SPATIAL CLUSTER MODELLING

Edited by

Andrew B. Lawson

Department of Mathematical Sciences
University of Aberdeen
Aberdeen, UK

David G.T. Denison

Department of Mathematics
Imperial College of Science, Technology and Medicine
London, UK

CHAPMAN & HALL/CRC

A CRC Press Company

Boca Raton London New York Washington, D.C.

CHAPTER 10

Accounting for Absorption Lines in Images Obtained with the Chandra X-ray Observatory

D.A. van Dyk   *C.M. Hans*

10.1 Statistical Challenges of the Chandra X-ray Observatory

In recent years, new telescopes have dramatically improved the ability of astrophysicists to image X-ray sources. The Chandra X-ray Observatory, for example, was launched by the space shuttle *Columbia* in July 1999 and provides a high resolution tool that produces images at least thirty times sharper than any previous X-ray telescope. The X-rays themselves are produced by matter that is heated to millions of degrees, e.g., by high magnetic fields, extreme gravity, or explosive forces. Thus, the images provided by such high resolution instruments help astrophysicists to understand the hot and turbulent regions of the universe.

Unlocking the information in these images, however, requires subtle analysis. The detectors aboard Chandra collect data on each X-ray photon that arrives at the detector. Specifically, the (two-dimensional) sky coordinates, the energy, and the time of arrival of each photon are recorded. Because of instrumental constraints each of these four variables is discretized; the high resolution of Chandra means that this discretization is much finer than what was previously available. For example, one of the instruments aboard Chandra has 4096×4096 spatial pixels and 1024 energy bins. Because of the discrete nature of the data, it can be compiled into a four-way table of photon counts. We refer to this four dimensional data as the image; it is a moving 'colored' picture. (Because of the high energy of X-rays the 'colors' are not in the visible spectrum.) Spectral analysis models the one-way marginal table of the energy data; spatial analysis models the two-way marginal table of sky coordinates; and timing analysis models the one-way marginal table of arrival times. As we shall see, however, because of subtleties in the instrumentation and the data itself, spatial and spectral analysis cannot be fully separated and both are crucial for a full understanding of the image; in other settings, similar concerns arise for spatial and temporal analysis (see Chapter 12).

The data gathered by Chandra, although high-resolution, present a number of statistical challenges to the astronomer. For spatial data, the images are convolved with instrument characteristics which blur the image. For example, the point spread function characterizes the probability distribution of a photon's recorded pixel location relative to its actual sky coordinates. The situation is further complicated because the shape of the scatter distribution varies across the detector; it is symmetric and relatively tight in the center and becomes more asymmetric, irregular, and diffuse towards the edge. Moreover, the scatter distribution can vary with the energy of the incoming photon. Like the sky coordinates the energy of a photon is subject to "blurring"; there is a distribution of potential recorded energies given the actual energy of a particular photon. Thus, we have three dimensional blurring of the image. Given the sky coordinates and energy of a photon, there is a distribution of the recorded sky coordinates and recorded energy of the photon. Since there are, for example, 4096×4096 pixels on the detector and 1024 energy bins, the resulting blurring matrix can have over 2.9×10^{20} cells. 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. Also the blurring matrix can be taken to be 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 the shape of the energy cross sky coordinate scatter is constant in each of the resulting 64 cells. Such techniques aim at computational efficiency while hoping the compromise in precision is minor. A careful analysis of this trade off has yet to be tackled.

Another complication for image 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 inter-stellar media between the source and the detector. Effective area is an instrumental effect, but has a similar effect on the data—the probability that a photon is recorded by the detector depends on its energy. Because the spectrum can vary across the source, the rate of this stochastic censoring also varies and can distort the image. Again, this emphasizes that a careful analysis of the image must involve spectral analysis. In this paper, we take up the task of modeling photon absorption—with the understanding that it has direct implications for spatial analysis. In particular, we introduce new models that aim to account for absorption lines in spectra. An absorption line is a narrow range of energies where a relatively high proportion of photons are absorbed. Because these lines are caused by an abundance of a particular element near the surface of the source, it is possible for the intensity of an absorption line to vary across the source, thus distorting the image.

The data are also degraded by the presence of background counts, X-ray photons which arrive at the detector but do not correspond to the source of interest. In spectral analysis, a second observation that consists only of background counts is compared with the primary observation. The background observation is obtained by looking into an area of space near the source but which contains no apparent X-ray sources. After adjusting for exposure time, in some analyses the background observation is subtracted from the source observation, and the result is analyzed as if it were a source observation free of background. This procedure is clearly questionable, especially when the number of counts per bin is small. It often leads to the rather embarrassing problem of negative counts and can have 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 1993, van Dyk et al. 2001).

A final degradation of the data is known as pile up and poses a particularly challenging statistical problem. Pile-up occurs in X-ray detectors (generally charged coupled devices, i.e., CCDs) when two or more photons arrive at the same location in the detector (i.e., in an event detection island, which consists of several pixels) during the same time bin. Such coincident events are counted as a single higher energy event or 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. Moreover 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. Model based methods for handling pile up are discussed in Kang et al. (2002); see also Davis (2001).

We propose using model-based Bayesian methods to handle these complex imaging problems; other Bayesian approaches to image analysis appear, for example, in Chapters 6, 7, 8, 11 and 14 of this volume. Models can be designed to handle not only the complexity of the data collection process (e.g., blurring, effective area of the instrument, background contamination, and pile up) but also the complex spatial and spectral structures of the sources themselves. For example, as discussed in Section 10.2, we are interested in clustering photons into spatial and spectral features of the source. Because of their complexity, the models are in turn complex and require sophisticated computational methods for fitting. A Bayesian perspective is ideally suited to such highly structured models in terms of both inference and computation. For example, the high dimensional parameter space and numerous nuisance parameters highlight the attraction of Bayesian marginal posterior distributions of parameters or groups of parameters. Astrophysicists are often interested in testing for the presence of a particular model feature. Many such tests, such as testing for the presence of an additional point source in an image, correspond to testing whether a parameter is on the boundary of its space. It is well known that the likelihood ratio and related tests fail in this setting. However, appropriate Bayesian model fitting and model checking procedures are readily available (e.g. Protassov et al. 2001).

From a computational point of view, such tools as the EM algorithm (Dempster et al. 1977a), the Data Augmentation Algorithm (Tanner and Wong 1987), the Gibbs sampler (e.g. Gelfand and Smith 1990, Smith and Roberts 1993), and other Markov chain Monte Carlo methods are ideally suited to highly structured models of this sort; see van Dyk (2002). The modular structure of these algorithms matches the hierarchical structure of our models. For example, the Gibbs sampler samples one set of model parameters from their conditional posterior distribution given all other model parameters. This allows us to fit one component of the overall model at a time, conditional on the others. Thus, a complex model fitting task is divided into a sequence of much easier tasks. This modular structure also allows us to take advantage of well known algorithms that exist for fitting certain components of our model. For example, using the EM algorithm to handle a blurring matrix and background contamination in Poisson image analysis is a well known (and often rediscovered) technique (Richardson 1972, Lucy 1974, Shepp and Vardi 1982, Lange and Carson 1984, Fessler and Hero 1994, Meng and van Dyk 1997). Even though this standby image reconstruction algorithm is unable to handle the richness of our highly structured model, we utilize it and its stochastic generalization as a step in our mode finding and posterior sampling algorithms.

Detailing how we (or how we expect to) handle all of the modeling, computational, and inferential aspects of image analysis of Chandra data is well beyond the scope of this chapter. Instead, we outline some of our models to give the reader a flavor of our Bayesian analysis and highly structured models. In some cases, the details can be found in one of several references; in other cases, methods are still being developed. To give the reader a flavor of the statistical details that are involved, however, we go into some depth in our description of absorption lines.

The remainder of this chapter is organized into four sections. In Section 10.2 we outline our marginal spatial and spectral models, paying particular attention to the model based clustering. We discuss absorption lines in Section 10.3, describing the science behind them, the models and computational methods we propose, and a simulation study which illustrates some of the statistical difficulties involved. In Section 10.4, we incorporate absorption lines into the spectral model discussed in Section 10.2 illustrating our methods with a data set. Concluding remarks appear in Section 10.5.

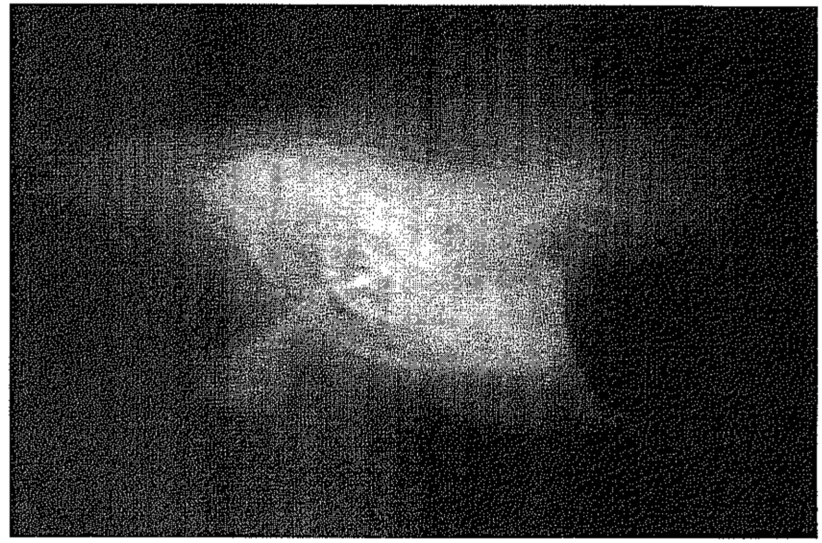

Figure 10.1 *This X-ray image of the Crab Nebula is one of the first images sent back by* Chandra. *The Crab Nebula is one of the youngest and most energetic of about 1000 known pulsars; in fact this supernova remnant produces energy at the rate of 100,000 suns. The image illustrates the extended irregular spatial structure that is typical of* Chandra *images (The image was adaptively smoothed; image credit: the National Aeronautics & Space Administration (NASA), the Chandra X-ray Center (CXC), and the Smithsonian Astrophysical Observatory (SAO)).*

10.2 Modeling the Image

10.2.1 Model-Based Spatial Analysis

We begin by modeling the true source counts in each pixel, $X = \{X_{i+}, i \in \mathcal{I}\}$, as independent Poisson random variables,

$$X_{i+} \sim Poisson(\Lambda_{i+}) \text{ for } i \in \mathcal{I}, \qquad (10.1)$$

where $\mathcal{I}$ is the set of pixels and the '+' in the subscript indicates that we are summing over the time and energy bins; in the remainder of Section 10.2.1 we suppress the '+'. Because the data are degraded by factors such as image blurring and background contamination as discussed in Section 10.1, X is not observed. Thus, we discuss both the constraints on Λ that we impose to represent structure in the image and how we model data distortion. We begin with Λ.

To motivate our parameterization of Λ we examine several Chandra images. Figure 10.1 illustrates an X-ray image of the Crab Nebula, the remnant of a supernova explosion that was observed on Earth in 1054 A.D; im-

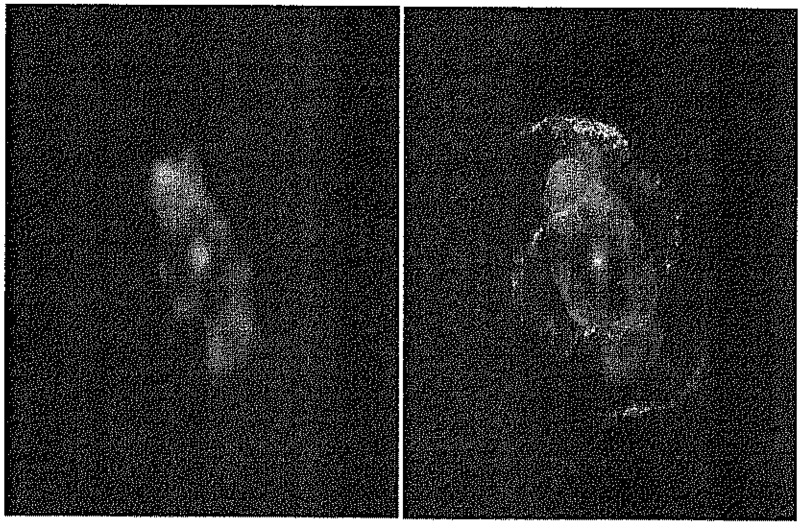

Figure 10.2 *X-ray image of the Cat's Eye Nebula (left) observed by* Chandra, *and a composite X-ray/optical observed by* Chandra *and the Hubble Space Telescope (HST) (right). The image on the left shows a bright central star, which corresponds to a high density cluster of photons. The composite image on the right illustrates the relative locations of hotter and cooler regions of the planetary nebula. (The X-ray image is adaptively smoothed. The color composite of optical and X-ray images was made by Zoltan G. Levay (Space Telescope Science Institute). The optical images were taken by J.P. Harrington and K.J. Borkowski (University of Maryland) with HST. Image credits: NASA, University of Illinois at Urbana-Champaign, Chu et al. (2001), and HST).*

age brightness corresponds to X-ray intensity. At the center of the nebula is a rapidly spinning pulsar that emits a pulse of photons thirty times per second. The Crab Nebula is a very bright X-ray source and illustrates the extended irregular structure that is typical of X-ray images. The structure in the extended source can sometimes be predicted from optical or radio images but often contains unique features. For example, the jet that extends towards the lower left from the center of the nebula was first observed by Chandra. Although model based methods are not required to identify some of the important structures in the Crab Nebula, such methods have broad application for analyzing weaker X-ray sources, understanding how the energy spectrum varies across a source, and identifying weak features in the source.

A second image appears in Figure 10.2 and illustrates X-ray (left panel) and optical (right panel) images of the Cat's Eye Nebula. A bright central

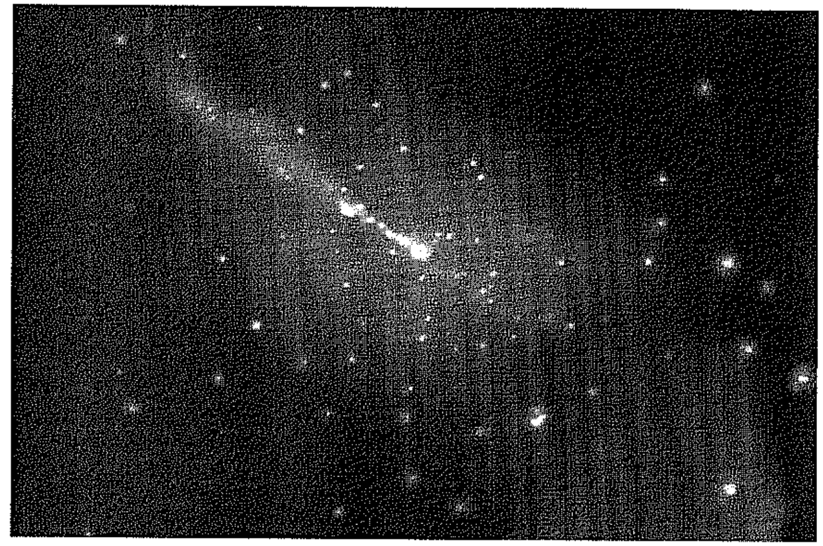

Figure 10.3 *Centaurus A, a nearby galaxy, as observed by* Chandra. *The several features of the galaxy – a super-massive black-hole, a jet emanating from the core and the point-like sources scattered about the image – have been clearly resolved for the first time due to the imaging resolution of* Chandra *and illustrate the variety of photon clusters that might be present in a* Chandra *image. (The image is adaptively smoothed and exposure corrected; image credit: NASA, SAO, and Kraft et al. (2001)).*

star is apparent in the center of the multi-million-degree nebula. Again the nebula exhibits extended irregular structure that is only partially predictable from the optical image. A final image appears in Figure 10.3, this one of a nearby elliptical galaxy, Centaurus A. The image shows a bright central source, which is believed to include a super-massive black hole, a jet emanating from the center, and numerous point-like X-ray sources all surrounded by a diffuse hot gas.

Figures 10.1-10.3 illustrate the wide variety of images that we hope to model. A useful model must both allow for the extended diffuse nebula with its irregular and unpredictable structure and include one or more highly concentrated X-ray emitter (i.e., point sources). An important objective is to cluster photons into these various sources. To accomplish this, we model X_i as a mixture of independent Poisson random variables. In particular,

$$\Lambda_i = \lambda_i^{\mathrm{ES}} + \sum_{k=1}^{K} \lambda_k^{\mathrm{PS}} p_{ik} \text{ for } i \in \mathcal{I}, \qquad (10.2)$$

where $\lambda_i^{\rm ES}$ represents the expected count due to smooth irregular extended source in pixel i, K is the number of point sources, $\lambda_k^{\rm PS}$ is the total expected count due to point source k, and p_{ik} is the proportion of point source k that falls in pixel i. The added point sources might be modeled as Gaussian distributions, in which case p_{ik} is a bivariate Gaussian integral. We can easily handle larger elliptical Gaussian point sources, or more irregular point source models. Information as to the location and spread of the point source is often forthcoming (e.g., from optical or radio observations) and thus informative prior distributions are often available for these parameters.

A Markov Random Field (MRF) model can be used to describe the irregular extended sources represented by $\lambda_i^{\rm ES}$ in (10.2) (and perhaps irregular added point sources). Our MRF models each pixel's log intensity as a Gaussian deviate centered at the mean of the log intensities of its neighboring pixels. In particular,

$$\log(\lambda_i^{\rm ES}) \sim \text{Normal}\left(\frac{1}{n_i}\sum_{j\in\partial(i)}\log(\lambda_j^{\rm ES}), \frac{v}{n_i}\right) \quad \text{for } i \in \mathcal{I}, \qquad (10.3)$$

where $\partial(i)$ is the set of pixels neighboring pixel i, n_i is the number of pixels in $\partial(i)$, and v is the user-specified between pixel variance. This specification allows for flexibility in defining each pixel's neighborhood, as well as specifying the variance of the Gaussian density. In principle, the variance parameter can be fit or can be allowed to vary across the detector field, so as to give the capacity for sharp edges. The variances themselves can then be modeled, perhaps via a common prior distribution with fitted hyperparameters, letting the data in effect determine the values of the smoothing parameter. An alternative to the MRF is the multiscale method proposed by Kolaczyk (1999), which aims to provide wavelet like models for Poisson data and has performed well in a variety of applications (Nowak and Kolaczyk 2000); the methods described in Chapters 6 and 7 may also be helpful in identifying regions of relative homogeneity in an extended source.

We turn now to models for the degraded observed data. As discussed in Section 10.1, the observed counts are blurred because of instrumental effects and contaminated by background counts. Thus, we modify (10.2) to model the observed count in pixel i, Y_i, as independent Poisson random variables,

$$Y_i \sim \mathit{Poisson}\left(\sum_{j\in\mathcal{I}} M_{ij}\Lambda_j + \theta_i^{\rm B}\right) \quad \text{for } i \in \mathcal{I}, \qquad (10.4)$$

where M_{ij} is the probability that a photon with actual sky coordinates corresponding to pixel j is recorded in pixel i, Λ_j is given in (10.2), and $\theta_i^{\rm B}$ the expected counts in pixel i attributed to background contamination. The blurring matrix (M_{ij}) and generally the background vector ($\theta_i^{\rm B}$) are assumed known from calibration. As discussed in Section 10.1, the model given in (10.4) is well known and has been the subject of much study. What is new here are the constraints we put on the source model and the spectral models to which we now turn.

10.2.2 Model-Based Spectral Analysis

In this section we briefly outline a class of spectral models; more details of the models and the algorithms used to fit them can be found in van Dyk et al. (2001), Protassov et al. (2001), and Sourlas et al. (2002). The spectral model aims to describe the distribution of the photon energies emanating from a particular source. Generally speaking, the distribution consists of several clusters of photons including a smooth *continuum* term and a number of added *emission lines*, which are narrow ranges of energy with more counts than would be expected from the continuum. The continuum is formed by radiation of heat from the hot center of stars to the cold space that surrounds them, a process known as thermal Bremsstrahlung. The emission lines are due to particular ions in the source and the abundance of the extra emission indicates the abundance of the ion in the source. Known spectral lines can tell us about the speed at which the source is moving by examining the Doppler shift of the line location.

Statistically, the models are designed to summarize the relative frequency of the energy of photons arriving at the detector and to separate the photons into clusters corresponding to the continuum and emission lines. Independent Poisson distributions are more appropriate to model the counts than the commonly used normal approximation (e.g., χ^2 fitting), especially for a high resolution detector. We parameterize the intensity in bin $j \in \mathcal{J} = \{1, \ldots, J\}$, as a mixture of the continuum term and K emission lines,

$$\Lambda_j = \delta_j \lambda^{\rm C}(\theta^{\rm C}, E_j) + \sum_{k=1}^{K} \lambda_k^{\rm E} p_{jk}, \quad j \in \mathcal{J}, \qquad (10.5)$$

where δ_j is the width of bin j, $\lambda^{\rm C}(\theta^{\rm C}, E_j)$ is the continuum term and is a function of the continuum parameter, $\theta^{\rm C}$, E_j is the mean energy in bin j, $\lambda_k^{\rm E}$ is the expected counts from emission line k, and p_{jk} is the proportion of emission line k that falls in bin j. (Here and in the rest of the chapter we refer to the spectral margin, although this is suppressed in the notation.) The smooth continuum term generally is parameterized according to physical models with several free parameters. Many of these models are amenable to standard statistical techniques, e.g., log linear models. Occasionally, a less parametric fit is desired, in which case a one dimensional Markov random field can be applied. The emission lines are generally parameterized as Gaussian or t distributions.

As discussed in Section 10.1, the counts are degraded by background contamination, instrument (i.e., the detector) response, and photon absorption. Instrument response is a characteristic of the detector that results in a blurring of the photons, i.e., a photon that arrives in bin j has probability M_{lj} of being detected in observed bin $l \in \mathcal{L} = \{1, \ldots, L\}$. The $L \times J$ matrix $\{M_{lj}\}$, which may not be square, is determined by calibration of the detector and presumed known. Because of these degradations, we model the observed counts as independent Poisson variables with parameters

$$\Xi_l = \sum_{j=1}^{J} M_{lj} \Lambda_j d_j \alpha(\theta^{\mathrm{A}}, E_j) + \theta_l^{\mathrm{B}}, \ \ l \in \mathcal{L}, \qquad (10.6)$$

where $\alpha(\theta^{\mathrm{A}}, E_j)$ is the probability that a photon of energy E_j is *not* absorbed, θ_l^{B} is the Poisson intensity of the background which is often known from calibration in space, and d_j is the known effective area for photons with energy E_j, with d_j normalized so that $\max_j d_j = 1$. In the absorption model $\alpha(\theta^{\mathrm{A}}, E_j)$ is typically taken to be a generalized linear model with θ^{A} denoting the model parameter. The absorbed photons form one or more clusters that are completely unobserved. An important special case involves so-called absorption lines, which are the topic of the remainder of the chapter.

10.3 Absorption Lines

10.3.1 Scientific Background

Absorption lines are downward spikes in the spectral continuum, which represent wavelengths where photons from the continuum have been absorbed by atoms of elements in the source. Because the specific energies at which photons are absorbed are unique to each element, examining absorption lines of a source can help to determine its composition. In order to motivate the physical models we employ and the statistical models we formulate, we begin with some scientific background. Photons are emitted from the hot center of a source (e.g., a star) in a continuous spectrum, and due to their high energy radiate toward the relatively colder region near the "surface" of the source (e.g., the corona of a star). In these cooler regions the continuum photons are in a higher energy state than their surroundings and thus are readily absorbed by surrounding atoms to keep the energy of the system in balance. When this occurs, the absorbing atom necessarily enters a higher, less stable, energy state. Any given atom, however, prefers a lower, more stable energy configuration, so with high likelihood the atom will shed the excess energy and return to its original state.

If the energy of the absorbed photon is eventually released by the absorbing atom, one may wonder why we observe a dip in the continuum. The answer is in part due to two processes called collisional deexcitation and radiative deexcitation. In collisional deexcitation the excited atom collides with another atom, and the "extra" energy due to the absorbed photon is

converted into kinetic energy. The previously excited atom is thus back to its original state, and we still observe a downward spike in the continuum. In radiative deexcitation the atom simply emits a photon of energy equal to the absorbed photon. However, the chance that we observe this photon with the detector is very small: there are many possible directions which the emitted photon can take, and the probability is minute that its path will be along our line of observation. Therefore, even though the excited atoms eventually return to lower energy states, we still observe an absorption line.

Absorption lines can be parameterized in terms of their location, width, and intensity. The location, μ, denotes the center of the absorption line and is of interest because the absorption wavelength (or equivalently, energy) indicates the absorbing element. Absorption lines generally have some positive width, σ^2, because they are "broadened" by several effects. One of these effects, Doppler broadening, occurs because the velocity at which a photon is moving when it is absorbed is a random variable, causing us to observe a Doppler "shifted" absorption energy. This causes some photons to appear to be absorbed at a slightly higher energy and others at a slightly lower energy, hence the broadening of absorption lines. The third parameter is the intensity parameter, λ^{A}. Astronomers often refer to the absorption mechanism that produces a line with small λ^{A} as being "optically thin," which means that the line does not absorb all of the continuum photons at its peak; an example appears in plot (a) of Figure 10.4. The intensity and width parameters together give an indication of the structure of the line, from which astronomers can learn about the relative concentration of the absorbing element in the source.

10.3.2 Statistical Models

In terms of the model specification in Section 10.2.2, formulating an absorption model requires us to specify the probability that a photon is *not* absorbed as a function of energy, i.e., $\alpha(\theta^{\mathrm{A}}, E_i)$ in (10.6). Generally there may be several types of absorption that act independently, e.g., absorption by the inter-stellar media over a broad energy range along with several relatively narrow absorption features. Thus, we can specify $\alpha(\theta^{\mathrm{A}}, E_i)$ as a product,

$$\alpha(\theta^{\mathrm{A}}, E_i) = \prod_{j=1}^{J^{\mathrm{A}}} \alpha_j(\theta_j^{\mathrm{A}}, E_i), \qquad (10.7)$$

where J^{A} is the number of independent absorption components.

For simplicity, we focus on the case when $J^{\mathrm{A}} = 1$ with the understanding that we can repeatedly apply the methods we describe to handle multiple components. This is a particularly useful exercise because of the modular structure of our model and fitting algorithms. For example, the Gibbs sampler fits the model one component at a time. Thus, a method for fitting the

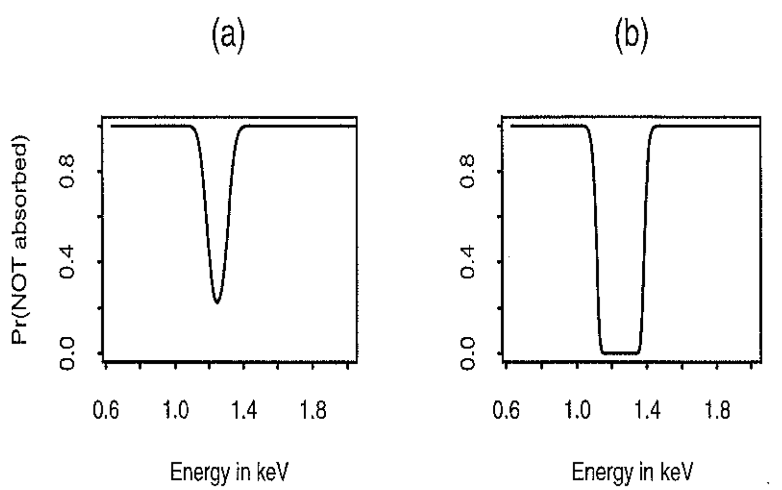

Figure 10.4 *Two examples of the physical flexibility of* $\alpha(E_j, \theta^A)$. *The vertical axis is the probability that a photon is* not *absorbed by the line, which is the probability that a photon is observed.*

spectral models described in Section 10.2.2 without absorption lines (see van Dyk et al. 2001) can be combined with the methods described here to fit the entire model, as we discuss below and illustrate in Section 10.4.

There are several models for absorption lines used by astrophysicists; some lines are modeled with flat edges on one side, and others are taken to have a symmetric or even Gaussian shape. Here, we limit our attention to a specific but important formulation of an absorption line; van Dyk et al. (2001) discuss models for absorption over a broad range of energy. Specifically, we consider the exponentiated Gaussian form described by Freeman et al. (1999),

$$\alpha(\theta^A, E_i) = \exp\left\{-\lambda^A \exp\left[\frac{-(E_i - \mu)^2}{2\sigma^2}\right]\right\}, \tag{10.8}$$

both because it is accepted as a useful description of physical phenomena and because, as discussed is Section 10.3.3, it is computationally tractable; see also Hans and van Dyk (2002). In this model, the absorption line parameter is $\theta^A = (\mu, \sigma^2, \lambda^A)$, where μ is the location parameter, σ^2 is the width parameter and λ^A is the intensity parameter. This parameterization is attractive for both physical and, as discussed below, statistical reasons. From a physical perspective, (10.8) provides a flexible way to describe the

absorption line. The inner exponential quantity gives the line a somewhat Gaussian shape, while the outer exponential and the intensity parameter control the strength of the line. Figure 10.4 illustrates this flexibility. Notice that in plot (a), the absorption line maintains a Gaussian shape but never reaches $\alpha(\theta^A, E_i) = 0$, where all the photons from the continuum would be absorbed. For this plot, $\mu = 1.25$, $\sigma^2 = 0.002$ and $\lambda^A = 1.5$. Plot (b) shows the outer limits of the line maintaining a Gaussian shape while all of the photons from the continuum are absorbed over the central bins; here, $\mu = 1.25$, $\sigma^2 = 0.002$ and $\lambda^A = 85$.

We specify the likelihood of the observed counts as

$$Y_l \overset{\text{indep.}}{\sim} Poisson(\Xi_l) \text{ for } l \in \mathcal{L}, \tag{10.9}$$

with Ξ_l given in (10.6). (Currently, it is common practice to account for absorption lines by modeling Y_l as independent Gaussian random variables with mean Ξ_l, for example, via χ^2 fitting. Such models are inappropriate for high-resolution, low-count detectors.) To complete the specification of the model, prior distributions must be assigned to the absorption line parameters. When prior information is available either from previous observations or other scientific knowledge, we use parameterized independent prior distributions on $(\sigma^2, \mu, \lambda^A)$; in particular we use scaled inverse χ^2, Gaussian, and gamma distributions respectively. Improper prior distributions should only be used with great care; there is a possibility of an improper posterior distribution when we consider the more general model described in Section 10.2.2.

Data Augmentation. We can augment model (10.9) to a larger, only partially observed sample space, which simplifies computation. The basic data augmentation is the idealized image that is undistorted by blurring, background contamination, or absorption,

$$X_i \overset{\text{indep.}}{\sim} Poisson(\Lambda_i) \text{ for } i \in \mathcal{I}, \tag{10.10}$$

with Λ_i given in (10.5). To account for absorption, we introduce an intermediate data augmentation, the idealized image *after absorption*,

$$Z_i | X_i, \theta^A \overset{\text{indep.}}{\sim} Binomial\left[X_i, \alpha(\theta^A, E_i)\right] \text{ for } i \in \mathcal{I}, \tag{10.11}$$

where $\alpha(\theta^A, E_i)$ is the probability that a photon in not absorbed and is given in (10.8). Combining (10.10) with (10.11) and marginalizing over X_i yields

$$Z_i | \theta^A \overset{\text{indep.}}{\sim} Poisson[\Lambda_i \alpha(\theta^A, E_i)] \text{ for } i \in \mathcal{I}. \tag{10.12}$$

Ordinarily we treat both $X = \{X_i, i \in \mathcal{I}\}$ and $Z = \{Z_i, i \in \mathcal{I}\}$ as missing data, along with a number of other quantities; see van Dyk et al. (2001). For the remainder of Section 10.3, however, we focus attention on absorption lines and treat Z as observed data and X as the unobserved idealized image. In particular, we act as if there were no blurring of photon energies or background contamination and assume $\Lambda = \{\Lambda_i, i \in \mathcal{I}\}$ is specified with no unknown parameters. All of these simplifications are to focus attention on absorption lines and will be relaxed in Section 10.4.

A Generalized Linear Model. The model specified in (10.11) can be formulated as a generalized linear model (McCullagh and Nelder 1989) using the link function, $\eta_i = -\log[-\log \alpha(\theta^{\mathrm{A}}, E_i)]$, where α is given in (10.8). In this case,

$$\eta_i = -\log\left(\Lambda^{\mathrm{A}}\right) + \frac{(E_i-\mu)^2}{2\sigma^2} = \left[-\log\left(\lambda^{\mathrm{A}}\right) + \frac{\mu^2}{2\sigma^2}\right] + \left[-\frac{\mu}{\sigma^2}\right] E_i + \left[\frac{1}{2\sigma^2}\right] E_i^2 \tag{10.13}$$

is linear in E_i and E_i^2. We can identify the coefficients of the generalized linear model with $\beta = (\beta_0, \beta_1, \beta_2)^{\mathrm{T}}$, via the invertible transformation

$$\beta_2 = \frac{1}{2\sigma^2}, \quad \beta_1 = -\frac{\mu}{\sigma^2}, \quad \beta_0 = -\log\left(\lambda^{\mathrm{A}}\right) + \frac{\mu^2}{2\sigma^2}. \tag{10.14}$$

10.3.3 Model Fitting

Our goal is to base inference on summaries of the posterior distribution,

$$p(\theta^{\mathrm{A}}|Z) = \int p(\theta^{\mathrm{A}}, X|Z)dX \propto \int p(Z|\theta^{\mathrm{A}}, X)p(X)p(\theta^{\mathrm{A}})dX, \tag{10.15}$$

where the factors under the final integral are given in (10.11), (10.10), and the prior distribution of θ^{A} respectively. Because of the complexity of (10.15) we resort to iterative methods to summarize the posterior distribution. Here we discuss both an EM algorithm that can be used to compute the posterior mode and MCMC methods that can be used to obtain a sample from the posterior distribution. Both of these methods are based on the data-augmentation scheme discussed in Section 10.3.2. In particular, both computational tools take advantage of the fact that the two conditional distributions, $p(X|Z, \theta^{\mathrm{A}})$ and $p(\theta^{\mathrm{A}}|Z, X)$, are well-known statistical models. Simple probability calculus shows that the first is

$$X_i|Z_i, \theta^{\mathrm{A}} \overset{\text{indep.}}{\sim} Z_i + \textit{Poisson}\left[\lambda_i(1-\alpha(\theta^{\mathrm{A}}, E_i))\right]. \tag{10.16}$$

The second is the posterior distribution under the generalized linear model described in Section 10.3.2.

EM Algorithm. The EM algorithm is a well-known iterative method for computing marginal posterior modes, such as the mode of $p(\theta^{\mathrm{A}}|Z)$ as expressed in (10.15). Starting with some starting value $\theta^{\mathrm{A}}_{(0)}$, EM proceeds by computing

$$\theta^{\mathrm{A}}_{(t+1)} = \mathrm{argmax}_{\theta^{\mathrm{A}}}\ E\left[\log p(\theta^{\mathrm{A}}|Z, X)\middle| Z, \theta^{\mathrm{A}}_{(t)}\right] \quad \text{for } t = 1, 2, \ldots \tag{10.17}$$

This procedure is guaranteed to increase the log posterior at each iteration and takes a particularly simple form in this case. The expectation in (10.17) can be written as

$$E\left[\sum_{i\in\mathcal{I}} X_i \log[\alpha(\theta^{\mathrm{A}}, E_i)] + (Z_i - X_i)\log[1-\alpha(\theta^{\mathrm{A}}, E_i)] + \log p(\theta^{\mathrm{A}})\middle| Z, \theta^{\mathrm{A}}_{(t)}\right]. \tag{10.18}$$

Since (10.18) is linear in X, we can simply replace the missing data by its expectation under model (10.16) and update the parameter by fitting the generalized linear model, e.g., via Newton-Raphson; see Hans (2001) for details.

MCMC Methods. The posterior distribution in (10.15) can be summarized via Monte Carlo by obtaining a sample from $p(\theta^{\mathrm{A}}, X|Z)$ and discarding the draws of X. We obtain a sample from the joint posterior distribution using the Gibbs sampler, an iterative algorithm that constructs a Markov chain which under mild regularity conditions converges to the joint posterior distribution (for convergence results see Roberts 1996).

We implement the following Gibbs sampler: given a starting value $\theta^{\mathrm{A}}_{(0)}$, we iterate,

Step1: Draw $X_{(t+1)}$ from $p(X|Z, \theta^{\mathrm{A}}_{(t)})$,

Step2: Draw $\theta^{\mathrm{A}}_{(t+1)}$ from $p(\theta^{\mathrm{A}}|Z, X_{(t+1)})$.

For sufficiently large T_0 we can consider $\{\theta^{\mathrm{A}}_{(t)}, X_{(t)}, t = T_0, \ldots, T\}$ to be a sample from (10.15) and summarize the posterior via Monte Carlo integration.

Step 1 in the Gibbs sampler can be easily accomplished according to (10.16). Although the probability distribution in Step 2 is not of a standard form, we can use the Metropolis-Hastings algorithm within each iteration of the Gibbs sampler to construct a Markov chain with stationary distribution as given in Step 2. We construct the Metropolis-Hastings jumping distribution using a wide-tailed approximation of the target density given in Step 2. The wide tails enable the sampler to jump across the parameter space, and if the approximation is good many proposals will be accepted. Our choice for the jumping density is a multivariate location-scale t-distribution with 4 degrees of freedom. We use the posterior mode of $p(\theta^{\mathrm{A}}|Z, X)$ (e.g, as computed in the M-step of EM) and the corresponding second derivative matrix to construct the center and scale of the jumping distribution respectively. Thus, the jumping distribution does not change within a single iteration of the Gibbs sampler. Because the Metropolis-Hastings algorithm is computationally quick once the jumping distribution has been computed, we iterate five times within each iteration of the Gibbs sampler, using the final draw as the draw for Step 2 of the Gibbs sampler. This strategy has negligible costs but potentially can improve the overall convergence properties of the Markov chain.

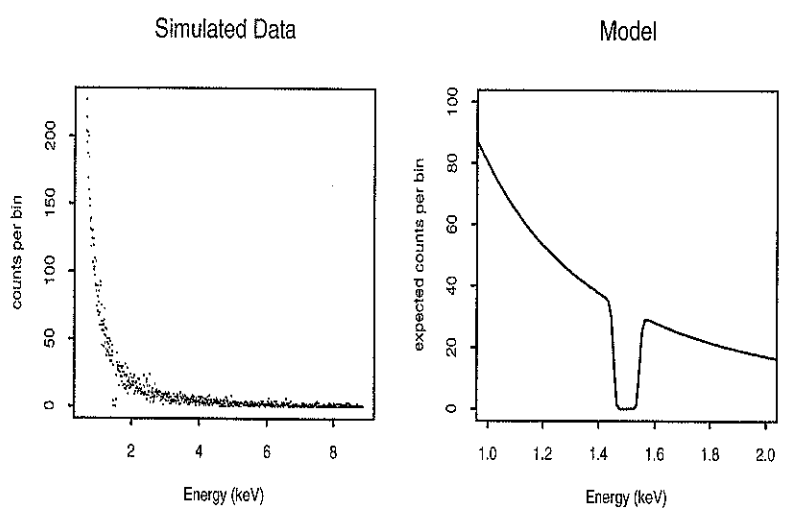

Figure 10.5 *Data simulated with $\lambda^A = 50$. The plot on the right shows the expected number of counts per bin in the area of the absorption line.*

10.3.4 A Simulation-Based Example

In this section we investigate how the characteristics of the fitted absorption line are affected by its actual parameters. We use a series of simulated data sets, generated according to (10.5) over the energy range [0.63 keV, 8.85 keV] using a bin width of 0.01 keV, giving a total of 822 bins. The continuum model was taken to be a power law, $\lambda^C(\theta^C, E) = \theta_1^C E^{-\theta_2^C}$, with the physically reasonable parameters $\theta_1^C = 80$ and $\theta_2^C = 2.23$. We simulated five datasets with the same line location, $\mu = 1.5$ keV, and width, $\sigma^2 = 0.0003$, but differing intensities, $\lambda^A = 1, 5, 10, 25$, and 50. The data set generated with $\lambda^A = 50$ is illustrated in Figure 10.5; the absorption line covers about ten bins. We used a flat prior distribution on θ^A in all analyses.

For each analysis we first ran the EM algorithm from five starting values dispersed about the parameter space to search for posterior modes. We then used these modes to select starting values for three MCMC samples. We begin with our analysis of the data set illustrated in Figure 10.5, generated with $\lambda^A = 50$. We fit the model to the data in two ways: first we allowed all three parameters in θ^A to be fit and second we fit only μ and λ^A, fixing σ^2 at 0.0003. The convergent values of EM for all five starting values for both fitting schemes appear in Table 10.1, which illustrates the

Run	Fit μ, σ^2, and λ^A				Fit only μ and λ^A		
	μ	$\sigma^2 \times 10^4$	λ^A	$\ell(\theta)$	μ	λ^A	$\ell(\theta)$
(i)	1.50	3.77	26.09	29936.3	1.30	0.23	29651.5
(ii)	1.50	3.77	26.09	29936.3	1.97	0.30	29650.6
(iii)	1.50	3.77	26.09	29936.3	1.50	52.09	29950.2
(iv)	1.50	3.77	26.09	29936.3	2.50	0.63	29654.1
(v)	1.50	3.77	26.09	29936.3	1.00	0.15	29650.5

Table 10.1 *Convergence of the EM algorithm for various starting values with $\lambda^A = 50$ for the underlying model $(\mu_{(0)}, \sigma^2_{(0)}, \lambda^A_{(0)})$ chosen as: (i) (1.25, 5×10^{-4},45); (ii) (2.0,2.0×10^{-3},80); (iii) (1.46,2.5×10^{-4},25); (iv) (2.5,8.0×10^{-2},100) and (v) (1.0,5.0×10^{-5},10). The left side of the table shows convergence for the model which fits all three parameters, and the right side shows convergence when σ^2 is fixed at 3×10^{-4}. The reported log-likelihood, $\ell(\theta)$, does not include the normalizing constant.*

multi-modal character of the posterior distribution. The natural Poisson variability of the photon counts can lead to random dips in the continuum which are not due to an absorption line, but create modes in the posterior distribution. Given our knowledge of the true model, the four small values of λ^A, and the value of the loglikelihood at each of these modes, it is evident that there is one major mode due to the absorption line and four minor modes that result from random fluctuations in the continuum. In practice these minor modes cause two difficulties. Computationally, MCMC samplers can get caught in a minor, relatively uninteresting mode. Thus, we generally recommend using the EM algorithm to first identify interesting modes and then construct MCMC starting values aiming to sample from these modes; see also Gelman et al. (1995). Secondly, it can be difficult to distinguish actual absorption lines from chance Poisson fluctuations in the continuum. The standard formulation of a formal hypothesis test involves a null value (i.e., no absorption line) that is on the boundary of the parameter space. Thus, the standard null distribution of the likelihood ratio test is inappropriate. In this case, we recommend using model checking techniques such as posterior predictive p-values to help distinguish between random fluctuations and weak lines; see e.g., Protassov et al. (2001).

A sample from the posterior distribution was generated by running each of three MCMC chains for 2500 iterations. We discarded the initial 500 draws of each chain and computed $\sqrt{\hat{R}}$ (Gelman and Rubin 1992) on the remaining draws to determine convergence to stationarity. Values of $\sqrt{\hat{R}}$ close to one signify that the several chains represent draws from the same distribution; values for three estimands (μ, σ^2 and $\log(\lambda^A)$) are reported in

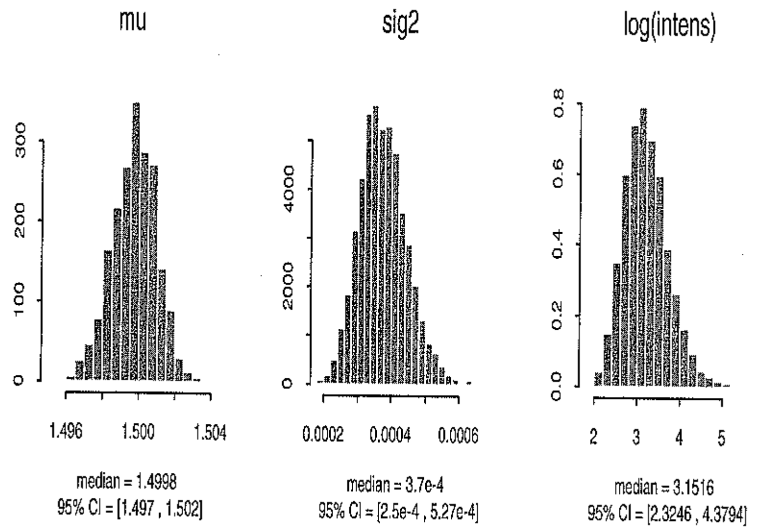

Figure 10.6 *6000 draws from the posterior distribution of θ^A for the data generated with $\lambda^A = 50$; the median and 95% credible intervals are reported. On the original scale, the posterior median for the intensity parameter is 23.374 with 95% CI = [10.223 , 79.794].*

Table 10.2. The $\sqrt{\hat{R}}$ statistic and visual inspection of the chains indicate good mixing. Figure 10.6 illustrates the marginal posterior distributions of μ, σ^2, and $\log(\lambda^A)$ and reports their median and 95% credible intervals; the true value of each is contained in its interval.

Finally, we repeat the above analyses for the datasets generated with $\lambda^A = 25, 10, 5,$ and 1; see Table 10.2. Figure 10.7 shows that 95% credible intervals cover the true parameter value in all cases. We also notice a negative association between the estimates of σ^2 and λ^A; when σ^2 is underestimated, λ^A is overestimated and vice versa. In either case, the expected absorption count is maintained.

10.4 Spectral Models with Absorption Lines

10.4.1 Combining Models and Algorithms

In this section we relax the model simplifications of Section 10.3.2, fitting absorption lines and the continuum jointly in the presence of background contamination, absorption due to the inter-stellar media, and blurring of photon energies. The idealized spectrum after absorption, Z, is treated as one level in a hierarchical data augmentation scheme; the observed data,

model	summary	μ	σ^2	$\log(\lambda^A)$	λ^A
$\lambda^A = 50$	median	1.4998	3.7×10^{-4}	3.1516	23.374
	lower	1.497	2.5×10^{-4}	2.3246	10.223
	upper	1.502	5.27×10^{-4}	4.3794	79.794
	$\sqrt{\hat{R}}$	1.0002	1.0000	1.0003	1.0023
$\lambda^A = 25$	median	1.5008	3.61×10^{-4}	2.7186	15.159
	lower	1.499	2.53×10^{-4}	2.0737	7.954
	upper	1.503	4.99×10^{-4}	3.6074	36.871
	$\sqrt{\hat{R}}$	1.0000	1.0000	1.0054	1.0182
$\lambda^A = 10$	median	1.4993	2.54×10^{-4}	2.4543	11.639
	lower	1.498	2.03×10^{-4}	1.8867	6.597
	upper	1.502	3.89×10^{-4}	3.1804	24.056
	$\sqrt{\hat{R}}$	1.0061	1.0000	1.0000	1.0001
$\lambda^A = 5$	median	1.4992	2.28×10^{-4}	1.9280	6.876
	lower	1.497	1.58×10^{-4}	1.5022	4.492
	upper	1.502	3.17×10^{-4}	2.5070	12.268
	$\sqrt{\hat{R}}$	1.0008	1.0000	1.0007	1.0006
$\lambda^A = 1$	median	1.5005	3.46×10^{-4}	0.02317	1.023
	lower	1.494	1.97×10^{-4}	-0.3654	0.694
	upper	1.507	6.22×10^{-4}	0.3469	1.415
	$\sqrt{\hat{R}}$	1.0026	1.0024	1.0006	1.0007

Table 10.2 *Posterior summaries for data generated according to five simulation models where lower and upper relate to the lower and upper bounds of the 95% confidence interval for the parameter.*

Y, is modeled as in (10.9). Because of the modular structure of our computational tools, it is not difficult to compose posterior sampling and mode finding algorithms in this more general setting. Using the notation of Section 10.3.2, the joint probability model factors as

$$p(X, Y, Z, \theta^A, \theta^{-A}) = p(Y|Z, \theta^{-A})p(Z|X, \theta^A)p(X|\theta^{-A})p(\theta^{-A})p(\theta^A), \tag{10.19}$$

where Y is the observed data, X is the ideal data, Z is the ideal data after absorption, and θ^{-A} are all model parameters not involved in the absorption line. The first factor on the right-hand side of (10.19) represents the effects of blurring, background, effective area of the instrument, and any other absorption components in the model; the second factor represents the

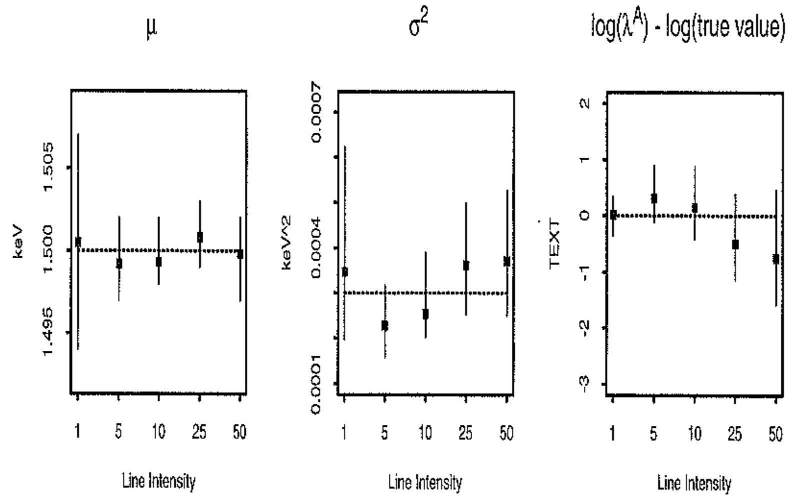

Figure 10.7 *Medians and 95% credible intervals for the five simulations. The horizontal axes represent the five simulations. Intervals in the third graph are translated (by substracting by the logarithm of the true value) so as to be comparable. In each case the dashed lines are the true values.*

absorption line; the third factor is the distribution of the ideal data given in (10.10); and the last two factors are independent prior distributions. Thus, we can construct a two step Gibbs sampler, perhaps with Metropolis-Hastings approximations, to obtain a sample from $p(X, Z, \theta^{-A}, \theta^{A}|Y)$ as follows:

STEP 1: Draw $(\theta^{A}_{(t)}, \theta^{-A}_{(t)})$ from

$$p(\theta^{A}, \theta^{-A}|X, Y, Z) = p(\theta^{A}|X, Y, Z)p(\theta^{-A}|X, Y, Z)$$

STEP 2: Draw $(X_{(t)}, Z_{(t)})$ from

$$\begin{aligned} p(X, Z|Y, \theta^{A}, \theta^{-A}) &= p(X|Y, Z, \theta^{A}, \theta^{-A})p(Z|Y, \theta^{A}, \theta^{-A}) \\ &= p(X|Z, \theta^{A}, \theta^{-A})p(Z|Y, \theta^{A}, \theta^{-A}) \qquad (10.20) \end{aligned}$$

The equalities follow from the factorization in (10.19). Because the draws of θ^{A} and of X are exactly as described in Section 10.3 while the draws of θ^{-A} and Z are given in van Dyk et al. (2001) we can easily implement this MCMC sampler. EM can be adapted using similar arguments.

10.4.2 An Example

To explore the effect of an absorption line on the spectral model of Section 10.2.2 we analyze the X-ray spectrum of Quasar S5 0014+813 (Kühr et al. 1981), using data observed with the ASCA instrument in 1993 (Elvis et al. 1994). We used all 512 of the instrument energy bins, except for the unreliable bins below ~ 0.5 keV or above ~ 10 keV. We model the continuum as a power law, i.e., $\lambda^{C}(\theta^{C}, E_j) = \theta^{C}_1 E_j^{-\theta^{C}_2}$, and add an inter-stellar absorption component, $\alpha_1(\theta^{A}_1, E) = \exp(-\theta^{A}_1/E)$. We account for background by setting θ^{B}_l equal to the (rescaled) counts in the corresponding bin of the background observation. Because the data are relatively informative for θ^{C} and θ^{A}_1, we use flat prior distributions on a variance stabilizing transformation of each parameter. A sample from the posterior distribution was obtained by running three MCMC chains according to the algorithm described in van Dyk et al. (2001) for 1000 iterations each. (We use the nesting methods described by van Dyk and Meng (2000) to reduce the autocorrelation in the resulting chains and produce three draws per iteration.) We then combined the last 2000 draws of each chain to form a sample of 6000 draws. The marginal posterior distributions of $\theta^{C}_1, \theta^{C}_2$ and θ^{A}_1 are represented by the shaded histograms in Figure 10.8.

To explore the effect of an absorption line on our analysis, we manually subtracted counts from nine adjacent bins near 1 keV to simulate an absorption line:

original counts	30	30	28	24	31	37	28	26	29
altered counts	15	10	5	0	0	0	5	10	15

Using the altered data we refit the model exactly as described above (not accounting for the absorption line), yielding the marginal posterior distributions depicted by the histogram with dashed lines in Figure 10.8. The presence of the unaccounted for absorption feature has both increased the posterior variance of all three parameters and has shifted the distributions away from their "true" values. Clearly inference based on this posterior is biased by the non-ignorable missing data caused by the absorption line. Thus, we refit the altered data, this time accounting for the absorption line component as described in Section 10.4. (We use a gamma prior on λ^{A} with $E(\lambda^{A}) = \text{var}(\lambda^{A}) = 2$.) The new marginal posterior (solid lines in Figure 10.8) match the original marginal posterior distributions closely; the bias caused by the absorption line has been removed.

Figure 10.9 shows the cumulative probability of membership of four clusters as a function of energy. The clusters correspond to the photons observed by the detector, photons lost to instrument response, photons absorbed by the inter-stellar media, and photons absorbed in the absorption line. Although these clusters are not specifically spatial in character, if the source were diffuse, we might expect their relative density to vary

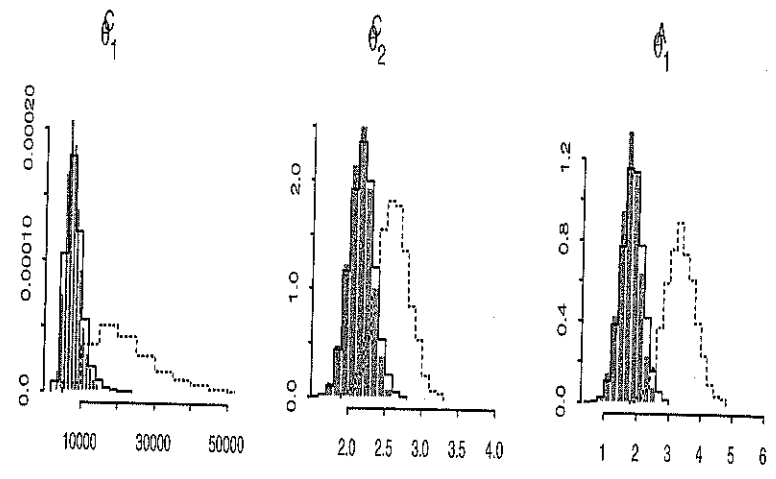

Figure 10.8 *Marginal posterior distributions. The shaded histograms represent the "true" model with no absorption line present. The dashed lines show the bias that an absorption line introduces to the parameter estimates, and the solid lines show the posterior distributions after the absorption line has been accounted for.*

across the source. Three of these clusters are completely unobserved—they are clusters within the idealized data, Y. Similar computations can separate the background, continuum, and emission line photon clusters, all of which are at least partially observed. If we confine our attention to the observed photons, the background, continuum, and emission line clusters are all sub-clusters of cluster 'A' in Figure 10.9. We use the posterior means of the model parameters to produce Figure 10.9; posterior variability can easily be used to compute error bars for the cluster probabilities.

10.5 Discussion

The statistical and computational challenges of image analysis in high energy astrophysics are truly immense. Accounting for the spatial, spectral, and temporal structure in the data along with the complexities in the photon redistribution matrix, pile-up, background contamination, and photon absorption requires highly structured models and sophisticated computing. Current work focuses on incorporating these complexities one at a time, taking advantage of the modular structure in both our models and computational techniques. The preliminary methods are useful for special classes

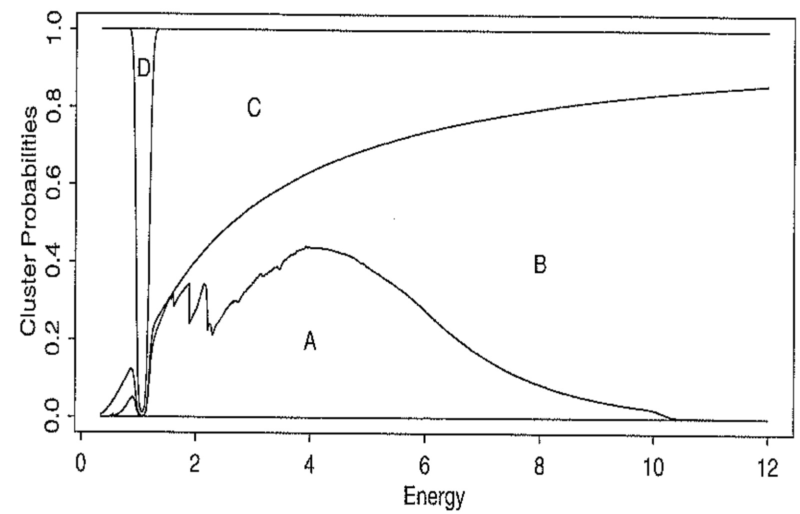

Figure 10.9 *Cumulative Cluster Probabilities. The figure displays the cumulative probabilities of four photon clusters as a function of Energy. Cluster 'A' contains observed photons; 'B' contains photons lost to instrumental response; 'C' contains photons absorbed by the smooth absorption term, α_1, e.g., due to the interstellar media, and 'D' contains photons absorbed in the absorption line. The relative size of the clusters vary dramatically as a function of Energy (the probabilities are computed using the posterior mean of the model parameters).*

of images (e.g., the spectral models can be applied directly to point sources) but need to be extended to be useful for more sophisticated images. Combining the spectral and spatial models is a particular area of active work.

Even within the much less complicated problem of accounting for absorption lines, there are sophisticated statistical and computational challenges. Handling the highly multi-modal posterior distribution will only become more complex as the overall model incorporates more of the features of the source and data collection mechanism. In some cases, the choice of prior distribution can be quite important and careful prior specification along with sensitivity analysis is required. In general, however, we expect the three steps of first exploring the posterior distribution with mode finding algorithms, second fitting the model via MCMC, and finally checking the model using posterior predictive checks to be a practical strategy. Thus far, we have found that directly modeling the stochastic features in the underlying images and data collection to be both powerful statistically and tractable computationally, and thus, a fruitful strategy for image reconstruction.

Acknowledgements

This work is a product of the collaborative effort of the Astro-Statistics Group at Harvard University whose members include A. Connors, D. Esch, P. Freeman, H. Kang, V. L. Kashyap, R. Protassov, D. Rubin, A. Siemiginowska, E. Sourlas , and Y. Yu. The authors gratefully acknowledge funding for this project partially provided by NSF grant DMS-01-04129 and by NASA Contract NAS8-39073 (CXC).

CHAPTER 11

Spatial Modelling of Count Data: A Case Study in Modelling Breeding Bird Survey Data on Large Spatial Domains

C.K. Wikle

11.1 Introduction

The North American Breeding Bird Survey (BBS) is conducted each breeding season by volunteer observers (e.g. Robbins et al. 1986). The observers count the number of various species of birds along specified routes. The collected data are used for several purposes, including the study of the range of bird species, and the variation of the range and abundance over time (Link and Sauer 1998). Such studies usually require spatial maps of relative abundance. Traditional methods for producing such maps are somewhat *ad hoc* (e.g., inverse distance methods) and do not always account for the special discrete, positive nature of the count data (e.g. Sauer et al. 1995). In addition, corresponding prediction uncertainties for maps produced in this fashion are not typically available. Providing such uncertainties is critical as the prediction maps are often used as "data" in other studies and for the design of auxiliary sampling plans.

We consider the BBS modeling problem from a hierarchical perspective, modeling the count data as Poisson, conditional on a spatially varying intensity process. The intensities are then assumed to follow a log-normal distribution with fixed effects and with spatial and non-spatial random effects. Model-based geostatistical methods for generalized linear mixed models (GLMMs) of this type have been available since the seminal work of Diggle et al. (1998). However, implementation is problematic when there are large data sets and prediction is desired over large domains. We show that by utilizing spectral representations of the spatial random effects process, Bayesian spatial prediction can easily be carried out on very large data sets over extensive prediction domains. General discussion of the role of such Bayesian hierachical random effect modelling is given in 1, and approaches to spatio-temporal modelling are found here in 12.