Bayesian Analysis (2007) **2**, Number 4, pp. 665–680

Estimation of Faraday Rotation Measures of the Near Galactic Sky Using Gaussian Process Models

Margaret B. Short*, David M. Higdon† and Philipp P. Kronberg‡

Abstract. Our primary goal is to obtain a smoothed summary estimate of the magnetic field generated in and near to the Milky Way by using Faraday rotation measures (RM's). Each RM in our data set provides an integrated measure of the effect of the magnetic field along the entire line of sight to an extragalactic radio source. The ability to estimate the magnetic field generated locally by our galaxy and its environs will help astronomers distinguish local versus distant properties of the universe. RM's can be considered analogous to geostatistical data on a sphere. In order to model such data, we employ a Bayesian process convolution approach which uses Markov chain Monte Carlo (MCMC) for estimation and prediction. Complications arise due to contamination in the RM measurements, and we resolve these by means of a mixture prior on the errors.

Keywords: Markov chain Monte Carlo, Gaussian process, error mixture model, spatial model

1 Introduction

There is widespread interest in understanding the role of magnetic fields on all scales and at all epochs in the history of the Universe. The strength and structure of large scale magnetic fields in astrophysics is just beginning to be probed with meaningful precision now that Faraday rotation measurements (RM; see Section 2), the prime detector of such fields, are available in sufficiently large numbers. This now justifies appropriate statistical methods to facilitate interpretation of the RM values in terms of cosmic magnetic fields. This paper describes the application of new statistical treatments to the largest RM data set available to date. These *ca.* 1600 RM measurements, spread over the spherical sky, contain multiple causes for their RM values, so that it is not trivial to *a priori* separate out different contributions to a given RM measurement.

Our Milky Way imposes a large imprint of foreground RM on the all-sky distribution of RM's. The aim of our analysis is to obtain the most statistically reliable estimate of this Milky Way foreground in the presence of "anomalous" values so that we can optimally remove it. This process is a prerequisite for analysing a variety of magnetic structures in the more distant universe.

*Department of Mathematics and Statistics, University of Alaska Fairbanks, Fairbanks, AK, ffmbs1@uaf.edu

†Statistical Sciences Group, Los Alamos National Laboratory, Los Alamos, NM, dhigdon@lanl.gov

‡Institute for Geophysics and Planetary Physics, Los Alamos National Laboratory, Los Alamos, NM kronberg@lanl.gov

To date, the attempts to produce a smooth map of RM's have consisted mostly of simple, heuristic algorithms such as the one described in Simard-Normandin and Kronberg (1980). Briefly, the method computes RM values locally in small circular regions of diameter D on the sky by obtaining a mean and standard deviation σ of the RM values of extragalactic radio sources within D – after iteratively discarding RM values that are over $x\sigma$ from the mean (x is typically 1.3 – 1.5). This method is essentially deterministic and although uncertainty estimates are not stated explicitly, a statistic such as 2σ might serve as an initial approximation. Figure 2 in Simard-Normandin and Kronberg (1980) shows estimates of smoothed RM using their procedure on a fixed grid of points. Several points on the grid have no estimate since those portions of the sky were lacking in adequate angular coverage of RM measurements.

Rotation measures are normally calculated at the galactic longitude-latitude positions (l, b) of discrete radio sources, and thus may be considered analagous to geostatistical data taken at irregularly spaced points on a sphere. In this sense they can be considered prototypical of other problems involving geostatistical data on a sphere, for example, atmospheric data. Previous efforts at modeling processes on a sphere include two papers using wavelet approaches; namely Oh and Li (2004), which estimates global temperature fields using scattered observations; and Frick et al. (2001) which estimates RM using 841 observations. While similar to these wavelet approaches, our process convolution model is embedded in a larger modeling framework that incorporates outliers and gives prediction uncertainties.

Another approach is possible, based on covariance functions. Such functions for a sphere have been described in the literature. See, for example, Gneiting (1999), Weber and Talkner (1993), Gaspari and Cohn (1999), and Yaglom (1987). These provide a basis for standard Bayesian statistical approaches to modeling Gaussian processes, which typically rely on a parametric covariance function to describe spatial dependence. Exploration of the resulting posterior is particularly difficult when the number of observations is moderately large ($n > 1000$) since evaluating the posterior requires a solve of an $n \times n$ covariance matrix.

Our method uses a process convolution model implementation which does not require factorizations of large matricies. Markov chain Monte Carlo (MCMC) is used for parameter estimation as well as prediction on a dense set of spatial locations. Strengths of our approach are its ability to readily produce uncertainty estimates, its easy incorporation of rich error structures, and its ability to handle moderately large data sets and dense sets of prediction locations.

Section 2 gives a description of the RM data set. Section 3 describes a process convolution-based Gaussian process model for the sphere, or any other compact manifold. Section 4 spells out the models and model components we are going to explore, followed by a sketch of our posterior sampling recipe based on MCMC. Section 5 describes results for the RM data set, while Section 6 discusses the sensitivity of our results to alternative model choices. Section 7 concludes with a brief discussion.

2 Faraday rotation measures

Faraday rotation measure is defined as the line integral over a path of length L (in parsecs, with 1 pc = $3{\times}10^{18}$ cm)

$$RM = k \int_0^L n_e \mathbf{B} \cdot ds, \tag{1}$$

where $k = 8.1 \times 10^5$, n_e is the electron density (measured in cm^{-3}), and $\mathbf{B}$ is magnetic field (in units of gauss); it is measured in units of radians per square meter of wavelength. Earth corresponds to one end point of the integral. Looking outward in any radial direction to a given radio source gives us another line integral at some position (l, b) on the galactic sky. Thus we may think of RM as a scalar quantity (positive or negative) which is a function of location on a sphere.

In fact, the locations on the sphere at which we are able to directly estimate RM are only in the selected directions where extragalactic radio sources such as radio galaxies and quasars happen to lie. Signals from these sources are usually highly polarized, and their polarization angle is rotated during passage through ionized gases in the presence of a magnetic field according to (1). For each of these radio sources, the data give a good approximation to the ideal Faraday rotation law, which states that

$$RM = \frac{\chi(\lambda_1) - \chi(\lambda_2)}{\lambda_1^2 - \lambda_2^2},$$

for every pair of wavelengths $\lambda_1 \neq \lambda_2$, where $\chi(\lambda)$ is the polarization angle associated with a given wavelength λ. This allows us to estimate rotation measure by measuring the polarization angle associated with various wavelengths, using a radio telescope.

Although we shall start referring to the RM values in our data set as observations, they are, in fact, values calculated from actual observations. A complete description of the algorithm used to estimate the RM's can be found in Simard-Normandin et al. (1981); that paper also gives associated uncertainty estimates for the observations. A condensed description of the algorithm is as follows. For each radio source, a collection of pairs $\{(\lambda_i, \chi(\lambda_i)\})$ measured over a range of wavelengths, λ_i, is used to estimate RM as the slope of a least squares fit to the line

$$\chi(\lambda) = \text{constant} + \text{RM} \cdot \lambda^2.$$

The number of usable pairs $(\lambda_i, \chi(\lambda_i))$ varies from one radio source to another. A minimum of four such pairs were used for each RM determination, and typically between four and ten pairs were used. We made no explicit effort to account for the original uncertainties in $\chi(\lambda_i)$, but rather took at face value the data we were given. In fact, many radio sources were omitted from the data set altogether due to investigator uncertainty over the reliability of the RM estimate.

Most of the data used in this paper are provided by one of us (PPK) and are not yet publicly available. Figure 1 shows these RM values, which correspond to 1566 distant

radio sources. Red circles indicate positive RM's, blue circles negative RM's. More intense reds and blues correspond to RM's having larger magnitude. (Please see the online version of the paper at ba.stat.cmu.edu if these colors are not apparent in your copy.) In this figure, the equator corresponds to the plane of the Milky Way. The center point $(l, b) = (0^o, 0^o)$ on the figure points directly at the center of the Milky Way. This is an Aitoff projection map, showing the entire sky. By convention, galactic longitudes l increase as one moves left. Figure 1 also includes a histogram of the RM data; the empirical distribution is approximately symmetric and has extremely long tails.

It is clear from Figure 1 that there are systematic, large scale variations, not yet entirely understood, often, but not always near the plane of the Milky Way in which we are embedded. There is also a substantial region close to the galactic plane ($b = 0^o$), and around the galactic center, having large negative and large positive rotation measures. As a soft rule, the farther away from the plane of the Milky Way, the smaller are the magnitudes of the RM's.

As mentioned earlier, we can directly estimate RM only in the directions of the extragalactic radio sources. It is a bit of a misnomer to say we wish to estimate a smooth RM field (although we shall continue in our abuse of the language in this matter). Rather, we wish to produce a smoothed map of $k \int_0^{L^*} n_e \mathbf{B} \cdot ds$, where, in each radial direction, the line integral extends not far beyond the limits of the Milky Way. For many but not all sources, that line integral is determined largely by the Milky Way.

We note in passing that RM's have also been measured for some pulsars which lie in a 3-D space within the Milky Way. But since the line integrals we wish to estimate $(\int n_e \mathbf{B} \cdot ds)$ contain contributions from beyond the limits of the Milky Way, the pulsar RM data need to be interpreted in a different context and thus we exclude them from our analysis.

If there are no substantial intervening magnetic fields between us and the radio source we are examining, then the major contributing factor to RM is usually the Milky Way itself. However, since RM is a line integral involving the magnitude of the magnetic field along the entire line of sight, sources having the longest extragalactic lines of sight are more likely than nearby sources to be intersected by intervening galaxies which may have substantial magnetic fields. Such interveners, for example, can produce the outliers that are seen by looking at a histogram of the raw data (Figure 1); the excessively long tails strongly suggest that a statistical model with simple independent and identically distributed (iid) errors will be inadequate. The outliers are thus independent of effects that cause systematic (l, b) variations of RM on the sky. In this context, we shall begin referring to data as being "contaminated" or "uncontaminated", contaminated RM's being those that are largely intrinsic to distant radio sources whose sightlines pass through distant intervening magnetized clouds, as illustrated in the sketch in Figure 2. In order to accommodate the varying degrees of contamination, we explored two mixture models for the errors. These are discussed in Sections 4 and 6.

The careful observer will note the presence of several extremely large negative and positive RM's close to the equator of the Milky Way, with latitudes $|b| < 5^o$. For each

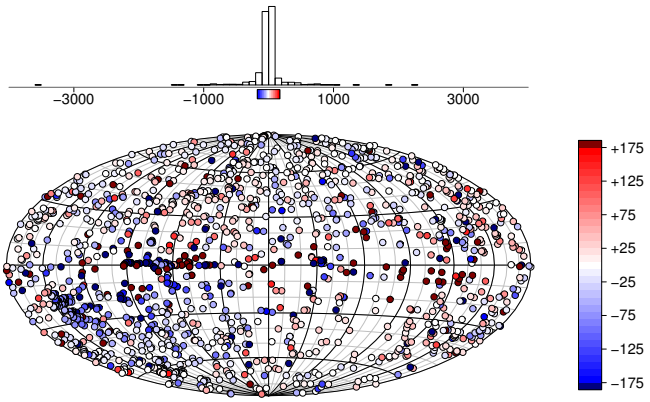

Figure 1: 1566 RM observations; units are rad/m^2. Large RM's (intense red and blue circles) tend to cluster along the plane of the Milky Way; higher latitudes exhibit smaller RM's. The histogram shows the extremely long tails, corresponding to "contaminated" observations. (Please see the online version of the paper at ba.stat.cmu.edu if the colors are not apparent in your copy.)

of these, the line of sight passes through strongly Faraday rotating zones close to the plane of the Milky Way. It is thus believed that the primary reason these RM's are so large is because of local effects due to the Milky Way.

3 Gaussian process models on a sphere

A fairly general Gaussian process (GP) model for a compact manifold $\mathcal{S}$ (e.g. a sphere, a torus, a potato) with distance metric $d(\cdot,\cdot)$ can be created by taking a collection of uniform, regularly spaced knot locations $\mathbf{w}_1,\ldots,\mathbf{w}_J$, and assigning to each of these locations a knot value $x_1,\ldots,x_J$ which are assumed to have iid $N(0,[J\lambda_x]^{-1})$ distributions. Convolving these knot values with a simple smoothing kernel $k(\cdot)$ then results in the GP model

$$z(\mathbf{s}) = \sum_{j=1}^{J} x_j k(d(\mathbf{s},\mathbf{w}_j)), \quad \mathbf{s}\in\mathcal{S}. \tag{2}$$

Figure 3 shows an example where $\mathcal{S}$ is the unit circle.

The covariance between any two locations is given by the formula

$$\mathrm{Cov}(z(\mathbf{s}_1),z(\mathbf{s}_2)) = [J\lambda_x]^{-1}\sum_{j=1}^{J} k(d(\mathbf{s}_1,\mathbf{w}_j))k(d(\mathbf{s}_2,\mathbf{w}_j)). \tag{3}$$

Since $\mathcal{S}$ is bounded, increasing J increases the density of knots. Provided the knots are added uniformly over $\mathcal{S}$ as $J\to\infty$, the covariance function (3) converges to a function

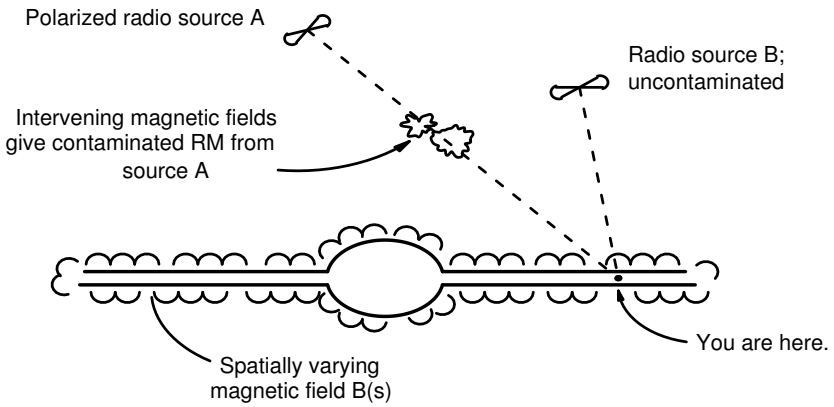

Figure 2: Cartoon diagram of the Milky Way and contributors to rotation measure. The RM obtained from radio source B is determined only by the Milky Way, whereas the RM from radio source A is contaminated by extragalactic interveners along the line of sight.

that depends only on $k(\cdot)$, the distance $d(\mathbf{s}_1, \mathbf{s}_2)$ and λ_x, so that $z(\mathbf{s})$ is a stationary GP in the limit. Note that this stationarity is with respect to the distance metric $d(\cdot, \cdot)$ on $\mathcal{S}$. Analogous process convolution models for $\mathbb{R}^d$ are described in Higdon (2002).

For a finite J, (2) is only approximately stationary. How close to stationary depends on the size of nearest knot-to-knot distances relative to the width of the convolution kernel $k(\cdot)$. Through empirical studies on a variety of manifolds using geodesic distance we have found that if $k(\cdot)$ is normal with SD σ_k, a knot-to-knot spacing of no more than σ_k leads to a very good approximation to the limiting stationary model on $\mathcal{S}$. In $\mathbb{R}^d$, this construction converges to a GP with the Gaussian covariance function Higdon (2002). Our initial investigations suggest this rule of thumb works for the circle, the sphere, the torus, and $\mathbb{R}^d$, $d \leq 3$, as well. This gives a recipe for a generalized Gaussian covariance function on these different manifolds, where σ_k controls the correlation distance. This recipe may be more general, we just haven't tried other manifolds $\mathcal{S}$. It is important to note that this discrete, process convolution construction is only useful for representing GP's with smooth realizations; using a very peaked smoothing kernel will require a very dense set of knot locations, rendering this discrete construction impractical.

For the application of this paper, we take $\mathcal{S}$ to be the unit sphere $S^2 = \{(x, y, z) : x^2 + y^2 + z^2 = 1\}$ corresponding to the RM sky, $k(\cdot; \sigma_k)$ to be a normal density with standard deviation σ_k, and $d(\cdot, \cdot)$ to be great circle distance. We use a recursive tessellation algorithm (http://www.fho-emden.de/∼hoffmann/ikos27042002.pdf, for example) to determine the knot locations that are fairly evenly distributed over the sphere. The algorithm begins by embedding an icosahedron in the sphere. This gives us a collec-

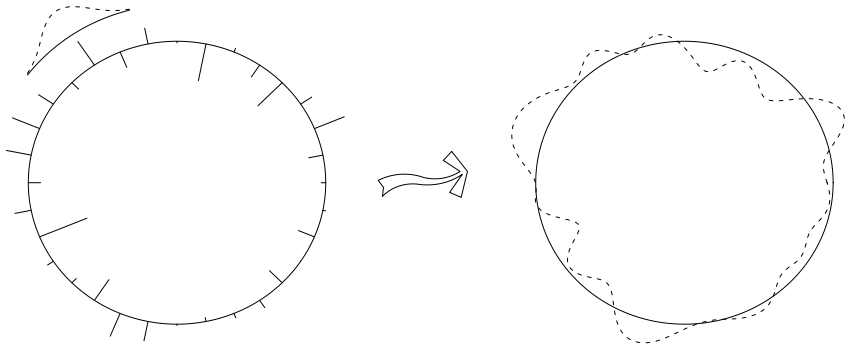

Figure 3: Construction of a realization of a GP on the unit circle. Left panel shows knot locations and values (positive and negative), along with a typical kernel. Right panel shows the resulting realization.

tion of 12 equally spaced locations (vertices) on the sphere. We subdivide each of the 20 equilateral triangular faces into four faces – no longer equilateral – by joining the midpoints of the sides and projecting the three new vertices out onto the surface of the sphere. This results in 42 vertices, approximately equally spaced, as well as 80 faces. The subdivision procedure is repeated as many times as necessary. Using $v^{(g)}$ and $f^{(g)}$ to denote the number of vertices and faces, respectively, after g iterations, it is easy to verify that $v^{(3)} = 642$, $v^{(4)} = 2562$, $f^{(4)} = 5120$, and $f^{(5)} = 20480$. We use the vertices as knot locations (and the centers of the faces for prediction locations). Figure 4 shows the 642 knot locations corresponding to the third level of subdivision.

For this application, we need enough knots J to ensure that the process convolution representation (2) gives an adequate approximation to the stationary model. This means that the neighboring knot-to-knot spacing needs to be no greater than the minimum plausible size for the kernel SD σ_k. In practice we select the number of knots by first making an initial specification for J. By looking at the estimated posterior distribution for σ_k, we can then decide if the knot density is adequate. In this application, we started with the 642 knots shown in Figure 4. Since the lower tail of the posterior distribution for σ_k did go below the knot-to-knot spacing, we opted to use the next iteration of the tessellation algorithm, giving $J = 2562$. In this case, the posterior distribution for σ_k stayed well above the knot-to-knot spacing, with the posterior mean for σ_k being about 1.9 times the knot-to-knot spacing. Of course, we could have used the knot tessellation produced by even finer subdivisions of the recursive algorithm. However this would greatly increase the computational effort, while giving only a slightly better approximation to the limiting, stationary model for $z(\mathbf{s})$.

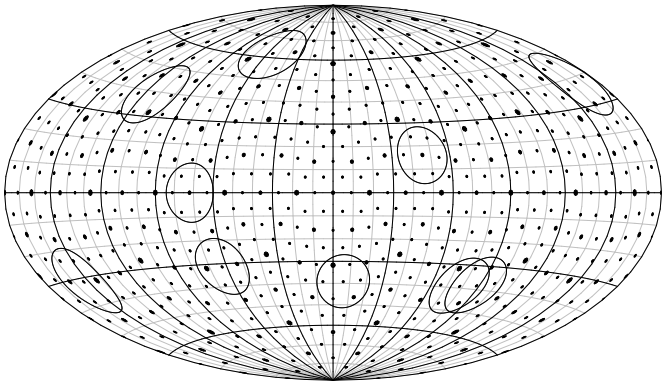

Figure 4: 642 knot locations. Large dots correspond to the 42 vertices after one subdivision, medium dots correspond to the next subdivision, small dots correspond to the third subdivision. Each of the large circles has a radius equal to 1.6 times the knot spacing.

4 Statistical model

We use the GP model over the sphere $z(\mathbf{s})$ described in the previous section to model the smooth RM field. This model uses $J = 2562$ knots distributed over the unit sphere, which give a neighboring knot-to-knot distance of approximately $2\pi/80$. The N observations taken at locations $\mathbf{s}_1, \ldots, \mathbf{s}_N$ are modeled as

$$\mathbf{Y}(\mathbf{s}_i) = z(\mathbf{s}_i) + \epsilon_i,\ i = 1, \ldots, N$$

where the errors ϵ_i are independent with $N(0, [\omega_i \lambda_\epsilon]^{-1})$ distributions. The ω_i's, which modify the error precision, account for the possibility that some measurements are contaminated by one (or more) intervening magnetic field zones. The effect of a given intervener can be similar to the effect of our local Milky Way galaxy (which is responsible for the RM field we are attempting to estimate), or can be an order of magnitude stronger. Hence, for a given observation i, the appropriate value for the precision modifier ω_i could range from slightly less than one to nearly zero. A priori, we expect somewhere around 80% of the measurements to be free of contamination. We explore the sensitivity of our analysis to some of these assumptions in the next section.

Our prior specification is summarized below. Additional detail follows.

$$\begin{aligned}
\mathbf{x}|\lambda_x &\sim MVN(0, {\lambda_x}^{-1} I_{J\times J}) \\
\sigma_k &\sim U(2\pi/80, 5\cdot 2\pi/80) \\
\lambda_x &\sim \Gamma(1, .001) \\
\lambda_\epsilon &\sim \Gamma(1, .001) \\
\omega_i &\overset{\text{iid}}{\sim} \begin{cases} \delta_1 & \text{with probability } .8 \\ \text{Uniform}(0,1) & \text{with probability } .2 \end{cases}.
\end{aligned}$$

Here the gamma priors for λ_x and λ_ϵ have a prior mean of 1000; δ_1 denotes a point mass at 1. Note that any overall mean for the RM field is to be absorbed in $z(\mathbf{s})$. This should be adequate since it appears from Figure 1 that the RM field is, for the most part, centered near zero. The prior for the kernel SD σ_k is uniform, with a lower limit that is set to the knot-to-knot distance of the $J = 2562$ knot tessellation, and an upper bound that is five times that spacing. This ensures that $z(\mathbf{s})$ contains no artifacts due to overly dispersed knot locations.

The resulting full conditional distributions are given by

$$x_j|\cdots \sim N(\frac{\lambda_\epsilon {k^j}^T \Sigma^{-1} r^j}{{k^j}^T \Sigma^{-1} k^j + \lambda_x}, \frac{1}{\lambda_\epsilon {k^j}^T \Sigma^{-1} k^j + \lambda_x})$$

$$\lambda_x|\cdots \sim \Gamma(1 + \frac{J}{2}, .001 + \frac{1}{2}\mathbf{x}^T\mathbf{x})$$

$$\lambda_\epsilon|\cdots \sim \Gamma(1 + \frac{N}{2}, .001 + \frac{1}{2}(\mathbf{Y} - K\mathbf{x})^T\Sigma^{-1}(\mathbf{Y} - K\mathbf{x}))$$

,

$$p(\omega_i|\cdots) \propto \omega_i^{\frac{1}{2}} \exp\left\{-\frac{1}{2}\omega_i\lambda_\epsilon(Y(s_i) - z(s_i))^2\right\}[.8\delta_1 + .2I[0 < \omega_i < 1]]$$

and

$$p(\sigma_k|\cdots) \propto \exp\left\{-\frac{\lambda_\epsilon}{2}(\mathbf{Y} - K\mathbf{x})^T\Sigma^{-1}(\mathbf{Y} - K\mathbf{x})\right\} I[2\pi/80, 10\pi/80]$$

where k^j is the j^{th} column of the matrix K, and r^j is an $n \times 1$ vector given by

$$r^j = \mathbf{Y} - \sum_{j' \neq j} k^{j'} x_{j'}$$

and

$$\Sigma = Diag(1/\omega_1, \ldots, 1/\omega_N).$$

5 Results

The model described in Section 4 was implemented in C++. Draws from the posterior distribution for the parameters were obtained via MCMC, using Gibbs steps for the

knot values $\mathbf{x}$ and for the precision parameters λ_x and λ_ϵ. Metropolis steps were used for the remaining parameters, i.e. $\{\omega_i\}$, and the kernel SD σ_k.

The first 50,000 MCMC iterations were discarded, as they occurred during an adaptive phase that was used to bring the Metropolis acceptance rates into the range of 30% to 60%. Of the next 200,000 iterations, 2000 equally spaced iterations were retained for posterior summarization. Visual inspection of trace plots of the parameters indicated that convergence had occurred. Additionally, we note that the MCMC error for each of the parameters was less than 2% of its posterior mean, and was often substantially less.

Prediction at spatial locations $(\tilde{\mathbf{s}}_1, \tilde{\mathbf{s}}_2, \ldots, \tilde{\mathbf{s}}_m)$, was accomplished by evaluating the expression $\tilde{\mathbf{Y}} = \frac{1}{G}\sum_{g=1}^{G} K^{(g)}\mathbf{x}^{(g)}$, where $\{\mathbf{x}^{(g)}\}_{g=1}^{G}$ are the posterior draws for $\mathbf{x}$, and the (i,j) entry of $K^{(g)}$ is given by $K_{ij}^{(g)} = k(d(\tilde{\mathbf{s}}_i, \mathbf{w}_j); \sigma_k^{(g)})$. Note that with this process convolution formulation, constructing posterior draws of $z(s)$ over a large prediction grid is trivial; this is a substantial undertaking under a standard covariance-based GP formulation.

Final results appear in Figures 5 – 8. Figure 5 shows predicted RM values at 5120 approximately equally spaced locations. The plot is generally consistent with the data (indicated by small circles). It shows blocks of large RM values along the galactic equator and small values near the poles. For this plot, there were $N = 1566$ observations, $J = 2562$ knots, $m = 5140$ prediction locations, and $G = 2000$ MCMC samples.

Figure 6 shows the width of 95% credible sets at the same set of 5120 locations. Not surprisingly, the widest sets occur where the non-contaminated data are sparsest. Figure 7 provides North Pole and South Pole perspectives, and is included to remove the high latitude distortion of the Aitoff projection used in Figure 5. Figure 8 shows the corresponding credible sets.

Posterior diagnostic plots are shown in Figure 9. In the left panel, which shows the posterior mean of the ω's versus observed RM's for which $|RM| < 1000$, we see that smaller RM's generally correspond to larger ω's and vice versa. This is in accordance with our model, since $\omega = 1$ implies uncontaminated data, and ω's progressively closer to 0 imply progressively larger degrees of contamination. The marginal posterior probability that a randomly chosen observation is an outlier is 0.23. The histogram in the middle panel shows the posterior mean value of the ω's for all 1566 RM values. As expected, a sizable majority of the observations are designated as uncontaminated, having values of ω close to 1. An empirical influence function appears in the right panel, showing standardized residuals, $[\hat{\lambda}\hat{\omega}_i]^{-\frac{1}{2}}(Y_i - \hat{Y}_i)$, versus residuals, $Y_i - \hat{Y}_i$. As expected from the model, contaminated observations correspond to large residuals, which are in turn downweighted by small $\hat{\omega}_i$'s.

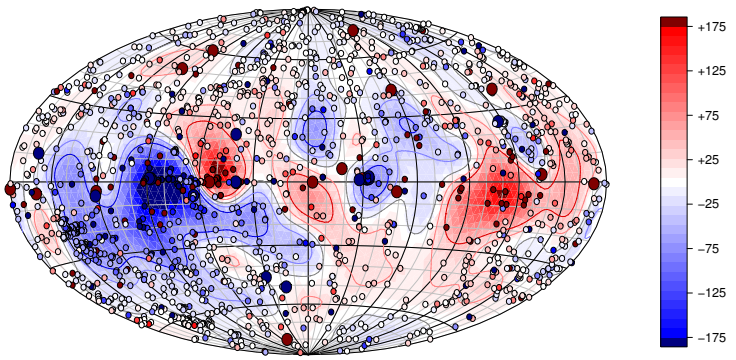

Figure 5: Posterior plot for the final model. Contour lines appear at levels corresponding to RM values of -50, -25, 0, 25 and 50 rad/m^2. The most extreme data are depicted by circles that are larger than those used for the remainder of the data.

6 Sensitivity analysis and model assessment

Here we assess the sensitivity of our analysis to modeling choices regarding the normal mixture model for the errors ϵ_i and the spatial process model $z(\mathbf{s})$ made in the previous section. As for the error model, the posterior influence shown in the right hand frame of Figure 9 shows how the data are being treated to estimate the RM field. The linear piece in the plot shows that the middle 77% of the deviations are basically treated as iid normals, with a common standard deviation. The most extreme residuals are being downweighted so that they have the effect of a residual that is slightly more than a standard deviation away from the fitted value. One may prefer that the outlier model give these residuals zero weight. This could be accommodated by specifying a prior for ω_i that has point masses at 0 and 1. The changing dimensionality of such a model makes it a bit more difficult to fit. We assessed the influence of these extreme weights by refitting the model after removing all of the RM data for which the posterior median for ω_i was not one. This was about 15% of the observations. As expected, the RM field estimated from our original analysis which includes all of the data shows a slight influence from the extreme observations. The average absolute deviation between the two posterior mean fields at the 5140 prediction locations was about 9 radians/m^2. This is small given the posterior mean field varies between ± 300, and the posterior mean for $(\lambda_\epsilon \omega_i)^{-\frac{1}{2}}$ for the uncontaminated observations is about 28.

The sensitivity to the prior probability of 80% that any given observation is contaminated was also investigated. This was done by refitting the model using the values 60%, 70% and 90%. In each case, the differences in the posterior mean field for $z(\mathbf{s})$ are very slight. Most differences appear near data points that are slightly extreme, but not obviously contaminated. The effect of the uncontaminated observations remains the same since the error precision λ_ϵ can compensate for changes in ω_i that are common to

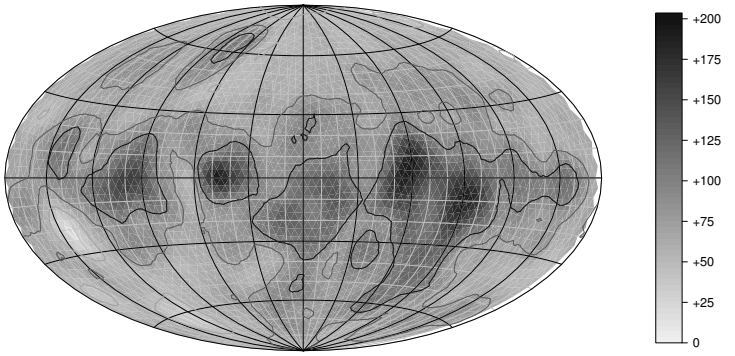

Figure 6: Pointwise 95% credible set width for the final model. Note that the widest intervals occur near the plane of the Milky Way, which also is where the most extreme RM values occur. Contour lines correspond to credible set widths of 25 (white), 50 (light gray), 75 (dark gray), and 100 (black) rad/m^2. As before, $(l, b) = (0, 0)$ points to the Galactic Center.

all observations.

The approximately stationary model for the RM field $z(\mathbf{s})$ is fairly simplistic. We also fitted more complicated models – one with a mean, and one that was the sum of two processes, one with a wide smoothing kernel, and one with a narrow kernel. Upon adding a constant mean term to the model (using a $N(0, 400^2)$ prior distribution), we found very little difference between the resulting model fit and our original one. This is not surprising since the posterior global mean for $z(\mathbf{s})$ was estimated to be not significantly different from 0. The additional computational burden of including this additional parameter (μ) was substantial since the single site MCMC scheme becomes far less efficient in exploring the posterior with the additional mean parameter. The mean parameter trades off with the knot values resulting in a chain that mixes much more slowly than with our original mean-free formulation.

We were concerned that our basic model might be tuning itself to fine scale features of the RM field, while neglecting important global features. To address this concern, we developed a multi-scale prior for $z(\mathbf{s})$ on the sphere which is analogous to the formulation for $\mathbb{R}^d$ given in Higdon (2002). Here $z(\mathbf{s})$ is represented as the sum of two independent process convolution fields

$$z(\mathbf{s}) = z^c(\mathbf{s}) + z^f(\mathbf{s})$$

where the coarse field $z^c(\mathbf{s})$ is constructed using a wide smoothing kernel, and the fine field $z^f(\mathbf{s})$ is constructed using a narrow smoothing kernel. We constrained the narrow kernel SD to be less than that of the wide kernel. The resulting posterior puts all of the variation in the fine process, while the coarse process is essentially 0. The kernel width for this fine process matched our estimate using our original model. We tried a number

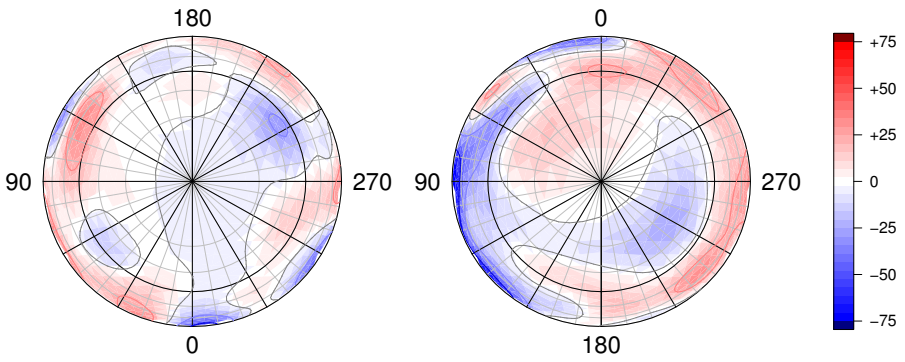

Figure 7: Posterior plots for the final model. Left panel: North Galactic Cap (NGC), in which $b \geq 30^o$N. Right panel: South Galactic Cap (SGC); $b \leq 30^0$S. The contour lines are the same as in Figure 5. The rotation measures at the highest latitudes tend to be much smaller than those near the plane of the galaxy.

of different initializations; still, each time the MCMC gave the same results. Hence we were satisfied that the single process convolution adequately models the RM field.

7 Summary and conclusions

In conclusion we emphasize that this analysis gives a principled, statistical estimate of the RM field in the near galactic sky while accounting for contamination from extra-galactic sources, using the most complete set of RM measurements to date. This enables us to start to distinguish certain near and distant features of the universe. The large number of observations in this current collection of measurements has forced us to make modeling choices that adequately address the specific features of this application while still making a fully Bayesian estimation analysis feasible via MCMC. The process convolution model for the RM field allows the MCMC scheme to avoid computationally costly inverses of large matrices. It also facilitates the inclusion of non-standard error models in the analysis.

The computational burden of our approach is moderate, considering the size of our data set, which has $N = 1566$ data points. A typical MCMC run of 250,000 iterations, including posterior prediction for a final smoothed map, required slightly over 3 days on a desktop PC. Had we gone with an approach based on covariance functions, the factorizations of 1566×1566 matrices at both the MCMC and prediction steps would have stopped us in our tracks. In particular, given the mixture model used for the errors, in order to sample the ω's, we would have needed as many as 1566 factorizations per MCMC iteration. Since new surveys will likely result in tens of thousands of RM observations, it is crucial that our approach accommodate large numbers of observations,

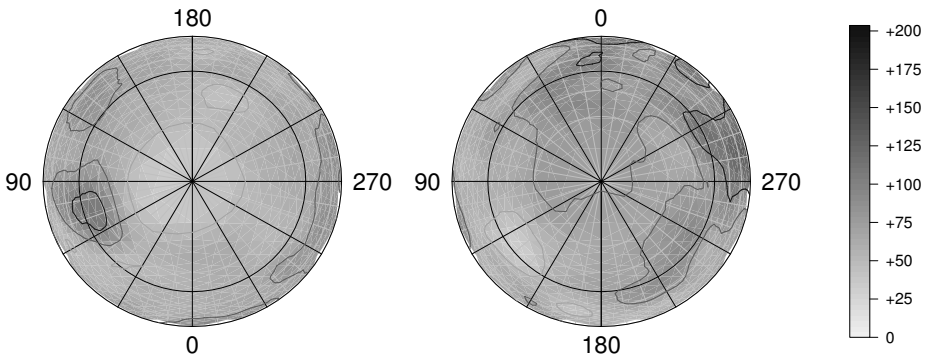

Figure 8: Credible set widths. As in Figure 7, the NGC is on the left and the SGC is on the right. The contour lines are the same as in Figure 6.

while automatically accounting for outliers.

We expect future investigations to focus in two opposite directions – mapping out the magnetic field of our local galaxy, and investigating magnetic fields in the distant universe. Within the Milky Way, more localized RM's can be obtained from pulsars. Since the spatial location of these objects can be estimated independently of their RM values, these observations lend additional spatial information regarding the magnetic fields within the Milky Way. Combining these pulsar RM's with future large scale RM surveys may give sufficient information for a 3-D reconstruction of the Milky Way's magnetic field. Rotation measures also serve as probes for identifying magnetic fields in the more distant universe. In this case, the effect of the Milky Way on RM's obtained from very distant sources is a nuisance, perturbing the signal from distant universe. Improving our estimate of the local RM sky will enhance our ability to detect magnetic fields in the distant universe in the future.

References

Frick, P., Stepanov, R., Shukurov, A., and Sokoloff, D. (2001). "Structures in the rotation measure sky." *Monthly Notices of the Royal Astronomical Society*, 325(2): 649–664. 666

Gaspari, G. and Cohn, S. (1999). "Construction of correlation functions in two and three dimensions." *Quarterly Journal of the Royal Meteorological Society*, 125: 723–757. 666

Gneiting, T. (1999). "Correlation functions for atmospheric data analysis." *Quarterly Journal of the Royal Meteorological Society*, 125: 2449–2464. 666

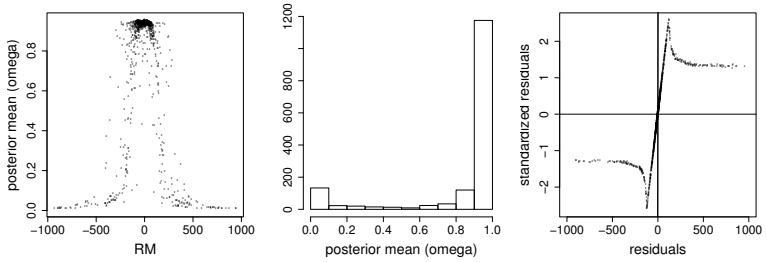

Figure 9: Diagnostic plots. The left panel shows the posterior mean of the ω's versus the RM's in the event that $|RM| < 1000$. In general, observations with small $|RM|$ have mean(ω) close to one, i.e. are likely uncontaminated according to our model; conversely, observations with large $|RM|$ tend to have mean(ω) close to zero. The histogram in the middle panel is of the posterior mean of the ω_i's. Note the bimodal distribution; a fair number of the observations are being flagged as contaminated, and many of these have mid-range RM values. An empirical influence function appears in the right panel, showing standardized residuals versus residuals. As expected from the model, contaminated observations correspond to large residuals, which are in turn downweighted by small $\hat{\omega}_i$'s.

Higdon, D. M. (2002). "Space and space-time modeling using process convolutions." In Anderson, C., Barnett, V., Chatwin, P. C., and El-Shaarawi, A. H. (eds.), *Quantitative Methods for Current Environmental Issues*, 37–56. London: Springer Verlag. 670

Oh, H. and Li, T. (2004). "Estimation of global temperature fields from scattered observations by a spherical-wavelet-based spatially adaptive method." *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 66(1): 221–238. 666

Simard-Normandin, M. and Kronberg, P. (1980). "Rotation Measures and the Galactic Magnetic Field." *The Astrophysical Journal*, 242: 74–94. 666

Simard-Normandin, M., Kronberg, P., and Button, S. (1981). "The Faraday rotation measures of extragalactic radio sources." *The Astrophysical Journal Supplement series*, 45: 97–111. 667

Weber, R. and Talkner, P. (1993). "Some remarks on spatial correlation function models." *Monthly Weather Review*. 666

Yaglom, A. M. (1987). *Correlation Theory of Stationary and Related Random Functions, volume I*. Springer-Verlag. P. 385. 666

Acknowledgments

The work of the first author was supported by DOE and UAF. The work of the second author was supported by DOE. The work of the third author was supported by DOE and the Natural

Sciences and Engineering Research Council of Canada (NSERC). The authors are grateful for helpful comments provided by the reviewers.