Bayesian Analysis (2006) **1**, Number 2, pp. 241–248

Rejoinder

David A. van Dyk* and Hosung Kang†

We would like to thank Professor Loh and Professor Gelman for their thoughtful comments. Some of the detailed points that the discussants bring up remind us to emphasize that in many regards our article offers only an overview of the scientific questions we aim to address and of the methods, models, and algorithms that we employ. Some technical details are omitted in the interest of readability and space constraints. We hope that readers interested in the problems we discuss will consider reading the several related articles cited here and in the main article.

We have attempted to group Loh's and Gelman's comments into two broad categories: model checking, which we address primarily in the context of our method for DEM reconstruction, and issues specific to our image reconstruction. We discuss these two categories in turn.

1 Model Checking

Model diagnostics are an important part of any model-based statistical analysis, and especially so in the context of complex models of the sort described in our article. Ideally such diagnostics should investigate both internal consistency and objective outside evaluation of the results. Outside evaluations might compare predictions under the model with data not used to fit the model. Examples include cross-validation and the use of comparable data that is available from other sources. Our comparison of the X-ray image reconstruction of NGC 6240 with the optical image produced by the *Hubble Space Telescope* in Figure 9 illustrates this strategy and offers confirmation of our model, algorithms, and methodology. In a Bayesian data analysis, internal consistency is often investigated by comparing the observed data with the posterior predictive distribution. Gelman et al. (1996) describe how one can quantify and assess discrepancies between the two. Such posterior predictive checks are a standard component of our methodology for the parameterized spectral analysis described in Section 3.3. Although we did not describe model checking in the context of spectral analysis in the current article, interested readers are referred to van Dyk and Kang (2004) and van Dyk and Park (2004). These articles show how the posterior predictive distribution can be used to assess the magnitude of the inherently heteroskedastic residuals under Poisson models.

The models for image reconstruction and DEM reconstruction described in Sections 3.4 and 3.5 rely more heavily on blurring matrices (the point spread function and the emissivity matrix, respectively) than does the parametric spectral model. Thus, we expect our results to be more sensitive to misspecification of these matrices. To explore this, we generated several replicate data sets from the posterior predictive distribution

*Department of Statistics, University of California, Irvine, CA, http://www.ics.uci.edu/~dvd/
†Department of Statistics, Harvard University, Boston, MA, http://www.people.fas.harvard.edu/~hkang/

 ba0002

under the DEM reconstruction model. Four replicate data sets are compared with the observed Capella data in Figure 18. The basic structures and line locations of the four replicate data sets appear to be very similar to those of the observed data. Thus, in general terms our model seems appropriate for the data. Two discrepancies, however, appear in the magnitude of the spectral emission lines near 108 and 128 Angstroms; the former appears stronger in the observed data than in the replicated data while the latter appears weaker.

A higher resolution diagnostic can be constructed by looking at residual plots. Figure 19 plots the difference between the observed count in each channel and the center of a 95% posterior predictive interval for that count. The intervals were computed by sampled 2000 replicate data sets from the posterior predictive distribution. The vertical range of the shaded area corresponding to each channel in Figure 19 represents the prediction interval for that channel. The heteroskedastic nature of the residuals is evident. The posterior predictive intervals can be used to assess the magnitude of the residuals. Ideally, we would expect about 5% of the observed counts to fall outside the shaded area. Unfortunately, the plot indicates that the residuals tend to be more dispersed than we would expect under the model: about 14% of the observed counts fall outside the shaded area. The lack of fit does not appear to be simple over-dispersion. Rather, there appears to be structure in the spectrum that is not fully represented by the model: the magnitude of the spectral emission lines near 108 and 128 Angstroms again stand out.

The most likely explanation for this lack of fit is that the precision of the Emissivity matrix is inadequate. Indeed, it is well known that this matrix is recorded with error. In an attempt to quantify the extent of the problem, its authors compute error bars for each element of the matrix. (See the Atomic Database (ATOMDB)[1].) Unfortunately, it is also known that the errors in the matrix are highly correlated and that these correlations are not readily available. Nonetheless, we have made some progress in accounting for imprecisions in the emissivity matrix by fitting the location (in wavelength) of some of the stronger spectral lines. The emissivity matrix provides a strong prior distribution for these locations and the data provides enough information to tweak the locations enough to improve the fit. In addition, we resample the emissivity matrix at each iteration to account for uncertainty in its elements. These innovations to the model were inspired by residual plots of the sort illustrated in Figure 19. The plots in this rejoinder reflect these improvements to the model, while those in the main article do not. These are some of the ways that our model and algorithms were updated during the year and a half between our presentation at *Case Studies in Bayesian Statistics Workshop 7* and this writing.

Judging from Figure 19, there is still room for improvement in our model. There are two basic strategies for accomplishing this improvement. First, we can attempt to further model uncertainties in the emissivity matrix (or elsewhere in the model) by including additional free parameters, perhaps with highly informative prior distributions. We expect that adding flexibility to the model in this way also adds variability to the

[1]URL: http://cxc.harvard.edu/atomdb/ (Smith et al. 2001)

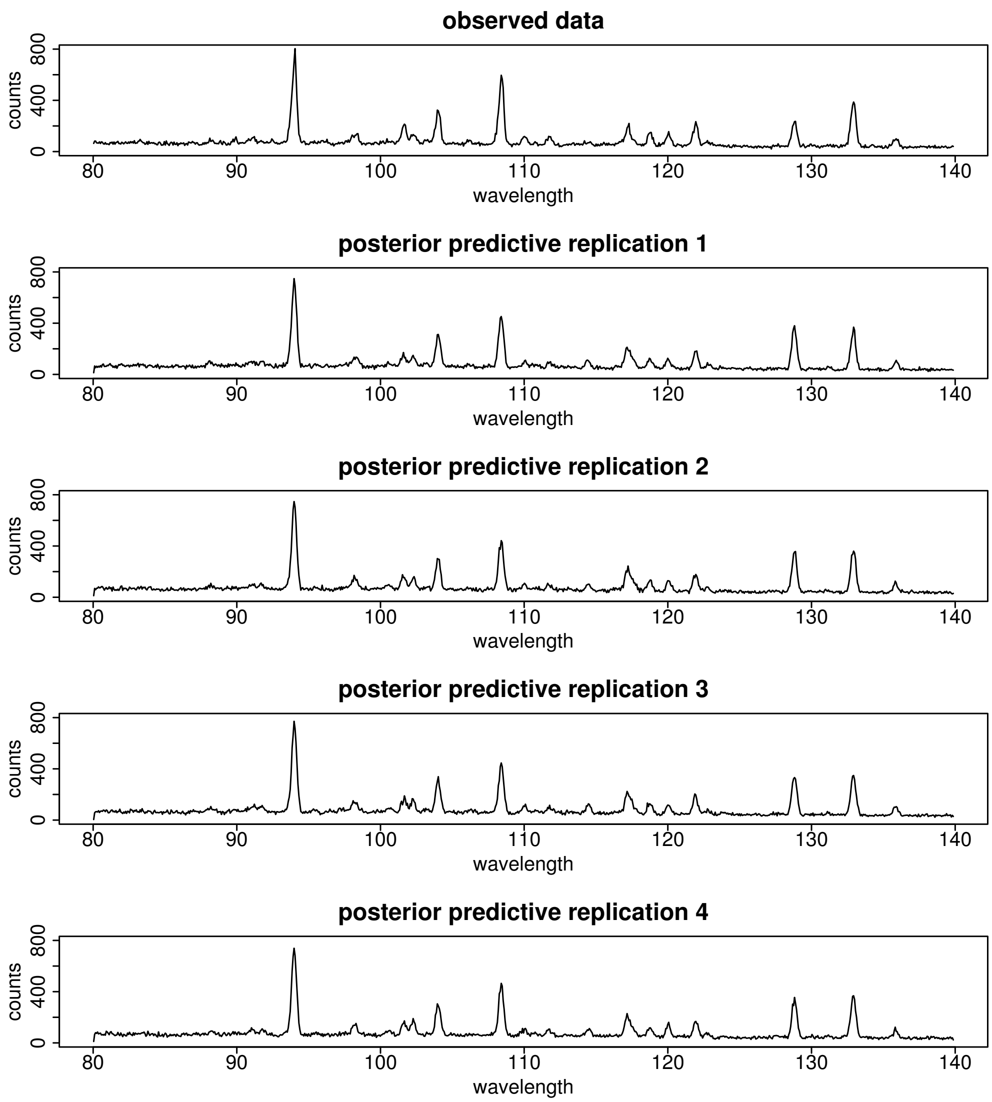

Figure 18: Comparing the Posterior Predictive Distribution with the Observed Capella Data. The first panel shows the observed X-ray photon count data for Capella. The remaining four panels illustrate replicate data sets sampled from the posterior predictive distribution under a DEM reconstruction model. The basic structures and line locations of the four replicate data sets appear to be very similar to those of the observed data.

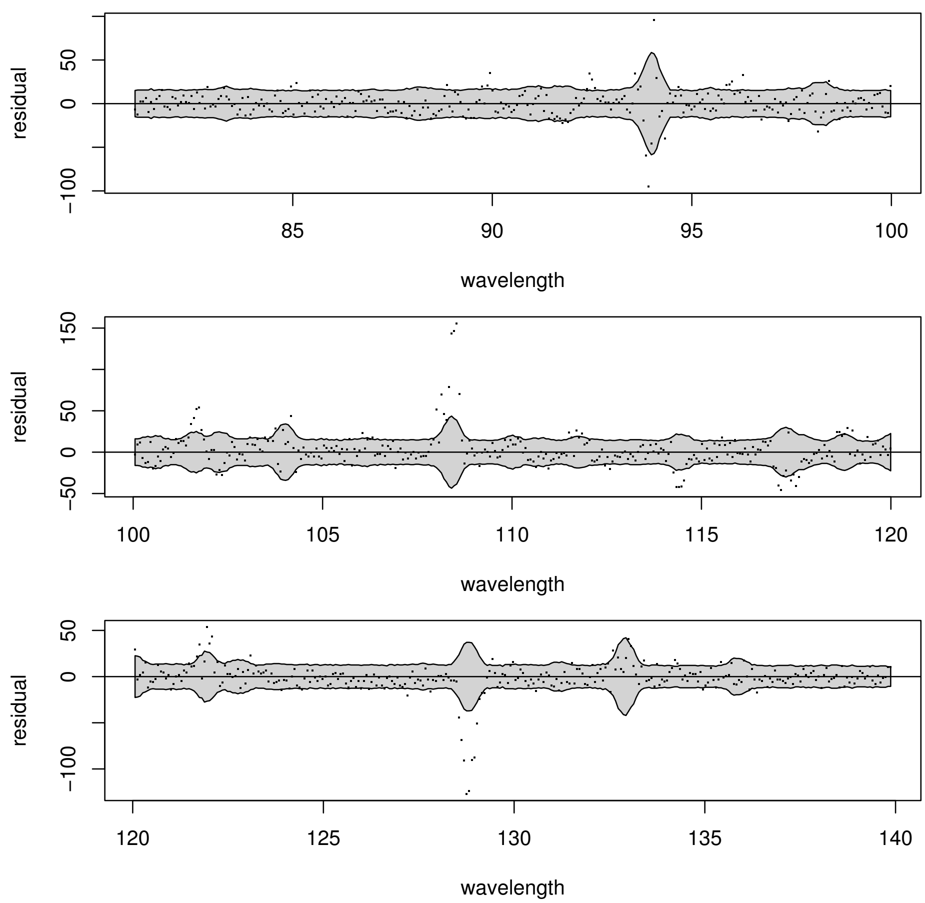

Figure 19: Residual Plot for the Reconstructed DEM of Capella. We plot the difference between the observed channel counts and the center of a 95% posterior predictive interval for that count as a function of channel wavelength. The vertical range of the shaded area corresponds to 95% Monte Carlo prediction intervals for each of the channel counts. Ideally, we would expect about 5% of the observed counts to fall outside the shaded area. That about 14% of the counts fall outside the shaded area indicates that the model does not fully account for the observed variability in the data.

predictive distributions. The second strategy is the one suggested by the discussants: to add an omnibus over-dispersion component to the model. Our colleagues in astronomy have a strong preference for the former strategy because it fosters understanding of the physical processes that give rise to the data. Although our primary goal is to reconstruct the DEM, the emissivity matrix is of interest in and of itself. If we are able to pin down the emissivity matrix or any physical component of the model by reconstructing the DEM, this would be an important scientific contribution. We emphasize that the residual plot in Figure 19 is already an important step in this direction in that the residual plot identifies particular potential misspecifications in the emissivity matrix.

One reason there appears to be little generic over-dispersion in the residuals is that the model already includes components that can at least partially accommodate over-dispersion. A standard conjugate strategy for introducing over-dispersion into Poisson models is to mix the Poisson parameter over its conjugate prior gamma distribution. If the gamma parameters are viewed as model parameters to be fit to the data, this strategy replaces the one-parameter Poisson sampling distribution with a two-parameter negative binomial sampling distribution. In our case, we have multiple conditionally independent Poisson distributions that we parameterize in terms of nested binomial or multinomial distributions. The probability parameters are in turn modeled using conjugate beta or Dirichlet distributions. The result is a mixture of Poisson models that is a more flexible marginal sampling distribution.

2 Image Reconstruction

The discussants astutely observed that the multi-scale prior distribution we describe in Section 3.5 should induce boundary effects in the image reconstructions of NGC 6240. Indeed, the prior distribution has a tendency to produce artifacts that appear in the reconstruction as checkerboard-like patterns; Figure 20 shows a (black and white) reconstruction where these artifacts are clearly present. The reason there are no boundary effects in the reconstructions in Figures 8–10 is that the prior distribution we use is actually somewhat more complex than described in Section 3.5. Algorithmically, we randomly shift the multi-scale representation to a different location on the image at each iteration of the MCMC sampler. This strategy is known as cycle spinning (Coifman and Donoho 1995). Cycle spinning can be obtained as a principled Bayesian technique under a mixture prior distribution. For an $n \times n$ image, there are n^2 possible prior distributions of the sort described in Section 3.5, each centered at a different origin. In effect, we are specifying a prior distribution that is an equally weighted mixture of these n^2 prior distributions. More details of our prior distribution, the hyper-prior distribution, and our sensitivity analyses can be found in Esch et al. (2004).

In their fifth and seventh points, the discussants are concerned with the visual character of the reconstructed images. In particular, they note that the image obtained with Gaussian smoothing (right hand side of Figure 5) more clearly depicts the three (two?) point sources near the center of the galaxy, while our reconstruction (upper right panel of Figure 8) more clearly depicts the loops of hot gas. Because of this difference the

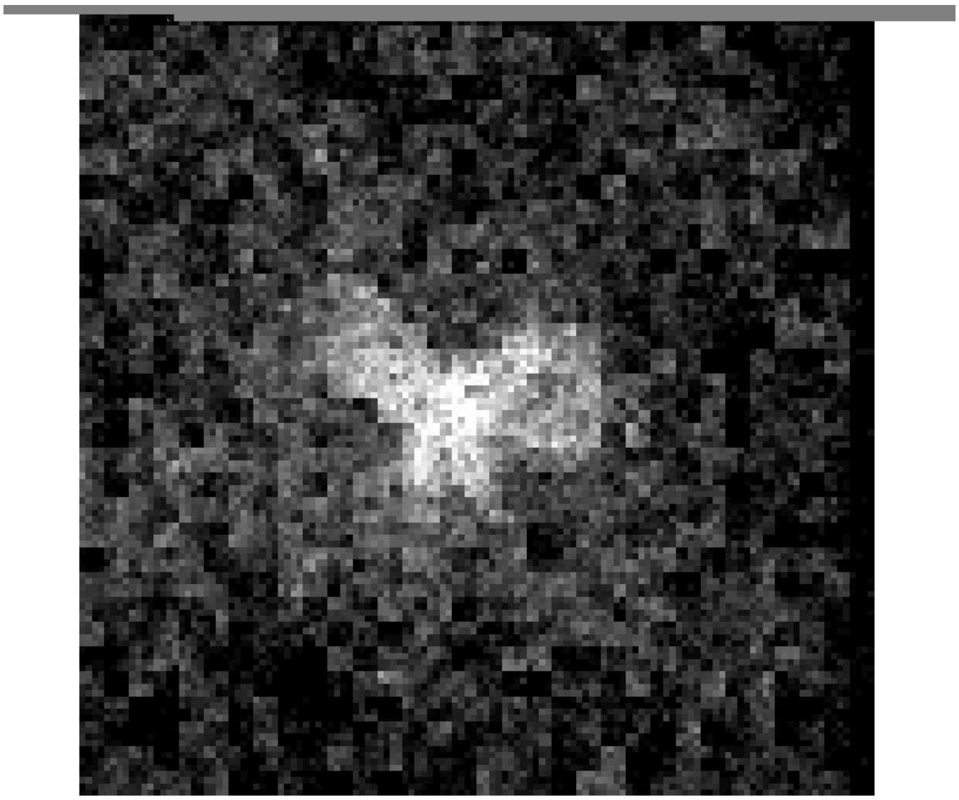

Figure 20: The Advantage of Cycle Spinning. The plot shows an image reconstruction of NGC 6240 obtained without cycle spinning. The checkerboard artifacts from the prior distribution are evident.

discussants conclude, “It is not clear what to believe.” We find this conclusion quite puzzling.

Deciding which method is more reliable should not be based, as the discussants seem to suggest, on which of the resulting images shows more structure. We believe there are two more relevant criteria. First is a thoughtful consideration of the relative merit of the principles behind the methods, models, and algorithms that underlie the reconstructions. It is not surprising that the two methods give different results. Our method accounts for the Poisson nature of the data, incorporates the point spread function, and attempts to quantify uncertainty in the resulting reconstruction; Gaussian smoothing, on the other hand, merely locally blurs the data in order to produce a prettier picture. (Is this the “desirable and attractive” feature that the discussants refer to?) The question should be: which of these strategies is more likely to isolate actual structures in the astronomical sources? The second criterion for comparing image reconstruction techniques is to employ objective outside standards. Simulation studies, where the true image is known, are valuable tools that we use to evaluate our methods (Esch et al. 2004). Comparisons with much higher-resolution optical images is another important tool that is described in the article (see Figure 9). We emphasize that we do not mean to imply that Gaussian smoothing is not a useful technique; it can be an important tool in exploratory analyses and in producing visually stunning images. However, it is less useful as a scientific tool for addressing the substantive questions we outline in the article. Although we certainly believe there is room for improvement in our methods as tools for addressing these scientific questions, we do not view Gaussian smoothing as a gold standard!

We should also note that the image obtained with Gaussian smoothing is actually a composite of three images: the red image smooths the low energy counts, the green image smooths the mid-energy counts, and the blue image smooths the high energy counts. Thus, the two point sources near the center are highlighted because there is both a large number of photons in these regions and the photons are of very high energy; the point sources are believed to be black holes. Our reconstruction, on the other hand, marginalizes over the photon energies and thus is not a representation of this extra information. Nonetheless, the two point sources are local maxima in our reconstruction. One is clearly visible as a white pixel in Figure 10; the other is not white in the color map used for the posterior mean, but is white in the color map used in the significance maps.

We agree with the discussants that the choice of color map can greatly affect the appearance of the images. This, however, is more of an issue for publications than for the scientists. Interactive software is available that allows scientists to change the color map and judge the sensitivity of their conclusions to the choice of map. A more pressing issue for Bayesian statisticians is how to represent and quantify structures in the high-dimensional posterior and posterior predictive distributions. The complex structures of interest to astronomers are difficult to describe mathematically, and thus their significance under these distributions is difficult to quantify. Perhaps the best strategy is to make movies that follow a Markov chain on its tour of the distribution.[2] Although such movies do give some insight into these high-dimensional distributions, it remains difficult to use them to quantify the degree of evidence for a particular structure in an astronomical source.

References

Coifman, R. R. and Donoho, D. L. (1995). "Translation-Invariant De-Noising." In *Wavelets and Statistics* (Editors: A. Antoniadis and G. Oppenheim), 125–150. New York: Springer–Verlag. 245

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. 245, 247

[2]URL: http://www.ics.uci.edu/~dvd/Astro/EMC2/EMC2movies.html

Gelman, A., Meng, X.-L., and Stern, H. (1996). "Posterior Predictive Assessment of Model Fitness (with Discussion)." *Statistica Sinica*, 6: 733–807. 241

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. 242

van Dyk, D. A. and Kang, H. (2004). "Highly Structured Models for Spectral Analysis in High Energy Astrophysics." *Statistical Science*, 19: 275–293. 241

van Dyk, D. A. and Park, T. (2004). "Efficient EM-Type Algorithms for Fitting Spectral Lines in High-Energy Astrophysics." In *Missing Data and Bayesian Methods in Practice: Contributions by Donald Rubin's Statistical Family* (Editors: A. Gelman and X.-L. Meng), 285–296. New York: Wiley & Sons. 241