BAYESIAN STATISTICS 7, pp. 000–000
J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid,
D. Heckerman, A. F. M. Smith and M. West (Eds.)

A Nonparametric Bayesian Approach to Inverse Problems

ROBERT L. WOLPERT
Duke University, USA
wolpert@stat.duke.edu

KATJA ICKSTADT
Universität Dortmund, Germany
ickstadt@statistik.uni-dortmund.de

MARTIN B. HANSEN
Aalborg Universitet, Denmark
mbh@math.auc.dk

SUMMARY

We propose a new method for making inference about an unknown measure $\Gamma(d\lambda)$ upon observing some values of the Fredholm integral $g(\omega) = \int k(\omega, \lambda)\Gamma(d\lambda)$ of a known kernel $k(\omega, \lambda)$, using Lévy random fields as Bayesian prior distributions for modeling uncertainty about $\Gamma(d\lambda)$. Inference is based on simulation-based MCMC methods. The method is illustrated with a problem in polymer chemistry.

Keywords: GAMMA PROCESS; LÉVY PROCESS; POLYMER; RANDOM FIELD; REVERSIBLE JUMP MCMC; RHEOLOGY.

1. INTRODUCTION

Fredholm (1900) initiated the formal study of integral equations of the first kind, in which we try to impute an unknown measure $\Gamma(d\lambda)$ from finitely many observed values of the integrals

$$G(\omega_i) = \int_\Lambda k(\omega_i, \lambda)\,\Gamma(d\lambda) \tag{1}$$

of a known kernel $k(\omega, \lambda)$. The problem is difficult in part because the integral operator $K:\ \Gamma \mapsto G$ is *smoothing*, making the "inverse problem" $K^{-1}: G \mapsto \Gamma$ ill-posed in the sense that small changes in G may be associated with large changes in Γ.

The most common approaches to solving Equation (1) for the unknown Γ begin by approximating this infinite-dimensional continuous problem with the finite-dimensional discrete one

$$G_i = \sum_{j\in J} k_{ij}\,\Gamma_j. \tag{2}$$

The approximate solution of Equation (2) is available from the normal equations as $\Gamma \approx [K'K]^{-1}K'G$ (see Kirsch (1996) for a discussion of such discretization methods and of the numerical obstacles that arise in trying to solve the resulting linear systems). Typically the matrix $[K'K]$ is ill-conditioned and so Equation (2) remains ill-posed—if J is small there are no solutions, while if J is sufficiently large there are infinitely many which differ wildly. Commonly this is addressed through the inclusion of some form of "roughness penalty" (see for example the *method of regularization* of Tikhonov (1963)).

2. A NEW BAYESIAN NONPARAMETRIC APPROACH

Our approach is to treat the solution of Equation (1) as a Bayesian statistical inference problem, that of estimating the uncertain element Γ of the space $\mathcal{M}_+(\Lambda)$ of positive measures on a set Λ upon observing, perhaps with error, the quantities $G_i \approx G(\omega_i) \in \mathcal{G}$ at some finite set of points $\{\omega_i\}_{i\in I} \subset \Omega$. To complete the Bayesian model specification we must select a *prior distribution* $\pi(d\Gamma)$ on $\mathcal{M}_+(\Lambda)$, making Γ a random measure, and we must select a *measurement error model* for G_i given $g_i \equiv G(\omega_i)$, leading to a likelihood function $L(\Gamma)$.

In many applications (including ours in Section (3) below) localization arguments suggest that the uncertain positive measures $\Gamma(A)$ and $\Gamma(B)$ assigned to disjoint sets $A,\, B \subset \Lambda$ may be regarded as stochastically independent *a priori*. Under mild regularity conditions this leads to a Lévy-Khinchine-like representation for stochastic integrals $\Gamma[\phi] \equiv \int_\Lambda \phi(\lambda)\,\Gamma(d\lambda)$ of measurable functions $\phi : \Lambda \to \mathbb{R}$ of the form

$$\log \mathrm{E}\left[e^{i\Gamma[\phi]}\right] = \iint\limits_{\mathbb{R}_+\times\Lambda} \left(e^{iu\phi(\lambda)} - 1\right)\, \nu(du\, d\lambda)$$

for some positive measure $\nu(du\, d\lambda)$ on $\mathbb{R}_+\times\Lambda$ satisfying the integrability condition $\iint_{\mathbb{R}_+\times K}(1\wedge u)\nu(du\, d\lambda) < \infty$ for compact $K \subset \Lambda$. Jacod and Shiryaev (1987, Chapter II, §4c) give details about this generalization of the usual Lévy-Khinchine formula to non-stationary processes and random fields. The Inverse Lévy Measure (ILM) algorithm of Wolpert and Ickstadt (1998a, 1998b) offers an explicit construction of such random fields, predicated on the representation

$$\Gamma[\phi] = \iint\limits_{\mathbb{R}_+\times\Lambda} u\phi(\lambda)\, H(du\, d\lambda) = \sum_{j\in J} u_j \phi(\lambda_j)$$

of $\Gamma(d\lambda)$ in terms of a Poisson measure $H(du\, d\lambda)$ on $\mathbb{R}_+\times\Lambda$ with Lévy mean measure $\mathrm{E}[H(du\, d\lambda)] = \nu(du\, d\lambda)$; here $\{u_j, \lambda_j\}_{j\in J}$ represents an instance of the (at most countable) random support of $H(du\, d\lambda)$.

If $\nu(\mathbb{R}_+\times\Lambda) < \infty$ then $H(du\, d\lambda)$ (and hence $\Gamma(d\lambda)$) will have only finitely many points of support. If $\nu(du\, d\lambda)$ has a density function $\nu(u,\, \lambda)$ with respect to some finite reference measure $m(du\, d\lambda)$, then Γ will have a probability density function

$$\pi(\Gamma) = \left[\prod_{j\in J} \nu(u_j,\, \lambda_j)\right]\, e^{m\left(\mathbb{R}_+\times\Lambda\right) - \nu\left(\mathbb{R}_+\times\Lambda\right)}$$

with respect to the random field with Lévy measure m.

If recorded measurements $G_i \in \mathcal{G}$ may be taken to differ only by independent measurement errors from the true values

$$g_i \equiv G(\omega_i) = \int_\Lambda k(\omega_i, \lambda)\, \Gamma(d\lambda) = \sum_{j\in J} k(\omega_i, \lambda_j)\, u_j,$$

with probability density functions $f(G_i \mid g_i)$, then the likelihood function is simply $L(\Gamma) = \prod_{i\in I} f(G_i \mid g_i)$ and by Bayes' theorem the posterior distribution for Γ has a probability density function

$$\pi(\Gamma \mid \{\vec{G}_i\}_{i\in I}) \propto \Big[\prod_{j\in J} \nu(u_j,\, \lambda_j)\Big] \Big[\prod_{i\in I} f(G_i \mid g_i)\Big]\, e^{m\left(\mathbb{R}_+\times\Lambda\right)-\nu\left(\mathbb{R}_+\times\Lambda\right)}. \tag{3}$$

This posterior distribution forms the basis for statistical inference about the solution Γ of the inverse problem in Equation (1). Features of Γ that are well-determined by the data (or the prior) will show little posterior variation, while the system's ill-posedness will be expressed in wide posterior variability of features that are undetermined by the prior and data.

The role of the Lévy prior distribution is analogous to that of the roughness penalty in conventional regularization methods, resolving features left unspecified by the data, but with the important benefit of easy interpretability and coherence.

3. A RHEOLOGY EXAMPLE

A Newtonian fluid suspended between two horizontal plates exhibits *viscous* behavior: a tangential force applied to one of the plates leads to a velocity gradient in the fluid proportional to the force per unit area, $\tau = \eta\frac{\partial v}{\partial y}$ (here τ represents the stress, or force per unit area, and v the horizontal velocity at any height y). The proportionality constant in this linear relationship, the viscosity η, is measured in pascal-seconds ("pascal" is the SI unit for pressure or stress, equal to one newton per square meter or one kilogram per meter per second squared, so one pound per square inch (p.s.i.) is about 6.89 kPa). The viscosity of familiar fluids ranges from about 10^{-5} Pa·s for air to 10^{-3} Pa·s for water to 1 Pa·s for glycerine.

A tangential force applied to one side of a springy or *elastic* substance induces a proportional deformation, $\tau = G\gamma$, where (unitless) γ represents the relative length change induced and the proportionality constant G (measured in Pa) is called the elastic modulus.

Polymers are gooey non-Newtonian compounds whose behavior lies in between highly viscous Newtonian fluids, with viscosities in the range of 10^2–10^5 Pa·s, and elastic compounds. Boltzmann (1876) had the idea of modeling these *visco-elastic* compounds by introducing time-dependence to the stress $\tau(t)$, elastic modulus $G(t)$, and deformation $\gamma(t)$ (with time-derivative $\dot\gamma(t)$), and relating all of them by a time-dependent extension of the elasticity equation, $\tau(t) = \int_{-\infty}^{t} G(t-s)\dot\gamma(s)\,ds$ or, upon changing variables,

$$\tau(t) = \int_0^\infty G(\omega)\,\dot\gamma(t-\omega)\,d\omega. \tag{4}$$

Boltzmann took G to have a "fading memory," i.e., to be completely monotonicly decreasing and so, by Bernstein's Theorem (see Feller (1971), §XIII.4) representable in the form

$$G(\omega) = \int_0^\infty e^{-\omega/\lambda}\,\Gamma(d\lambda)/\lambda \tag{5}$$

as the Laplace transform of some positive measure $\Gamma(d\lambda)$ on $\mathbb{R}_+$ called the *relaxation spectrum*. For infinitesimal stresses the behavior is approximately that of a viscous fluid with *zero-shear viscosity* $\nu_0 = \int G(\omega)\,d\omega = \Gamma(\mathbb{R}_+)$. For periodic strains $\gamma(\omega) = \gamma_0\sin(\omega t)$ the solution to Equations (4,5) is available in closed form:

$$\tau(\omega) = \gamma_0 \Big[G'(\omega)\sin(\omega t) + G''(\omega)\cos(\omega t) \Big]$$

where the elastic, in-phase, energy-conserving *storage modulus* and the viscous, out-of-phase, energy-dissipating *loss modulus* are given (respectively) by

$$G'(\omega) \equiv \int_0^\infty \frac{\omega^2\lambda}{1+\omega^2\lambda^2}\,\Gamma(d\lambda) \qquad\qquad G''(\omega) \equiv \int_0^\infty \frac{\omega}{1+\omega^2\lambda^2}\,\Gamma(d\lambda). \tag{6}$$

For small periodic strains both $G'(\omega)$ and $G''(\omega)$ can be measured experimentally using an oscillatory shear rheometer; our analysis below is based on measurements of Berger (1988), reproduced in Table (1) and plotted on both linear and logarithmic scale in Figure (1).

Table 1. *Experimental measurements of storage modulus $G'(\omega_i)$ and loss modulus $G''(\omega_i)$, both in Pa, at various frequencies ω_i (s^{-1}) for a polybutadiene melt at 23° C, from Berger (1988).*

ω	$G'(\omega)$	$G''(\omega)$	ω	$G'(\omega)$	$G''(\omega)$
2.493×10^0	2052	34526	7.680×10^1	432105	359952
3.670×10^0	4156	50445	1.144×10^2	534678	343388
5.373×10^0	8847	73294	1.654×10^2	619214	327629
7.864×10^0	18834	105329	2.433×10^2	701325	307419
1.144×10^1	37737	149699	3.539×10^2	772708	290069
1.695×10^1	74730	206936	5.238×10^2	841878	278292
2.451×10^1	136257	266220	7.529×10^2	897344	264055
3.608×10^1	223611	313420	1.114×10^3	956262	249131
5.218×10^1	324937	345321			

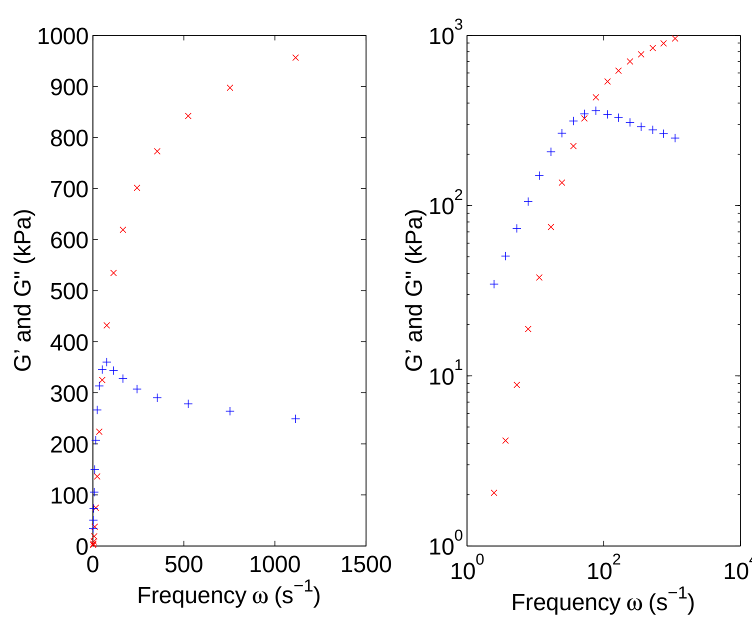

Figure 1. *Measured $G'(\omega_i)$ (×) and $G''(\omega_i)$ (+) from Berger (1988)*

Equation (6) is a two-dimensional Fredholm inverse problem of the form of Equation (1) with $\omega \in \Omega \equiv \mathbb{R}_+$, $G(\omega) \equiv [G'(\omega), G''(\omega)] \in \mathcal{G} \equiv \mathbb{R}_+^2$, $\Lambda = \mathbb{R}_+$, $k : \Omega \times \Lambda \to \mathcal{G}$ with $k(\omega, \lambda) = [\omega^2\lambda,\ \omega](1+\omega^2\lambda^2)^{-1}$, and $\Gamma(d\lambda)$ unknown. To proceed with the Bayesian modeling approach of Section (2) we will need to specify a measurement-error model for $\{G_i = G(\omega_i)\}_{i\in I}$ and a prior distribution for Γ.

Berger's choice of sampling frequencies ω_i and the general form of both $G'(\omega)$ and $G''(\omega)$ in Figure (1) suggest that the logarithmic transformation simplifies the relation and may stabilize the variance, leading to our choice of a bivariate lognormal measurement error model

$$\log G_i' \sim \mathrm{N}(\log g_i', \sigma^2) \qquad \log G_i'' \sim \mathrm{N}(\log g_i'', \sigma^2)$$

or, more succinctly, $\vec{G}_i \sim \mathrm{LN}(\vec{g}_i, \sigma^2 I_2)$ for some fixed $\sigma^2 > 0$ with $\vec{G}_i \equiv [G_i', G_i'']$. In the absence of repeated measurements which might let us validate the log-normal model and help select a value for σ^2, we based our choices on an exploratory study of the residuals $\{(\vec{G}_i - \hat{G}_i)\}_{i\in I}$ from the best fit $\{\hat{G}_i\}_{i\in I}$ of Equation (6) to Berger's data.

For a prior distribution we have chosen the Gamma random field with Lévy measure

$$\nu(du\, d\lambda) = \alpha\, u^{-1}\lambda^{-1}\, e^{-\beta u}\, du\, d\lambda$$

with uniform shape measure (on a logarithmic scale) on the interval $\Lambda \equiv (\lambda_-, \lambda^+)$ with $\lambda_- = e^{-7}$, $\lambda^+ = e^{-1}$, a convenient approximation to the interval $1/\omega_{\max} < \lambda < 1/\omega_{\min}$ that the localization principle argument of Davies and Anderssen (1997) suggests is the widest range on which we can hope to learn from the data about $\Gamma(d\lambda)$. The localization principle is also the basis for our choice of an independent-increment prior distribution.

With our choice of prior distribution the zero-shear viscosity $\int_0^\infty G(t)\, dt = \Gamma(\mathbb{R}_+)$ has a $\mathrm{Ga}(\alpha \log(\lambda^+/\lambda_-), \beta)$ distribution. Values for the parameters α, β were chosen to ensure that the mean and variance of $\Gamma(\mathbb{R}_+)$ would be approximately $6\alpha/\beta \approx 13,500$ and $6\alpha/\beta^2 \approx 50^2$, as suggested by the empirical evidence of Berger's observations (the zero-shear viscosity is also the slope $dG''/d\omega$ at $\omega = 0$). See Wolpert, Hansen and Ickstadt (to appear) for a wider range of prior distributions and measurement error models and see Anderssen and Hansen (to appear) for a more specific rheological treatment and discussion of the sampling localization theorem in the context of relaxation spectral analysis.

In summary, the complete Bayesian model specification is:

$$\begin{array}{rrcl}
\text{Data:} & \vec{G}_i & \sim & \mathrm{LN}(\vec{g}_i, \sigma^2 I_2) \text{ in } \mathcal{G} = \mathbb{R}_+^2 \\
\text{Model:} & \vec{g}_i & = & \left[\int_0^\infty \frac{\omega_i^2\lambda}{1+\omega_i^2\lambda^2}\,\Gamma(d\lambda),\ \int_0^\infty \frac{\omega_i}{1+\omega_i^2\lambda^2}\,\Gamma(d\lambda)\right] \\
\text{Prior:} & \Gamma(d\lambda) & \sim & \mathrm{Levy}(\nu(du,\, d\lambda)), \qquad u \in \mathbb{R}_+,\ \lambda \in \Lambda \equiv (\lambda_-, \lambda^+) \\
 & & = & \mathrm{Ga}(\alpha\lambda^{-1}\, d\lambda, \beta)
\end{array}$$

and our goal is to estimate $\Gamma(d\lambda)$ upon observing $\vec{G}_i \equiv [G'(\omega_i),\ G''(\omega_i)] \approx \vec{g}_i$ for several frequencies ω_i.

4. COMPUTATIONS

Our choice of the Gamma Lévy prior features an infinite Lévy measure $\nu(\mathbb{R}_+\times\Lambda)=\infty$ and, therefore, almost surely there are infinitely many terms in the representation

$$\vec{g}_i=\sum_{j\in J}\,[\omega_i^2\lambda_j,\omega_i]\,(1+\omega_i^2\lambda_j^2)^{-1}\,u_j.$$

For any $\epsilon>0$ the number $M_\epsilon=|J_\epsilon|$ of points with $u_j>\epsilon$ (indexed by $J_\epsilon=\{j\in J:\ u_j>\epsilon\}$) is a random variable whose prior distribution is Poisson with mean

$$\mathrm{E}[M_\epsilon]=\int_{\lambda_-}^{\lambda^+}\int_\epsilon^\infty \alpha\,u^{-1}\lambda^{-1}\,e^{-\beta u}\,du\,d\lambda=\alpha E_1(\beta\epsilon)\log\frac{\lambda^+}{\lambda_-}<\infty,$$

where $E_1(x)\equiv\int_x^\infty t^{-1}e^{-t}\,dt$ is the exponential integral function (Abramowitz and Stegun (1964), §5.1). The expected total mass $\sum_{j\in J_\epsilon}\{u_j:\ u_j\le\epsilon\}$ of all points (u_j,λ_j) with $u_j\le\epsilon$ is only

$$\mathrm{E}\Big[H\Big((0,\epsilon]\times\Lambda\Big)\Big]=\int_{\lambda_-}^{\lambda^+}\int_0^\epsilon\ \alpha\,\lambda^{-1}\,e^{-\beta u}\,du\,d\lambda=\alpha\beta^{-1}\log\frac{\lambda^+}{\lambda_-}(1-e^{-\beta\epsilon}),$$

a fraction $(1-e^{-\beta\epsilon})$ of the total prior expected mass $\mathrm{E}[\Gamma(\mathbb{R}_+)]=\alpha\beta^{-1}\log\frac{\lambda^+}{\lambda_-}$. In our implementation we select ϵ small enough that this represents 0.5% of the total mass, and include only the mass points $u_j\in U_\epsilon\equiv(\epsilon,\infty)$. The posterior distribution of M_ϵ is not Poisson, of course.

The space Θ of configurations we model may be represented as the countable union of Cartesian powers

$$\Theta=\cup_{M=0}^\infty(\mathbb{R}_+\times\Lambda)^M,$$

where the measure $\Gamma_\theta(d\lambda)\in\mathcal{M}_+(\Lambda)$ associated with index $\theta\in\Theta$ is

$$\Gamma_\theta(d\lambda)=\sum_{j=1}^M u_j\delta_{\lambda_j}(d\lambda),$$

the sum of M point masses of magnitudes $u_j\in\mathbb{R}_+$ at points $\lambda_j\in\Lambda$. From Equation (3) we can compute the probability density function of the posterior distribution of $\theta\in\Theta$ (with respect to the Poisson random measure on Θ with rate $m(du\,d\lambda)$) upon observing the $N\equiv|I|$ vectors $\vec{G}_i=[G_i',\,G_i'']$:

$$\begin{aligned}\log\pi(\theta\,|\,\{\vec{G}_i\}_{i\in I})&=c+\sum_{j\in J}\log\nu(u_j,\,\lambda_j)\ +\sum_{i\in I}\log f(\vec{G}_i\,|\,g_i)\\&=c+M\log\alpha-\sum_{j\in J}\log u_j\lambda_j-\beta\sum_{j\in J}u_j\\&\qquad-N\log 2\pi\sigma^2-\frac{1}{2\sigma^2}\sum_{i\in I}\Big[\log^2\frac{G_i'}{g_i'}+\log^2\frac{G_i''}{g_i''}\Big]\end{aligned}$$

where $c=m(U_\epsilon\times\Lambda)-\nu(U_\epsilon\times\Lambda)$ does not depend on θ.

To implement the Metropolis-Hastings version of the Markov Chain Monte Carlo method (see, e.g., Tierney (1994)) we must select an irreducible transition probability distribution $Q(d\theta^* \mid \theta)$ on Θ. Our choice reflects our intention to *model* uncertainty about $\Gamma(d\lambda)$ using the Gamma random field with its infinite Lévy measure, even though our *implementation* permits us to simulate only the finite number M_ϵ Gamma masses of magnitude $u_j > \epsilon$.

The heuristic behind our proposal distribution is to imagine infinitely many particles at locations $(u_j, \lambda_j) \in \mathbb{R}_+ \times \Lambda$ all undergoing simultaneous ergodic diffusion with the posterior as a stationary distribution. At any given time only finitely many points M_ϵ will lie above the line $u_j > \epsilon$; now and then one of these will diffuse below that line, causing M_ϵ to fall by one, while now and then one of the infinitely many points below the line will rise above it, increasing M_ϵ by one. Sampled at discrete times this would be a random walk similar to that we propose below, with three types of steps—those with M_ϵ unchanged ($\Delta M_\epsilon = 0$) and those where M_ϵ increases or decreases by one ($\Delta M_\epsilon = \pm 1$). For fixed logarithmic step size δ (we use $\delta = 0.25$), re-entry probability $0 < p < 1$ (we use $p = 0.01$), and re-entry distribution with density $f(u, \lambda)$ on $(0, \infty) \times \Lambda$ (see below), the move proposals are:

- $\Delta M_\epsilon = \;\; 0$ With probability $1-p$, choose j uniformly from the integers $1{:}M$ and propose the lognormal step $u_* = u_j \exp(\delta Z_1)$, $\lambda_* = \lambda_j \exp(\delta Z_2)$ with $Z_1, Z_2 \sim \mathrm{N}(0,1)$. Reflect at the boundaries if necessary to ensure that $\lambda_* \in \Lambda = (\lambda_-, \lambda^+)$. If $u_* > \epsilon$, then the proposed new M_ϵ remains unchanged. Otherwise,
- $\Delta M_\epsilon = -1$ If $u_* \le \epsilon$ above, remove j from J and decrease M_ϵ by one; the resulting proposal is to delete the single point (u_j, λ_j) from the ensemble θ.
- $\Delta M_\epsilon = +1$ With probability p, increment M_ϵ by one and introduce a new index M to J and draw a new mass point $(u_M, \lambda_M) \sim f(u_M, \lambda_M)$ from the re-entry distribution.

For our re-entry distribution we draw λ from the uniform distribution on a logarithmic scale on Λ and, independently, draw u from the exponential distribution with mean $\mu = \mathrm{E}[H]/\mathrm{E}[M_\epsilon]$, conditioned to satisfy $u > \epsilon$, giving $f(u, \lambda) = (\lambda \mu \log \frac{\lambda^+}{\lambda_-})^{-1} \exp\left(\frac{\epsilon - u}{\mu}\right)$ on $U_\epsilon \times \Lambda$.

The conditional p.d.f. $Q(\theta_* \mid \theta)$ of the proposal transition probability distribution $Q(d\theta_* \mid \theta)$ (again, with respect to the $\mathrm{Po}(m(du\, d\lambda))$ Poisson random measure) is easily calculated from this prescription. Finally the MCMC algorithm proceeds as follows:

0. Initialize $t = 0$, $M^{(0)} \sim \mathrm{Po}(\mathrm{E}[M_\epsilon])$, $J = \{1, ..., M^{(0)}\}$, $\{(u_j, \lambda_j)\}_{j \in J} \sim f(u, \lambda)$, and set $\theta^{(0)} = \{(u_j, \lambda_j)\}_{j \in J}$.
1. Find a proposed new point $\theta_* \sim Q(d\theta_* \mid \theta^{(t)})$ and compute the Metropolis-Hastings log acceptance probability

$$\zeta^{(t+1)} = \log \pi(\theta_* \mid \{\vec{G}_i\}) + \log Q(\theta^{(t)} \mid \theta_*) - \log \pi(\theta^{(t)} \mid \{\vec{G}_i\}) - \log Q(\theta_* \mid \theta^{(t)})$$

2. Generate a standard exponential random variable $Z \sim \mathrm{Ex}(1)$ and set

$$\theta^{(t+1)} = \begin{cases} \theta_* & \text{if } Z + \zeta^{(t+1)} \ge 0 \\ \theta^{(t)} & \text{if } Z + \zeta^{(t+1)} < 0 \end{cases}.$$

   Adjust $M^{(t)}$ and J if necessary. Increment $t \leftarrow t + 1$.

3. Periodically (e.g. at 100 evenly-spaced times following "burn-in") store $\theta^{(t)}$.
4. If $t < TMAX$, repeat steps 1–4.

The parameters δ, μ, etc. are adjusted in trial runs to ensure that the rate of accepting proposed moves is approximately 20-50%. On contemporary small computers (2GHz dual-processor Unix workstations), our MatLab implementation can complete approximately two million steps per hour.

5. RESULTS

Figure (2) shows a representation of the posterior distribution of the model's predictions $G(\omega) = [G'(\omega),\ G''(\omega)]$, with the measurements ($\times$ for $G'(\omega_i)$, $+$ for $G''(\omega_i)$); the figure shows the 25%, 50% and and 75% percentile bands along with 100 MCMC iterations (equally spaced from among one million). These curves lie so close together that it is difficult to distinguish them in the plot, showing that there is little posterior uncertainty about $G(\omega)$.

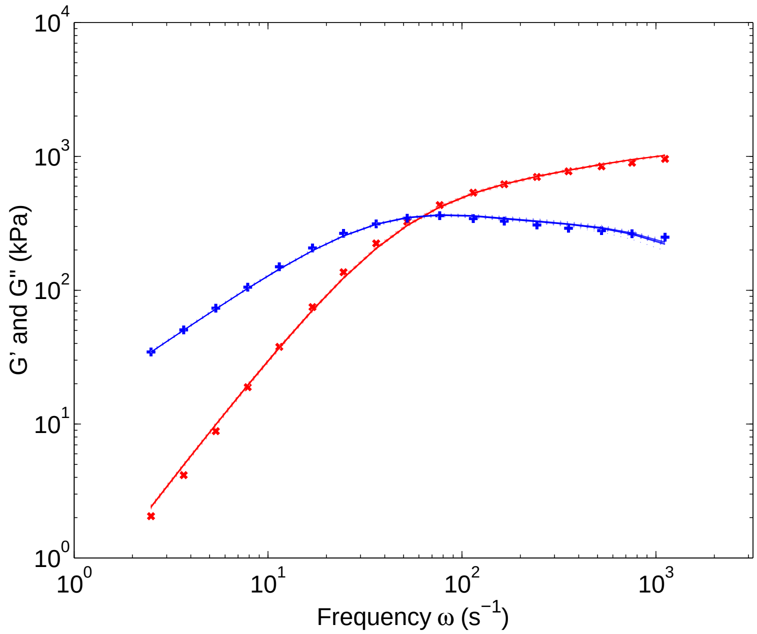

Figure 2. *Posterior distributions for $G'(\omega_i)$ $(\times)$ and $G''(\omega_i)$ $(+)$.*

Figure (3) shows the prior (dotted line) and posterior means of the spectral density $\Gamma(d\lambda)/d\lambda$, and Figure (4) the prior (again, dotted) and posterior cumulative spectral distribution $H[(0,\lambda)]$ with the 25%, 50% and and 75% posterior percentiles and 100 MCMC iterations to illuminate the distribution. Evidently the spectral density is unimodal (or, if not, has no strong second mode), centered at about $\lambda \approx 0.020s$, with half of its mass in the interval $[0.015, 0.027]$ and 90% in the interval $[0.004, 0.086]$.

Figure (5) displays the posterior distribution of M_ϵ, the number of mass points for the spectral measure of magnitude $u > \epsilon$, with the posterior mean $\mathrm{E}[M_\epsilon \mid \{\vec{G}_i\}_{i\in I}] = 303.9$ indicated with a vertical line (the prior distribution was $M_\epsilon \sim \mathrm{Po}(118.0)$).

Figure (6) shows a single state $\theta^{(t)}$ from the simulation (the final one, with $t = 1,000,000$). This step features $M_\epsilon = 290$ mass points (u_j, λ_j) (dots represent recently-moved points, stars represent recently-added points from the re-entry distribution). Following the 100,000-step burn-in phase in this simulation run, M_ϵ ranged from a

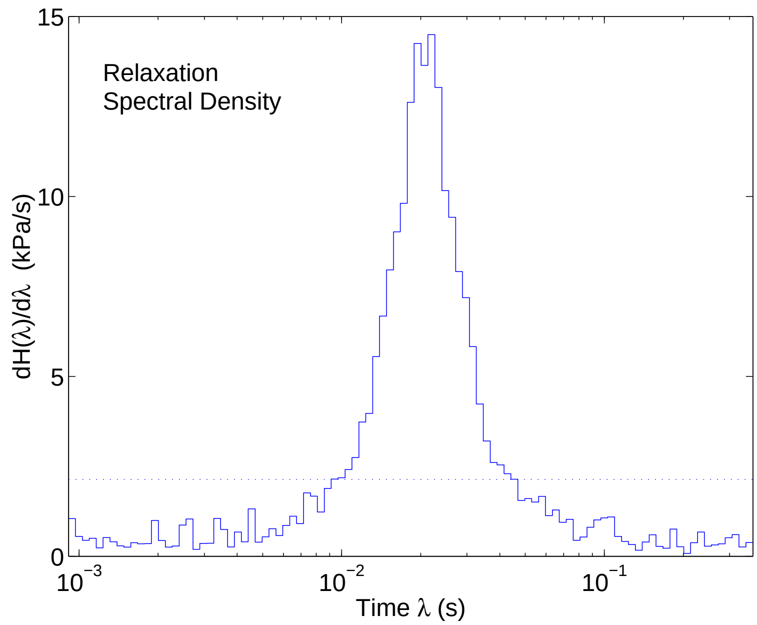

Figure 3. *Prior and Posterior Spectral Density*

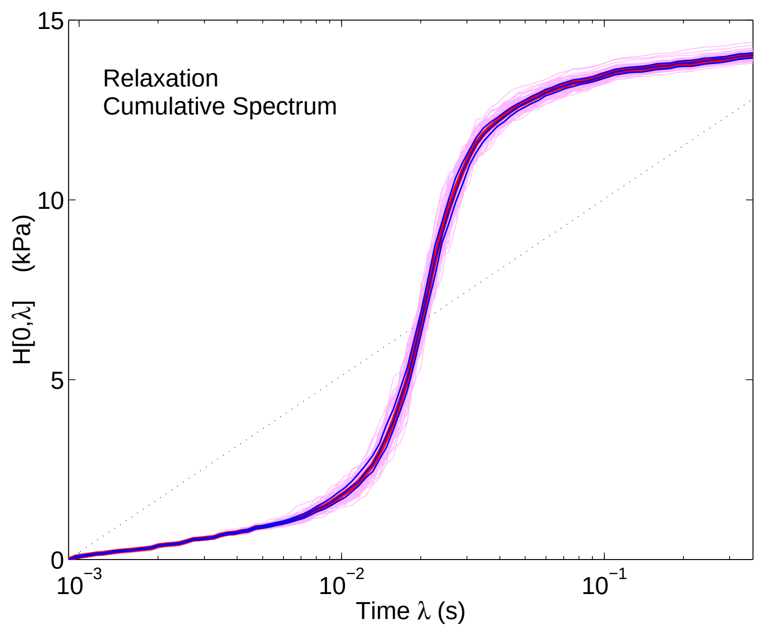

Figure 4. *Prior and Posterior Cumulative Spectrum*

minimum of 247 to a maximum of 458, with a mean of 303.9. Magnitudes u_j are represented on a logarithmic scale, so a very large fraction of the mass is represented by the largest few points; the 30 points above $u > 100$ hold 50% of the total mass, and the 186 points above $u > 10$ hold about 95%. Here $\epsilon = 2.57$ was chosen to ensure that 99.5% of the prior mass exceeded ϵ.

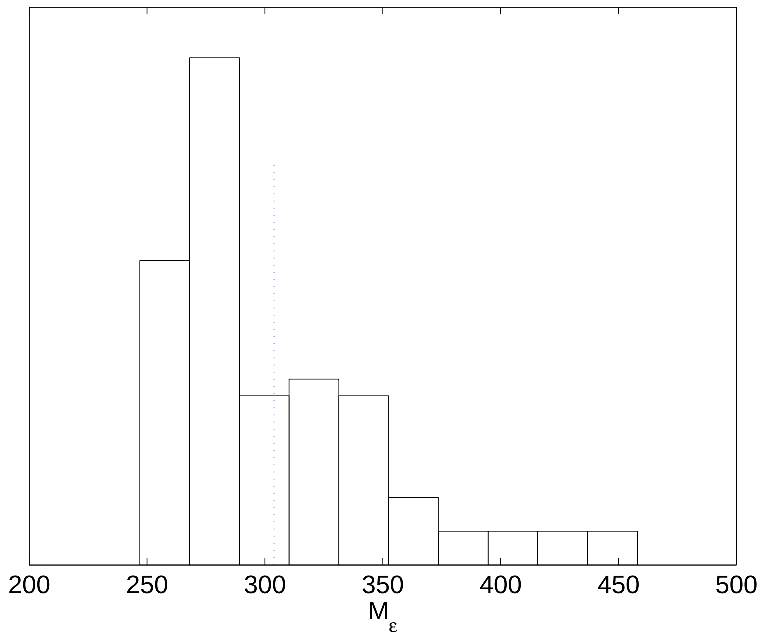

Figure 5. *Posterior Distribution of M_ϵ*

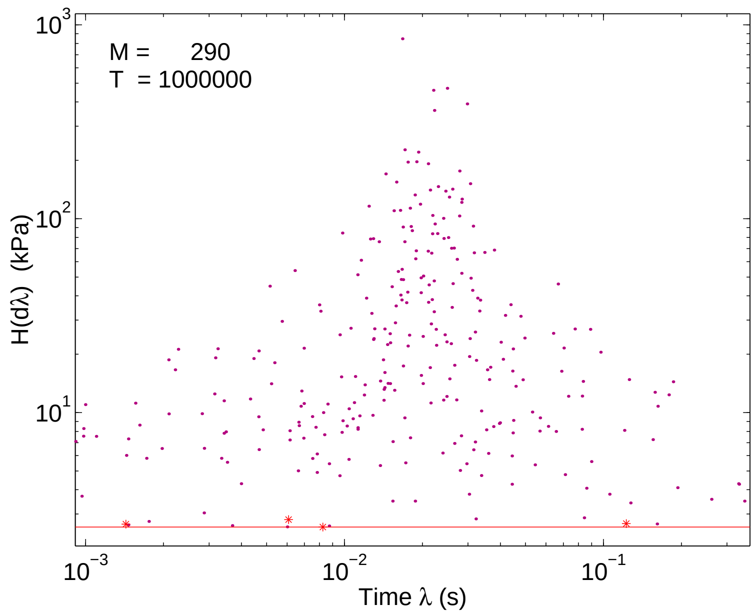

Figure 6. *A single configuration $\theta^{(t)}$ from the simulation*

6. DISCUSSION

The nonparametric Bayesian approach to inference in inverse problems allows us to model explicitly our prior beliefs or expert understanding about features of the solution $\Gamma(d\lambda)$, and to represent honestly and coherently whatever uncertainty remains about these features following the observation of our data. The prior distribution serves the same role as the roughness penalty in the regularization approach, but with the added benefits of interpretability and coherence.

The specific choice of Lévy prior distributions leads both to (usually welcome) prior independence of the spectral measure of disjoint sets and to tractable computational problems. Note that the *posterior* distribution of $\Gamma(d\lambda)$, while still discrete, is *not*

Lévy— the point process $H(du\,d\lambda)$ in the representation $\Gamma(d\lambda) = \int_{\mathbb{R}_+} u\,H(du, d\lambda)$ assigns independent random variables $H(A_k) \sim \mathrm{Po}(\nu(A_k))$ to disjoint sets $A_k \subset \mathbb{R}_+ \times \Lambda$ but the likelihood function induces dependence among the $\{H(A_k)\}$.

Our nonparametric Bayesian approach offers a number of advantages over earlier methods, including

Flexibility : Different choices for the Lévy measure $\nu(du, d\lambda)$ will lead to "smooth" or "bumpy" measures, allowing the analysis to reflect any expert opinion about features of $\Gamma(d\lambda)$ such as smoothness, uni- or multi-modality, zero-shear viscosity $\Gamma(\mathbb{R}_+)$, etc.

Tractability : The MCMC approach described here works equally well for any of these choices of Lévy measure (and measurement-error model). Posterior distributions of any quantity $\Gamma[\phi] = \int \phi(\lambda)\,\Gamma(d\lambda)$, including interval measures $\Gamma(A)$ and zero-shear viscosity $\Gamma(\mathbb{R}_+)$, are easily computed.

Parsimony : Some choices of $\nu(du,\, d\lambda)$ will almost-surely have finite numbers M of mass points, and can even have $\mathrm{E}[M]$ as small as two or three, leading to strikingly parsimonious representations of $\Gamma(d\lambda)$ as a sum of a small number of point masses, similar to the representations of Anderssen and Davies (2001).

In our work with the Gamma prior we found no apparent sensitivity to the choice of the cut-off $\epsilon > 0$. We settled on a value small enough that our truncated approximation includes 99.5% of the prior expected mass, but in a sensitivity analysis we varied ϵ over a wide range.

We also explored a similar algorithm modeling a fixed number M of mass points (u_j, λ_j) diffusing over $U_\epsilon \times \Lambda$ with *reflecting* boundary conditions at $u = \epsilon$, rather than the *free* boundary conditions of the present implementation. Model fit and posterior distributions of $\Gamma(A)$ for intervals A were very similar to those found in the present study, for a wide range of values of M.

We continue to explore the prior elicitation issues that arise in this modeling approach, studying a range of different Lévy measures and seeing how they affect posterior inference. We are also exploring inference for other inverse problems; some of this work will be described in Wolpert, Hansen and Ickstadt (to appear).

REFERENCES

Abramowitz, M. and Stegun, I. A. (1964). Handbook of Mathematical Functions With Formulas, Graphs and Mathematical Tables. Washington, D.C.: National Bureau of Standards.

Anderssen, R. S. and Davies, A. R. (2001). Simple moving-average formulae for the direct recovery of the relaxation spectrum. *J. Rheol.* **45**, 1–27.

Anderssen, R. S. and Hansen, M. B. (to appear). Bayesian relaxation spectrum recovery.

Berger, L. (1988). Untersuchung zum rheologischen Verhalten von Polybutadienen mit bimodaler Molmassenverteilung. Ph.D. Thesis, Eidgenössische Technische Hochschule (ETH), Zurich.

Boltzmann, L. (1876). Zur Theorie der elastischen Nachwirkung. *Ann. Phys. Chem.* **7**, 624–654.

Davies, A. R. and Anderssen, R. S. (1997). Sampling localization in determining the relaxation spectrum. *J Non-Newtonian Fluid Mech.* **73**, 163–179.

Feller, W. (1971). An Introduction to Probability Theory and its Application (vol 2), second edition. Chichester: Wiley.

Fredholm, Erik I. (1900). Sur une nouvelle méthode pour la résolution du problèm de Dirichlet. In *Œuvres complètes: publiées sous les auspices de la Kungliga svenska vetenskapsakademien par l'Institut Mittag-Leffler*, 61–68, reprinted 1955. Malmö: Litos reprotryck.

Jacod, J. and Shiryaev, A. N. (1987). Limit Theorems for Stochastic Processes. Berlin: Springer.

Kirsch, A. (1996). An Introduction to the Mathematical Theory of Inverse Problems. New York: Springer.

Tierney, L. (1994). Markov chains for exploring posterior distributions (with discussion). *Ann. Statist.* **22**, 1701–1762.

Tikhonov, A. N. (1963). Solution of incorrectly formulated problems and the regularization method. *Soviet Doklady* **4**, 1035–1038.

Wolpert, R. L., Hansen, M. B. and Ickstadt, K. (to appear). Inverse problems and a Lévy process solution.

Wolpert, R. L. and Ickstadt, K. (1998a). Poisson/gamma random field models for spatial statistics. *Biometrika* **85**, 251–267.

Wolpert, R. L. and Ickstadt, K. (1998b). Simulation of Lévy random fields. In Practical Nonparametric and Semiparametric Bayesian Statistics (D. Dey, P. Müller and D. Sinha, eds.), New York: Springer, 227–242.

ACKNOWLEDGMENTS

The authors acknowledge gratefully the support of US NSF grant DMS-9626829, of US EPA grant R828686-01-0, of EC TMR grant ERB-FMRX-CT96-0096, and of the Centre for Mathematical Physics and Stochastics (MaPhySto), funded by the Danish National Research Foundation. We would like to thank Robert S. Anderssen for suggesting the problem and Jesper Møller and the Department of Mathematical Sciences of Aalborg University for facilitating our collaboration.