illustrates the relative abundance of matter in a region of the solar corona with high sunspot activity as a function of temperature; this is a plot of the solar DEM in an active region. Figure 13 is the same plot, but in a region with no sunspot activity. Notice that there is relatively less very hot matter in the quiet region of the solar corona.

Although impressive images of the solar corona are available from the Solar & Heliospheric Observatory (SOHO)[20], very little is known about the temperature structure of stellar coronae. The stars being very distant, their disks cannot be resolved with the resolution of even the best existing telescopes. We can, however, infer their structures indirectly by examining the temperature structure. There are clues in the emission lines of stellar X-ray spectra that can be unlocked using prior information obtained from detailed quantum mechanical computations and ground-based laboratory measurements. A stellar corona is made up of various ions which can be recognized in a spectrum by their identifying emission lines. If the corona is relatively hot, we expect the emission lines that correspond to more energetic quantum states to be relatively strong. Thus, the relative strength of the emission lines corresponding to a particular ion carries information as to the temperature of the source.

Model Formulation. Unlocking this information is equivalent to solving a Poisson image reconstruction problem of the type formulated in (6). Suppose a stellar corona were composed completely of a single ion at a particular temperature. Quantum physical calculations can predict exactly which emission lines are possible and the relative probability that an emitted photon appears in each of the possible lines; this is a multinomial probability vector. The same can be done if the relative abundance of the various ions is known by simply mixing the multinomial distributions with weights equal to the relative abundance of the ions. As a starting point in practice, we use the relative abundance of ions in the solar corona. Thus, for a given temperature and for a large set of emission lines, we have the probability that an emitted photon falls into each line. This probability vector can be computed for each value of temperature on a grid of possible coronal temperatures (equally spaced on the $\log_{10}$ scale). We compile these vectors into a $J \times M$ *emissivity matrix*, $\mathbf{P}$, where J is the number of possible emission lines, and M is the number of temperature values in the grid. (The emissivity matrix can be compiled using the Atomic Database (ATOMDB)[21] provided by the *Chandra* X-ray Center.)

Using data collected with *Chandra*'s HRC with a grating in place (LETGS), we can obtain independent Poisson counts from tens of thousands of lines. (These are actually counts in very narrow energy bins that often contain two or more lines. This, however, poses no conceptual problem. It simply means that the instrumentation is

---

[20] URL: http://sohowww.nascom.nasa.gov
[21] URL: http://cxc.harvard.edu/atomdb/ (Smith et al. 2001)

at a lower resolution than the theoretical calculations and we must sum sets of rows of the emissivity matrix.) Because of the continuum, each count is actually a mixture of a line count and continuum count; for the moment we ignore the continuum. The expected count for each of the lines are collected into a vector, $\boldsymbol{\mu}$. This is the expected ideal count, because we are not yet accounting for the instrument effect, the effective area of the instrument, absorption, or background contamination. We wish to compute the expected count corresponding to each of the temperature bins. Recall that the temperature is not constant across the corona; see Figure 11. We compile the expected counts in each temperature bin into the vector $\boldsymbol{\mu}_{\mathrm{T}}$; this is our representation of the DEM.

The ML estimate of $\boldsymbol{\mu}$ can be computed by solving

$$\boldsymbol{\mu} = \mathbf{P}\boldsymbol{\mu}_{\mathrm{T}}, \tag{13}$$

which is a simplified version of (6). In practice, the vector $\boldsymbol{\mu}$ has tens of thousand of components, while $\boldsymbol{\mu}_{\mathrm{T}}$ has fewer than 100. We might expect this to be a very simple inversion, since there are far more data points than unknown parameters, but the information in the line counts for the DEM is very sparse: $\mathbf{P}$ can be nearly singular and the ML estimate can be very poorly behaved. Bayesian methods and informative prior information for $\boldsymbol{\mu}_{\mathrm{T}}$ are needed.

Because very little is known about the shape of the DEM, we impose simple smoothing prior distributions. We again implement the multi-scale prior distribution described in Section 3.4, this time in one dimension; we replace the multinomial and Dirichlet distributions with binomial and beta distributions, respectively.

Future Work. The model that we have described thus far is a simplified version of what is required for useful application in practice. For example, we are ignoring the continuum in our representation of the expected ideal counts in (13). A full analysis of the continuum would include a continuum emissivity matrix that quantifies how the DEM is manifested in the distribution of the energy of continuum emission. We are also not fully accounting for the data generation mechanism; in order to account for instrument response, the effective area of the instrument, and background contamination $\boldsymbol{\mu}$ must be substituted into (6) yielding a nested convolution model. This illustrates the power of the modular/hierarchical approach—one can add another layer to the model fairly simply. This will be important when we incorporate the effect of uncertainties in the emissivity matrix, $\mathbf{P}$; this matrix is only known with limited precision and its components are measured with correlated error. Another important outstanding extension to the model is to parameterize the stellar ion abundances; our emissivity matrix assumes solar abundances.

Example: DEM Reconstruction of Capella. One of the challenging aspects of DEM reconstruction from a statistical point of view is the near singularity of the dispersion matrix, $\mathbf{P}$. To illustrate this point, we begin with a simulation study that uses our best approximation of the emissivity matrix, $\mathbf{P}$, but uses simulated data from a known DEM,

$\boldsymbol{\mu}_T$. The simulation is based on model (13); it does not account for the continuum or the processes described in (6). The emissivity matrix (obtained from ATOMDB version 1.3) has 33,795 emission lines ranging 1.15 to 2715 Ångstrøms (Å) and 50 temperature bins; as is typical with grating data, we parameterize energy in terms of its reciprocal, wavelength, using Angstroms for the unit of length. For the simulation, we select 8,905 emission lines by taking the subset ranging from 5 to 25Å; the 8,905 emission lines are contained in 1,600 bins of equal width in units of wavelength (0.0125 Å). For convenience in the context of the hierarchical binomial splitting in the multiscale smoothing model, we expanded the partition of temperature bins into $64 = 2^6$ bins using linear extrapolation.

Figure 14 shows the results from four simulations corresponding to four known DEMs; the values of the DEM, $\boldsymbol{\mu}_T$, used to generate each of the datasets are represented by dotted lines. The MAP estimate of the DEM is illustrated using each of three multi-scale prior distribution for each of the four data sets. The prior distributions are specified in terms of the split probabilities at each level of resolution; we use the same conjugate beta(α, α) prior distribution at each level. The three prior distributions correspond to: (1) flat prior distributions with $\alpha = 1$, (2) moderate shrinking prior distributions with $\alpha = 2$, and (3) strongly shrinking prior distributions with $\alpha = 3$. The EM algorithms used to compute the MAP estimates were started with a flat DEM and run until convergence, as measured by the increase in the log of the posterior density evaluated at two consecutive iterates relative to the value evaluated at the first of the two iterates; convergence was called when this quantity was less than 10^{-5}. The results appear in Figure 14 and show that stronger prior distributions help significantly in the reconstruction of the smooth underlying DEM. In these simulations, the data are such that we only expect reliable results for $5.5 < \log(T) < 7.5$. Thus, we focus on this range when evaluating the reconstructions. We emphasize that even with 26,005 energy bins, the ML estimates are not satisfactory and the prior suspicion of a smooth DEM is helpful in achieving acceptable results.

The same models can be fit via MCMC to compute the posterior mean of $\boldsymbol{\mu}_T$ and to access its posterior variability. This time we use a larger range of emission lines; we select 26,005 emission lines by taking the subset ranging from 3 to 180Å; this corresponds to 14,160 bins each of width 0.0125Å. To accommodate the larger dataset we use a stronger prior distribution—the conjugate beta prior distribution with $\alpha = 10$. Figure 15 illustrates the result for simulations 3 and 4. Starting from a flat DEM for each dataset we ran the Markov chain for 300,000 iterations, saving every 100th draw, for a total sample size of 3000 draws, of which the final 2000 were used for Monte Carlo integration. Figure 15 compares $\boldsymbol{\mu}_T$ with its posterior mean and componentwise 95% posterior intervals for both data sets. Again in the range $5.5 < \log(T) < 7.5$, the reconstructions are very good.

Finally we attempt to estimate the DEM of the X-ray bright star, Capella. The raw spectrum of Capella collected with the HRT using the LETGS diffraction grating appears in Figure 16, which illustrates the many emission lines that compose the spectrum. We select 7,741 emission lines by taking the subset ranging from 6 to 20Å; this corresponds to 1,120 bins of width 0.0125Å. In our reconstruction, we include a

continuum term that is assumed known up to a normalizing constant and account for the processes described in (6); we include a term for background contamination that is also assumed known up to a normalizing constant. We compute the MAP estimate of the DEM using the EM algorithm, again starting from a flat DEM, and using the same convergence criterion as in the simulation study; α was set to three in the beta prior distributions. The model was also fit via MCMC. The posterior distribution of $\boldsymbol{\mu}_{\mathrm{T}}$ is summarized by its posterior mean and componentwise 95% posterior intervals in Figure 17. An important advantage of our model-based approach is a principled method for assigning errors to the fitted DEM. No previous attempt to reconstruct a stellar DEM has addressed uncertainty in the fitted DEM and the atomic data—indeed, this is the first method to perform global fits to the spectrum that uses all the pieces of available information in terms of the individual line emissivities, that allows for conceptually easy enhancements to include errors in atomic databases, and that allows for the inclusion of prior information (see also Kashyap and Drake 1998).

4   Discussion

In his seminal 1979 paper introducing the likelihood ratio to astrophysicists, Webster Cash began by remarking,

> As high-energy astronomy matures, experiments are producing data of higher quality in order to solve problems of greater sophistication. With the advent of the *HEAO* satellites, the quality of X-ray astronomy data is being increased again, and it is important that the procedures used to analyze the data be sufficiently sophisticated to make the best possible use of the results.

Cash's statement applies as well today as it did twenty-five years ago. The sophisticated equipment on board *the Chandra X-ray Observatory* offers a giant leap forward in the quality of spectral, spatial, and temporal data available to high-energy astrophysicists. Unlocking the information in this data, however, requires equally sophisticated statistical tools and methods. In the realm of sophisticated space observatories, *Chandra* is just the tip of the iceberg. Planned or proposed missions for high-energy detectors include *Constellation-X, GLAST, XMM, and Generation-X.* Although the lessons learned from *Chandra* will surely inform the design of these instruments and, one day, the analysis of their data, new challenges will certainly be encountered. High-energy astrophysics and other areas in astronomy offer many statistical challenges and fertile ground for statisticians and scientists interested in developing statistical tools and methods.

References

Babu, G. J. and Feigelson, E. D. (eds.) (1992). *Statistical Challenges in Modern Astronomy*. New York: Springer. 191

— (1997). *Statistical Challenges in Modern Astronomy II*. New York: Springer. 191, 203

Bechtold, J., Siemiginowska, A., Aldcroft, . T. L., Elvis, M., and Dobrzycki, A. (2001). "Chandra Detection of X-Ray Absorption Associated with a Damped Lyα System." *The Astrophysical Journal*, 562: 133–138. 196

Bijaoui, A. B. (1971a). "Étude électronographique du centre de l'amas globulaire M 13. I." *Astronomy and Astrophysics*, 13: 225–231. 190

— (1971b). "Étude électronographique du centre de l'amas globulaire M 13. II." *Astronomy and Astrophysics*, 13: 232–236. 190

Cash, W. (1979). "Parameter estimation in astronomy through application of the likelihood ratio." *The Astrophysical Journal*, 228: 939–947. 219

Connors, A. (2003). "Power from Understanding the Shape of Measurement: Progress in Bayesian Inference for Astrophysics." In *Statistical Challenges in Modern Astronomy III* (Editors: E. Feigelson and G. Babu), 19–39. New York: Springer–Verlag. 190

Davis, J. E. (2001). "Event Pileup in Charged-coupled Devices." *The Astrophysical Journal*, 562: 575–582. 200

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). "Maximum Likelihood from Incomplete Data Via the EM Algorithm (with Discussion)." *Journal of the Royal Statistical Society, Series B, Methodological*, 39: 1–37. 191

Dupree, A. K., Brickhouse, N. S., Doschek, G. A., Green, J. C., and Raymond, J. C. (1993). "The Extreme Ultraviolet Spectrum of Alpha Aurigae (Capella)." *The Astrophysical Journal-Letters*, 418: L41–L44. 235

Esch, D. N. (2003). "Extensions and Applications of Three Statistical Models." Ph.D. thesis, Harvard University, Dept. of Statistics. 213

Esch, D. N., Connors, A., Karovska, M., and van Dyk, D. A. (2004). "An Image Reconstruction Technique with Error Estimates." *The Astrophysical Journal*, 610: 1213–1227. 213

Fabian, A. C., Sanders, J. S., Ettori, S., Taylor, G. B., Allen, S. W., Crawford, C. S., Iwasawa, K., Johnstone, R. M., and Ogle, P. M. (2000). "Chandra Imaging of the Complex X-Ray Core of the Perseus Cluster." *Monthly Notices of the Royal Astronomical Society*, 318: L65–L68. 197

Feigelson, E. D. and Babu, G. J. (eds.) (2003). *Statistical Challenges in Modern Astronomy III*. New York: Springer. 191

Fessler, J. A. and Hero, A. O. (1994). "Space-Alternating Generalized Expectation-Maximization Algorithm." *IEEE Transactions on Signal Processing*, 42: 2664–2677. 207

Friedman, H. (1960). "Recent Experiments from Rockets and Satellites." *Astronomical Journal*, 65: 264 – 271. 193

Gelman, A., Meng, X.-L., and Stern, H. (1996). "Posterior Predictive Assessment of Model Fitness (with Discussion)." *Statistica Sinica*, 6: 733–807. 215

Gelman, A. and Rubin, D. B. (1992). "Inference from Iterative Simulations Using Multiple Sequences (with Discussion)." *Statistical Science*, 7: 457–472. 214

Giacconi, Gursky, Paolini, and Rossi (1962). "Evidence for X Rays from Sources Outside the Solar System." *Physical Review Letters*, 9: 439 – 443. 193

Gregory, P. C. and Loredo, T. J. (1992). "A new method for the detection of a periodic signal of unknown shape and period." *The Astrophysical Journal*, 398: 146–168. 202

Hans, C. M. and van Dyk, D. A. (2003). "Accounting for Absorption Lines in High Energy Spectra." In *Statistical Challenges in Modern Astronomy III* (Editors: E. Feigelson and G. Babu), 429–430. New York: Springer–Verlag. 208

Hughes, J. P., Ghavamian, P., Rakowski, C. E., and Slane, P. O. (2003). "Iron-rich Ejecta in the Supernova Remnant DEM L71." *The Astrophysical Journal-Letters*, 582: L95–L99. 209

Kang, H., van Dyk, D. A., Yu, Y., Siemiginowska, A., Connors, A., and Kashyap, V. (2003). "New MCMC Methods to Address Pile-up in the Chandra X-ray Observatory." In *Statistical Challenges in Modern Astronomy III* (Editors: E. Feigelson and G. Babu), 449–450. New York: Springer–Verlag. 200

Kashyap, V. and Drake, J. J. (1998). "Markov-Chain Monte Carlo Reconstruction of Emission Measure Distributions: Application to Solar Extreme-Ultraviolet Spectra." *The Astrophysical Journal*, 503: 450–466. 219

Kashyap, V. L., Drake, J. J., Güdel, M., and Audard, M. (2002). "Flare Heating in Stellar Coronae." *The Astrophysical Journal*, 580: 1118–1132. 202

Kolaczyk, E. D. (1999). "Bayesian Multi-Scale Models for Poisson Processes." *Journal of the American Statistical Association*, 94: 920–933. 211

Kolaczyk, E. D. and Dixon, D. D. (2000). "Nonparametric Estimation of Intensity Maps Using Haar Wavelets and Poisson Noise Characteristics." *The Astrophysical Journal*, 534: 490–505. 202, 203

Kolaczyk, E. D. and Nowak, R. D. (1999). "Deconvolution of Poisson-Limited Data Using a Bayesian Multi-Scale Model." *Bulletin of the American Astronomical Society*, 31: 734–+. 203

Komossa, S., Burwitz, V., Hasinger, G., Predehl, P., Kaastra, J. S., and Ikebe, Y. (2003). "Discovery of a Binary Active Galactic Nucleus in the Ultraluminous Infrared Galaxy NGC 6240 Using Chandra." *The Astrophysical Journal-Letters*, 582: L15–L19. 210

Lampton, M., Margon, B., and Bowyer, S. (1976). "Parameter Estimation in X-ray Astronomy." *The Astrophysical Journal*, 208: 177–190. 201

Lange, K. and Carson, R. (1984). "EM Reconstruction Algorithms for Emission and Transmission Tomography." *Journal of Computer Assisted Tomography*, 8: 306–316. 207

Leahy, D. A., Elsner, R. F., and Weisskopf, M. C. (1983). "On Searches for Periodic Pulsed Emission - The Rayleigh Test Compared to Epoch Folding." *The Astrophysical Journal*, 272: 256–258. 202

Lochner, J. C., Swank, J. H., and Szymkowiak, A. E. (1989). "A Search for a Dynamical Attractor in Cygnus X-1." *The Astrophysical Journal*, 337: 823–831. 202

Loredo, T. J. (1992). "Promise of Bayesian Inference for Astrophysics." In *Statistical Challenges in Modern Astronomy* (Editors: E. Feigelson and G. Babu), 275–306. New York: Springer-Verlag. 200

Loredo, T. J. and Chernoff, D. F. (2003). "Bayesian Adaptive Exploration." In *Statistical Challenges in Modern Astronomy III* (Editors: E. Feigelson and G. Babu), 57–70. New York: Springer–Verlag. 191

Lucy, L. B. (1974). "An Iterative Technique for the Rectification of Observed Distributions." *The Astronomical Journal*, 79: 745–754. 191, 202

Lyons, L., Mount, R., and Reitmeyer, R. (eds.) (2004). *Proceedings of the Conference on Statistical Problems in Particle Physics, Astrophysics, and Cosmology*. Menlo Park, CA: SLAC Technical Publications. 192

McHardy, I. and Czerny, B. (1987). "Fractal X-ray Time Variability and Spectral Invariance of the Seyfert Galaxy NGC5506." *Nature*, 325: 696–698. 202

Meng, X.-L. and van Dyk, D. A. (1997). "The EM Algorithm – an Old Folk Song Sung to a Fast New Tune (with Discussion)." *Journal of the Royal Statistical Society, Series B, Methodological*, 59: 511–567. 191, 207

Meredith, D. C., Ryan, J. M., and Young, C. A. (1995). "Searching for Multifractal Scaling in Gamma-Ray Burst Time Series." *Astrophysics and Space Science*, 231: 111–114. 202

Neugebauer, O. (1951). *The Exact Sciences in Antiquity*. Copenhagen: Ejnar Munksgaard. 190

Nichol, R. C., Chong, S., Connolly, A. J., Davies, S., C., G., Hopkins, A. M., Miller, C. J., Moore, A. W., Pelleg, D., Richards, G. T., Schneider, J., Szapudi, I., and Wasserman, L. (2003). "Computational AstroStatistics: Fast and Efficient Tools for

Analysing Huge Astronomical Data Sources." In *Statistical Challenges in Modern Astronomy III* (Editors: E. Feigelson and G. Babu), 265–278. New York: Springer–Verlag. 191

Nowak, R. D. and Kolaczyk, E. D. (2000). "A Bayesian Multiscale Framework for Poisson Inverse Problems." *IEEE Transactions on Information Theory*, 46: 1811–1825. 203, 211, 213

Noyes, R. (1982). *The Sun, Our Star*. Cambridge, MA: Harvard University Press. 193

Park, T., van Dyk, D. A., Kashyap, V., and Zezas, A. (2004). "Computing Hardness Ratios with Poissonian Error." *Manuscript in Progress.* 205

Perdang, J. (1981). "The Solar Power Spectrum - A Fractal Set." *Astrophysics and Space Science*, 74: 149–156. 202

Plackett, R. (1958). "Studies in the history of probability and statistics. VII. The Principle of the Arithmetic Mean." *Biometrika*, 45: 130–135. 190

Protassov, R., van Dyk, D. A., Connors, A., Kashyap, V., and Siemiginowska, A. (2002). "Statistics: Handle with Care – Detecting Multiple Model Components with the Likelihood Ratio Test." *The Astrophysical Journal*, 571: 545–559. 202

Richardson, W. H. (1972). "Bayesian-Based Iterative Method of Image Restoration." *Journal of the Optical Society of America*, 62: 55–59. 191, 202

Scargle, J. D. (1998). "Studies in Astronomical Time Series Analysis. V. Bayesian Blocks, a New Method to Analyze Structure in Photon Counting Data." *The Astrophysical Journal*, 504: 405–418. 203

Shepp, L. A. and Vardi, Y. (1982). "Maximum Likelihood reconstruction for Emission Tomography." *IEEE Transactions on Image Processing*, 2: 113–122. 207

Siemiginowska, A., Bechtold, J., Aldcroft, T. L., Elvis, M., Harris, D. E., and Dobrzycki, A. (2002). "Chandra Discovery of a 300 Kiloparsec X-Ray Jet in the Gigahertz-peaked Spectrum Quasar PKS 1127-145." *The Astrophysical Journal*, 570: 543–556. 196

Siemiginowska, A., Elivis, M., Alanna, C., Freeman, P., Kashyap, V., and Feigelson, E. (1997). "AXAF Data Analysis Challanges." In *Statistical Challenges in Modern Astronomy II* (Editors: E. Feigelson and G. Babu), 241–253. New York: Springer-Verlag. 201

Slezak, E., Bijaoui, A., and Mars, G. (1990). "Identification of Structures from Galaxy Counts - Use of the Wavelet Transform." *Astronomy and Astrophysics*, 227: 301–316. 202

Smith, R., Brickhouse, N., Liedahl, D., and Raymond, J. (2001). "Spectroscopic Challenges of Photoionized Plasmas." In *ASP Conference Series* (Editors: Gary Ferland and Daniel Wolf Savin), volume 247, 159. San Francisco: Astronomical Society of the Pacific. 216

Sourlas, E., van Dyk, D. A., Kashyap, V., Drake, J. J., and Pease, D. O. (2004). "Bayesian Spectral Analysis of Metal Abundance Deficient Stars." *Manuscript in Progress.* 208

Sourlas, N., van Dyk, D. A., Kashyap, V., Drake, J., and Pease, D. (2003). "Bayesian Spectral Analysis of "MAD" Stars." In *Statistical Challenges in Modern Astronomy III* (Editors: E. Feigelson and G. Babu), 489–490. New York: Springer–Verlag. 208

Stigler, S. M. (1986). *The History of Statistics: The Measurement of Uncertainty Before 1900*. Cambridge, Massachusetts: Harvard University Press. 190

Strauss, M. A. (2003). "Statistical and Astronomical Challenges in the Sloan Digital Sky Survey." In *Statistical Challenges in Modern Astronomy III* (Editors: E. Feigelson and G. Babu), 113–126. New York: Springer–Verlag. 191

Tucker, W. and Tucker, K. (2001). *Revealing the Universe: The Making of the Chandra X-ray Observatory*. Cambridge, MA: Harvard University Press. 194

van der Klis, M. (1997). "Quantifying Rapid Variability in Accreting Compact Objects." In *Statistical Challenges in Modern Astronomy II* (Editors: E. Feigelson and G. Babu), 321–331. New York: Springer-Verlag. 202

van Dyk, D. A. (2003). "Hierarchical Models, Data Augmentation, and Markov Chain Monte Carlo with discussion." In *Statistical Challenges in Modern Astronomy III* (Editors: E. Feigelson and G. Babu), 41–56. New York: Springer–Verlag. 200

van Dyk, D. A., Connors, A., Kashyap, V., and Siemiginowska, A. (2001). "Analysis of Energy Spectra with Low Photon Counts via Bayesian Posterior Simulation." *The Astrophysical Journal*, 548: 224–243. 201, 208

van Dyk, D. A. and Hans, C. M. (2002). "Accounting for Absorption Lines in Images Obtained with the Chandra X-ray Observatory." In *Spatial Cluster Modelling* (Editors: D. Denison and A. Lawson), 175–198. London: CRC Press. 208, 213

van Dyk, D. A. and Kang, H. (2004). "Highly Structured Models for Spectral Analysis in High Energy Astrophysics." *Statistical Science*, 19: 275–293. 208

Vardi, Y., Shepp, L. A., and Kaufman, L. (1985). "A Statistical Model for Positron Emission Tomography." *Journal of the American Statistical Association*, 80: 8–19. 207

Willett, R. M., Nowak, R. D., and Kolaczyk, E. D. (2002). "Multiscale Analysis of Photon-Limited Astronomical Signals and Images." *American Astronomical Society Meeting*, 201: 0–+. 203

Young, C. A., Meredith, D. C., and Ryan, J. M. (1995). "A Compact Representation of Gamma-Ray Burst Time Series." *Astrophysics and Space Science*, 231: 119–122. 203

Acknowledgments

---

[22]URL: http://www.ics.uci.edu/~dvd/astrostat.html

original   ***EMC2* image**

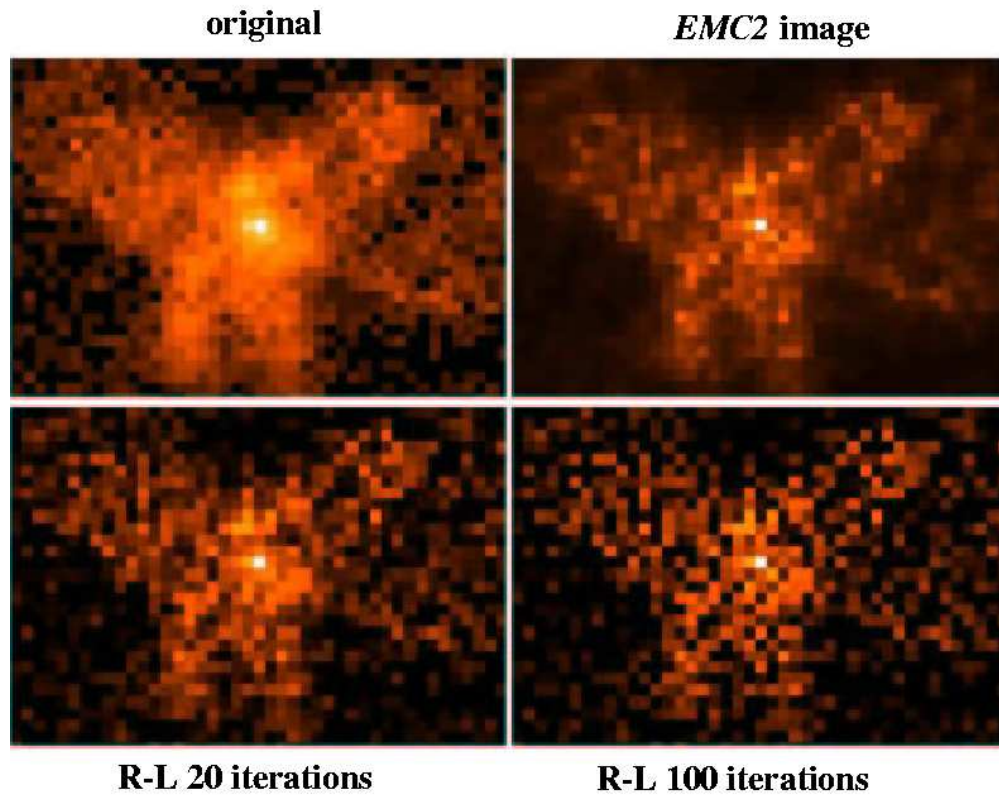**R-L 20 iterations**   **R-L 100 iterations**

Figure 8: Richardson-Lucy Reconstruction of NGC 6240. The upper left color image shows the original *Chandra* dataset. The upper right image illustrates the posterior mean reconstruction, showing two sources in the middle, surrounded by extended loops of hot gas. In the second row are the Richardson-Lucy reconstructions, stopped at 20 and 100 iterations, respectively. The Richardson-Lucy restorations under smooth the image, aggregating the intensity into E'clumps' in the image.
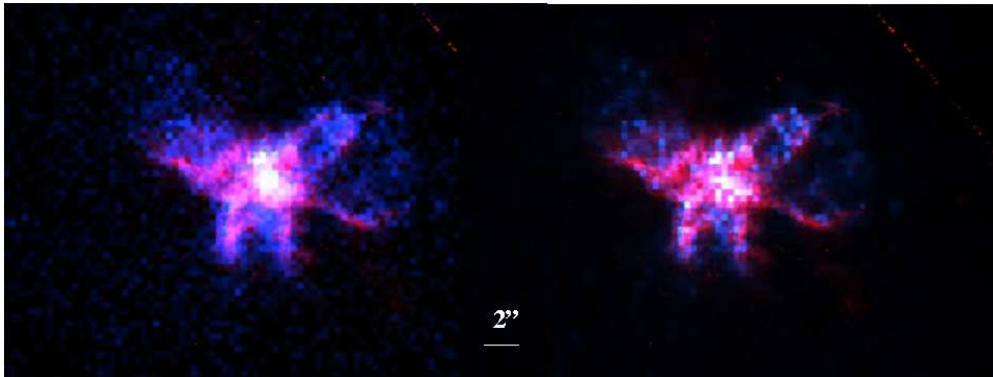
Figure 9: Comparing X-ray and Optical Images of NGC 6240. The figure shows a color overlay of the *Hubble* optical image first with the raw *Chandra* data and second with the posterior mean reconstruction. There is a strong correlation in the extended structures in the optical and X-ray bands, which provides additional verification for the posterior mean result.
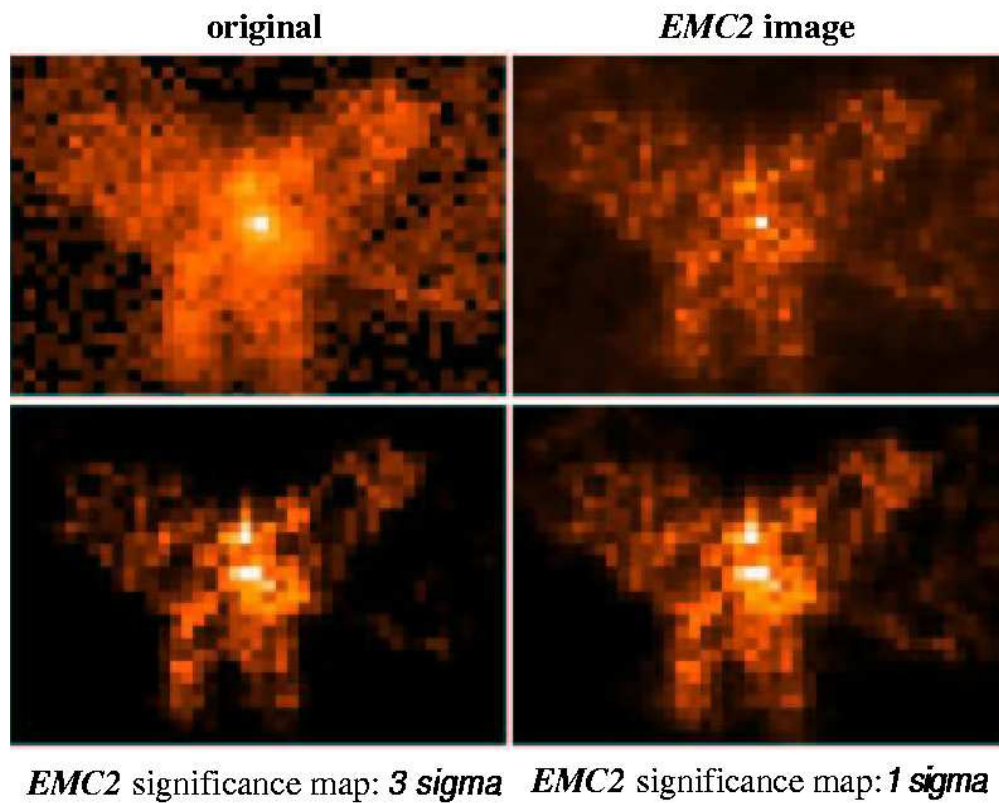
original   *EMC2* **image**

EMC2 significance map: **3 sigma**   *EMC2* significance map: **1 sigma**

Figure 10: Significance Maps for the Reconstruction of the X-ray Image of NGC 6240. The first row illustrates the raw data and the posterior mean reconstruction; the lower two images are the significance map, thresholded at three and one standard deviation, respectively. The significance maps indicate, for example, that the loop of hot gas in the upper right quadrant of the reconstructed image appears to be a feature in the source.
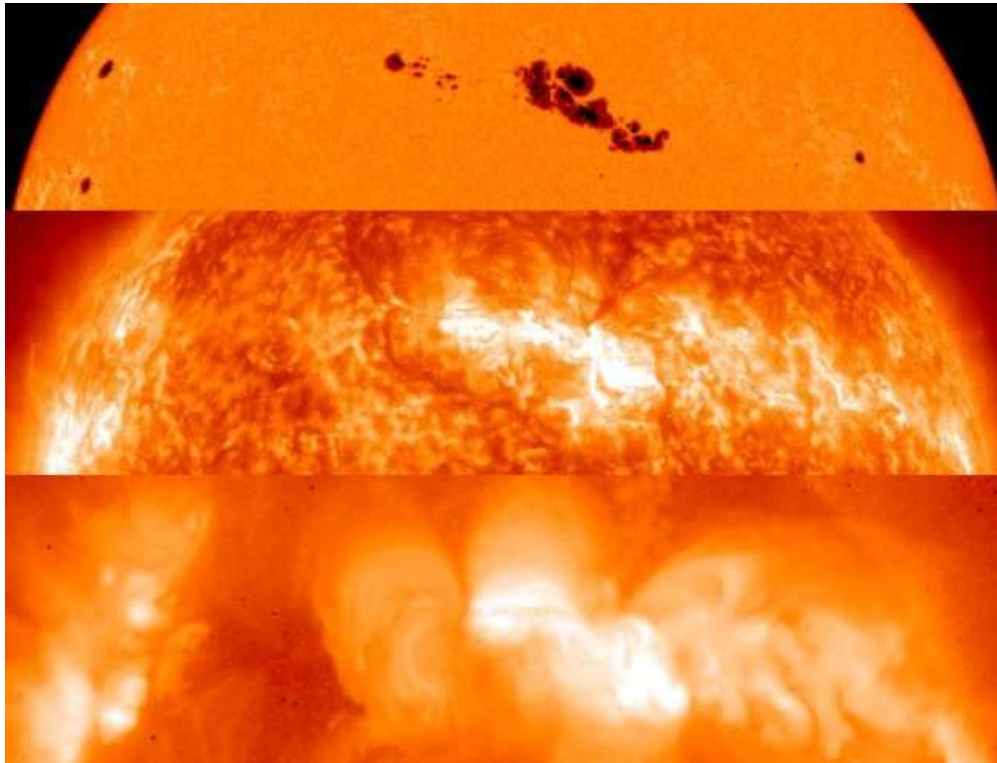
Figure 11: The Sun. The three images are, from top to bottom, optical, extreme ultraviolet, and X-ray images. The X-ray image illustrates the complex gradient of X-ray intensity and the corresponding temperature gradient across the solar corona. A stellar DEM is a representation of the distribution of the temperature of matter in a stellar corona. (Image Credit: SOHO - MDI / EIT Consortia, Yohkoh / SXT Project.)
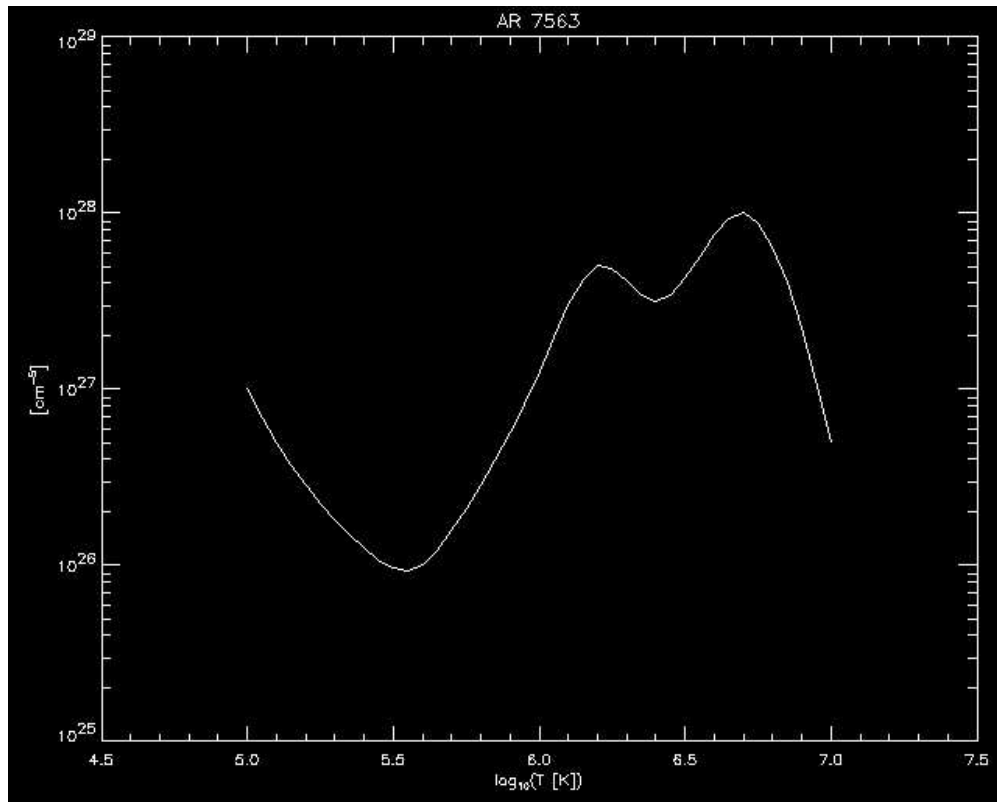
Figure 12: The Solar DEM in an Active Region. This is a plot of the relative abundance of matter in a region of the solar corona with high sunspot activity as a function of the log (base 10) of the temperature of the matter. The plot can be compared with that in Figure 13, which plots the solar DEM in a quiet region of the Sun.
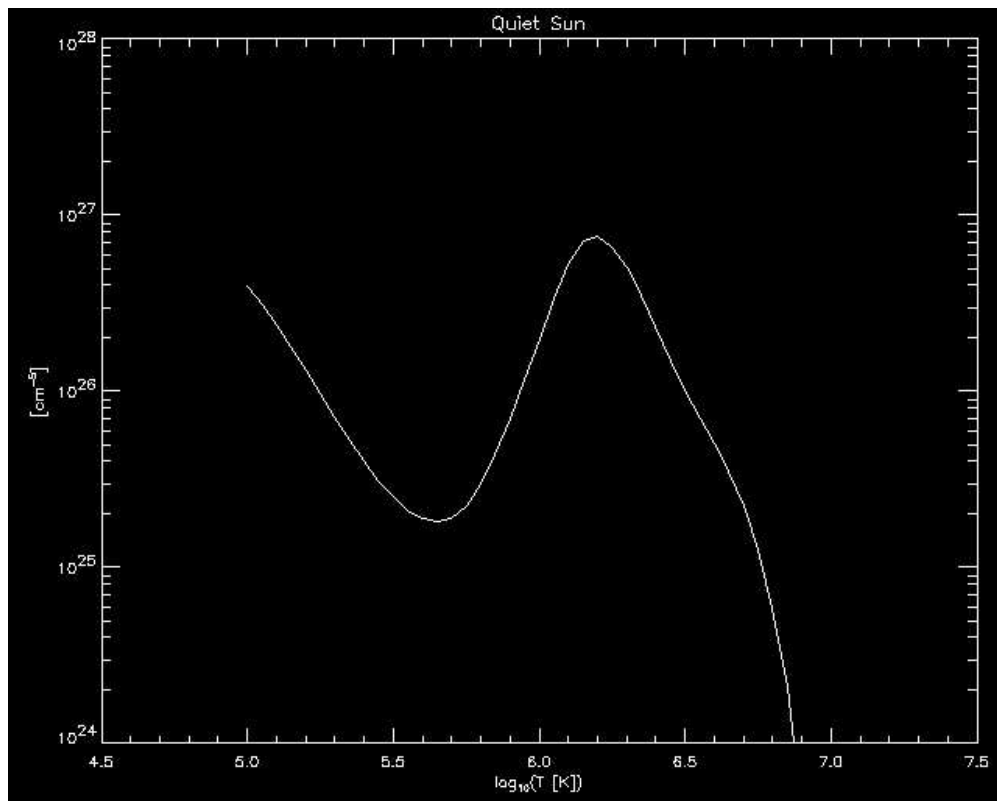
Figure 13: The Solar DEM in a Quiet Region. This is a plot of the relative abundance of matter in a region of the solar corona with no sunspot activity as a function of the log (base 10) of the temperature of the matter. The plot can be compared with that in Figure 12, which plots the solar DEM in an active region of the Sun.
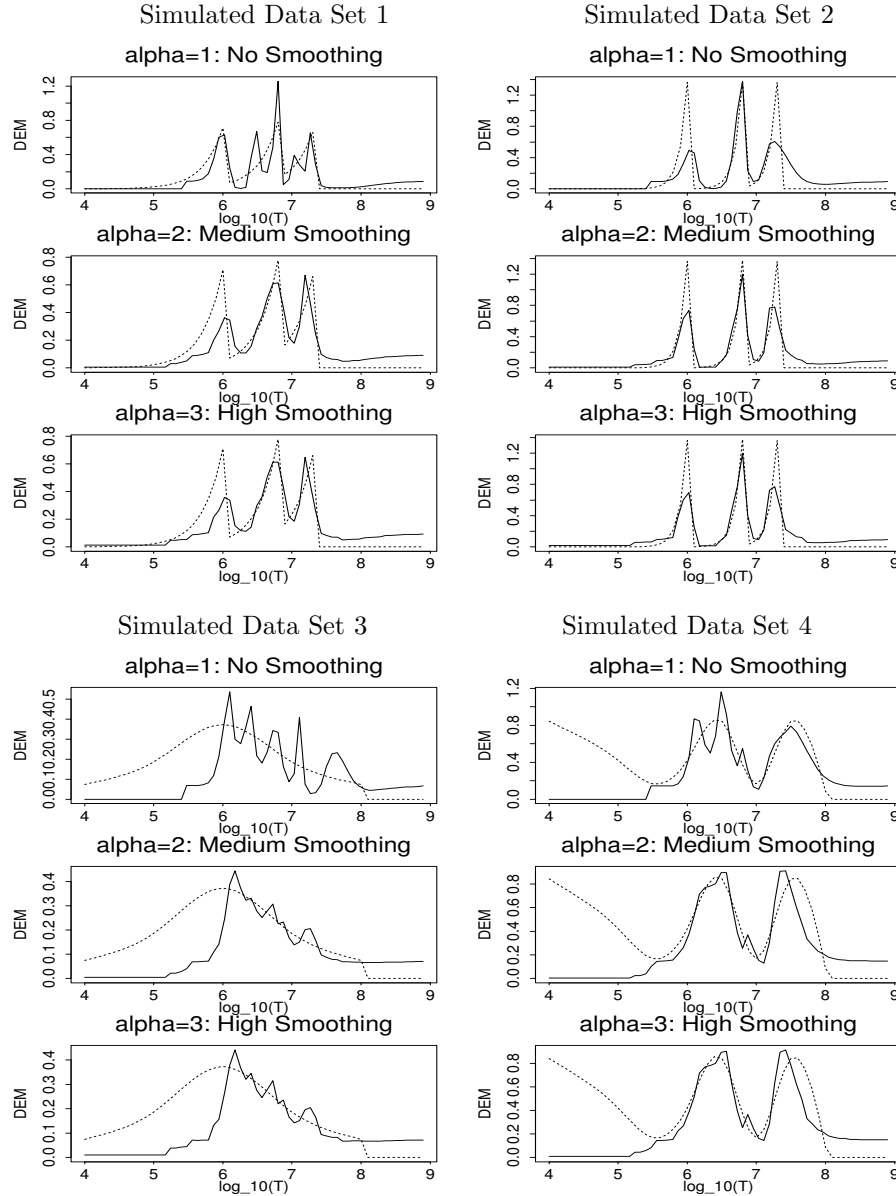
Simulated Data Set 1

alpha=1: No Smoothing

alpha=2: Medium Smoothing

alpha=3: High Smoothing

Simulated Data Set 2

alpha=1: No Smoothing

alpha=2: Medium Smoothing

alpha=3: High Smoothing

Simulated Data Set 3

alpha=1: No Smoothing

alpha=2: Medium Smoothing

alpha=3: High Smoothing

Simulated Data Set 4

alpha=1: No Smoothing

alpha=2: Medium Smoothing

alpha=3: High Smoothing

Figure 14: Fitted DEMs from the Simulation Study. Four DEMs were used to generate one data set each and appear as dotted lines in the plots. The MAP estimate of the DEM was computed using each of three prior distributions for each of the data sets; the MAP estimates are plotted as solid lines. The results improve with more heavily smoothing prior distributions. The fits are quite good in the reliable range of $5.5 < \log(T) < 7.5$ with the highly smoothing prior distributions. The vertical scales are in units of $10^1 7$.
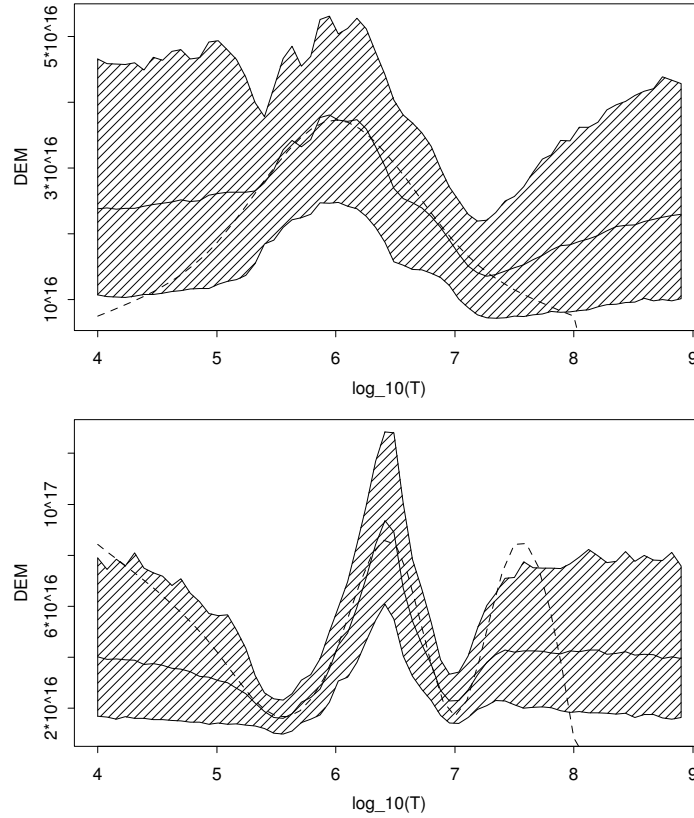
Figure 15: Error Bars for the Fitted DEMs from the Simulation Study. The dashed lines in the two plots represent the DEM that was used to generate the data; the solid line is the posterior mean of the DEM under a beta prior distribution with $\alpha = 10$; the shaded area represents componentwise 95% posterior intervals for the DEM. Note that for the reliable region of $5.5 < \log(T) < 7.5$, the DEM is very close to its posterior mean and is included in the posterior intervals.
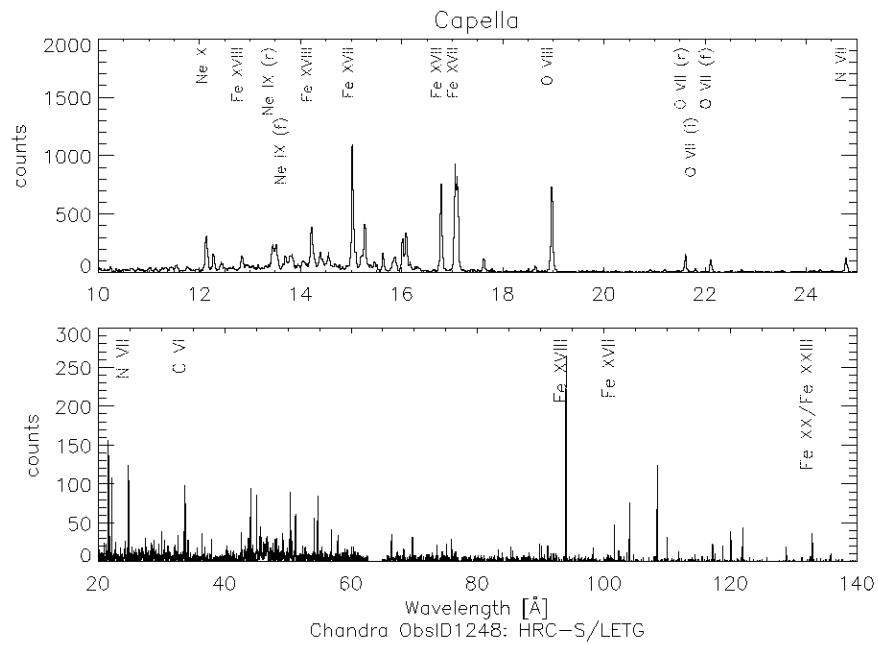
Figure 16: The Raw Spectrum of Capella (α Aur). This high resolution spectra was collected using *Chandra*'s HRC with the LETGS diffraction grating. The first panel magnifies the low-wavelength end of the spectrum. Notice the numerous emission lines that compose the spectrum; several important emission lines are labeled.
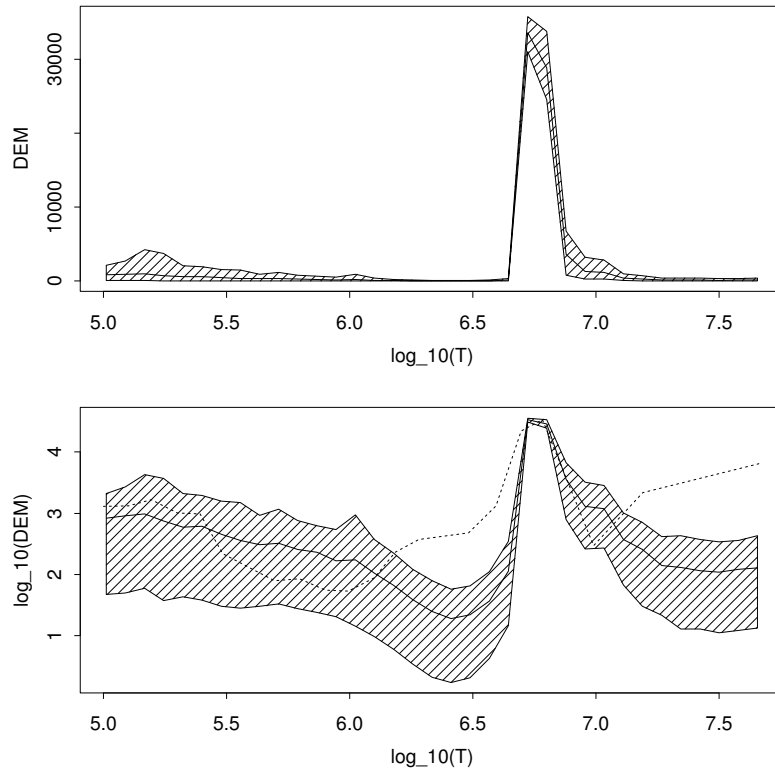
Figure 17: The Fitted DEM of Capella. The two plots represent the original scale and the log scale of the DEM, respectively. The solid line is the posterior mean of the DEM under a beta prior distribution with $\alpha = 3$, and the shaded area represents componentwise 95% posterior intervals for the DEM. The dotted line in the bottom plot is another DEM reconstruction of Capella by Dupree et al. (1993) using a different statistical method and a data set obtained at a different time epoch. Since the DEM varies over time, we expect the DEM corresponding to the two data sets to differ.