



## Are Maintenance Practices for Railroad Tracks Effective?

Jason R. W Merrick, Refik Soyer & Thomas A Mazzuchi

To cite this article: Jason R. W Merrick, Refik Soyer & Thomas A Mazzuchi (2005) Are Maintenance Practices for Railroad Tracks Effective?, Journal of the American Statistical Association, 100:469, 17-25, DOI: [10.1198/016214504000002104](https://doi.org/10.1198/016214504000002104)

To link to this article: <https://doi.org/10.1198/016214504000002104>



Published online: 31 Dec 2011.



Submit your article to this journal [↗](#)



Article views: 84



View related articles [↗](#)



Citing articles: 17 View citing articles [↗](#)

# Are Maintenance Practices for Railroad Tracks Effective?

Jason R. W. MERRICK, Refik SOYER, and Thomas A. MAZZUCHI

---

The Association of American Railroads wished to determine the effect of a maintenance practice known as grinding on the occurrence of rail fatigue defects and on the subsequent total traffic usage before a track must be replaced. Because a designed experiment was not practical, an analysis of historical data from the Canadian Northern Railroad is presented. In the analysis, certain covariate data are available, specifically the amount of grinding and some physical characteristics of the rail; other important covariate data are not available, however. A model for the number of defects as a function of traffic usage is developed based on a modulated Poisson point process. The model incorporates the effect of the available covariates and a mixture of Dirichlet processes set-up for the scale parameters of the individual rail sections that allows an assessment of the overall effect of the unavailable covariates. The model is then used to determine an optimal replacement period for a whole rail track. The analysis demonstrates that grinding reduces the expected number of defects and increases the optimal replacement interval.

KEY WORDS: Counting process; Markov chain Monte Carlo; Mixture of Dirichlet processes.

---

## 1. INTRODUCTION

Rail fatigue failure is a substantial economic burden and a threat to safe operation in the railroad industry (Cannon 1999). Cracks that begin as tiny flaws inside steel rails can grow into surface cracks and eventually cause the rails to fracture. The growth of internal cracks is propagated by contact with rolling stock, so the number of defects is a function of traffic usage, measured in millions of gross tons (MGT) of stock traversing the rail. In an effort to prevent derailments caused by rail fractures, the cracked rail can be either ground down, removing the metal surrounding the crack and leaving only solid metal, or welded, fusing the crack. In either case the repair is minimal, providing that the crack has not led to complete fracture of the rail. Grinding of the rails is performed by specialized cars pulled by powerful locomotives. Circular stones are rotated at high speed to wear away the surface metal. Hard stones are used initially to remove most of the steel surface, followed by softer stones used to smooth the surface and achieve the correct profile.

This type of corrective maintenance is costly for railroad companies (Stanford and Kalousek 2000), and because the repair is minimal, the frequency of the need for grinding may increase with rail age as measured in MGTs. As the expense of maintaining an aging rail becomes too high, the entire track must be replaced. During the 1990s, the railroad industry began a transition from the traditional corrective mode of grinding (i.e., grinding in response to discovered cracks) to a preventative mode using regular removal of the surface of the rail can to stop the propagation of existing surface-level cracks, stop the initiation of new cracks, and remove inactive cracks. Although the practice of preventative grinding is backed by sound engineering theory, there have been no empirical studies to illustrate whether this practice is effective in improving rail track reliability and prolonging the replacement period. In fact, the Transportation Research Board of National Academy

of Sciences' Committee on Railway Maintenance developed two specific research problem statements: (1) the "development of predictive track maintenance models which utilize field data" and (2) determining "when to renew or maintain track" (Transportation Research Board 2000). This article outlines a study of rail grinding performed for the Association of American Railroads (AAR).

The initial suggestion to the AAR was to use their test tracks to determine the effect of grinding by grinding only part of the test track, leaving some track untouched to act as a control. The problem with performing such a study is the time it would take to collect sufficient data for meaningful conclusions. Historical data from the Canadian National Railroad (CN) was made available by the AAR with different levels of grinding performed on different tracks of rail to determine the effect of grinding on the number of rail fatigue failures. However, because the data were not collected in a designed experiment, other factors varied along the length of the rail tracks, some available in the data and some not. These covariates had to be accounted for in determining the effect of grinding.

An inference model for predicting track failures based on this data was provided by Campodonico and Singpurwalla (1995) without covariate information. The analysis presented in this article extends these results to address the questions posed by the AAR and the two research problem statements from the Transportation Research Board of the National Academy of Sciences.

The article is organized as follows. In Section 2 the manner of data collection and the subsequent form of the data are discussed. In Section 3 a semiparametric setup is proposed to assess the effect of covariate data, including grinding, on the intensity of failures and whether covariates unavailable in the data have significant effects on the formation of cracks are assessed. The effect of unavailable covariate data is assessed in Section 4, before the effect of grinding and other available covariates are discussed in Section 5. In Section 6 optimal replacement strategies are developed for railroad tracks that may have different physical characteristics and maintenance practices along its length. The decision-theoretic approach is used to determine the effect of different levels of grinding on the useful life of rail.

---

Jason R. W. Merrick is Associate Professor, Department of Statistical Sciences and Operations Research, Virginia Commonwealth University, Richmond, VA 23284 (E-mail: [jrmerrick@vcu.edu](mailto:jrmerrick@vcu.edu)). Refik Soyer is Professor, Department of Management Science, George Washington University, Washington, DC 20052 (E-mail: [soyer@gwu.edu](mailto:soyer@gwu.edu)). Thomas A. Mazzuchi is Professor, Department of Engineering Management & Systems Engineering, George Washington University, Washington, DC 20052 (E-mail: [mazzuchi@seas.gwu.edu](mailto:mazzuchi@seas.gwu.edu)). This work was supported in part by the Association of American Railroads.

## 2. CANADIAN NATIONAL RAILROAD DATA

CN operates 15,500 route miles of track across North America and has been performing varying levels of preventive grinding for many years. In the past, rail surface defects were detected by inspections performed by human operators. Modern detection techniques included visual image analysis (Short 1993) and ultrasound. The detection equipment is housed in specialized cars that travel along the rail detecting and recording surface-level cracks and, more recently, internal cracks. However, when a crack is detected in a given run of the detection equipment, it is known only that the crack occurred at some point between the current run and the previous run. Because multiple cracks may be detected in a given run, the data are in the form of counts over a defined interval (interval-censored).

The AAR supplied failure count data from five CN tracks. The detection technology used to find cracks in these five tracks was the same over period during which these data were collected. Each of the tracks had a constant level of grinding across its length, either 0, .75, or 1 mm of steel removed each year. Because the grinding level is constant for a given track, no nominal covariate to identify the different tracks was included in the analysis, because this would be confounded with the grinding covariate. Other covariate data were available, specifically the curvature of the track, the weight of the track (in lb/yard), and the speed limit for traffic traversing the rail (in mph). Each of these other covariates varied along the length of a track; thus the data were made available for sections of each track that had constant covariate values, 132 sections in all with varying lengths, ranging from a few hundred feet to 2 miles. The AAR suggested converting the curvature of the rails from simple curvature in degrees to the percentage of the section length that was curved by more than 3 degrees. This measure varied from 0% to 91%, whereas the rail weight varied from 100 to 136 lb/ft and the maximum speed varied from 30 to 70 mph. The number of failures per mile between inspections ranged from 1 to 10. Rail wear is measured as a function of traffic usage, not chronological age (Campodonico and Singpurwalla 1995). Thus the intervals between inspections are defined by the number of MGTs at the time of each inspection. The lowest MGT value was 3, and the highest was 799. This implies that no failure data are available for new track; the newest track recorded had already had 3 MGTs of traffic usage. In other words, the data are left-censored as well as interval-censored.

As an illustration of the form of the data, Table 1 gives the number of failures per mile over the recorded inspection intervals for five rail sections that form a continuous segment of a single track. For each section, the start value indicates the MGT of traffic usage at the beginning inspection for that section. The values for interval 1 are then the MGT values at the next inspection and the number of defects per mile found during the inspection. These failures must have occurred between the two inspections, that is, in the first inspection interval. The covariate values for these sections are also given in Table 1. The experts at the AAR noted that some relevant covariate information was not available in the data. Examples given were the hardness of the steel used in the rail manufacture (measured by the Brinell hardness number), whether the track was lubricated, a wear-reducing maintenance strategy, and the environmental conditions to which the rails were exposed. This was of major concern to the AAR, because they did not want the effect of grinding to be masked by the unavailable covariates.

## 3. MODELING THE EFFECT OF GRINDING

First, we define the necessary notation. Let  $i = 1, \dots, 132$  index the rail sections and let  $r_i$  denote the total number of inspection runs in the data for the  $i$ th rail section. The inspection runs are performed at  $t_{i,0}, \dots, t_{i,r_i}$  MGTs and  $N_i(t_{i,1}), \dots, N_i(t_{i,r_i})$  defects are discovered. The covariate data for the  $i$ th section is constant, with  $Z_{i,1}$  denoting the level of grinding,  $Z_{i,2}$  the curvature,  $Z_{i,3}$  the rail weight, and  $Z_{i,4}$  the maximum speed. Campodonico and Singpurwalla (1995) modeled the number of defects occurring in rail tracks as a function of the cumulative traffic usage using a nonhomogenous Poisson point process (NHPP). They chose a parametric intensity function that, reparameterized for our analysis, may be written as

$$(N_i(t)|\alpha, \gamma, L_i) \sim \text{NHPP}(L_i\alpha\gamma t^{\gamma-1}) \quad (1)$$

for  $i = 1, \dots, 132$ , where  $L_i$  is the length of the section in miles. This is the well-known power-law form commonly used in reliability because the intensity function  $L_i\alpha\gamma t^{\gamma-1}$  can be increasing ( $\gamma > 1$ ), constant ( $\gamma = 1$ ), or decreasing ( $\gamma < 1$ ). An increasing intensity function would signify that the rate of occurrence of defects is increasing with usage; that is, the rail is degrading. When the form  $\alpha\gamma t^{\gamma-1}$  is chosen for the intensity function, the process is known as a Weibull process, because the parametric form is the same as the hazard rate of a Weibull distribution (Ebeling 1997). It should be noted, however, that the

Table 1. Failures per Mile, MGT Intervals, and Covariate Values for Five Rail Sections

		Intervals						Covariates			
Section		Start	1	2	3	4	5	Grind	Curve	Weight	Speed
1	MGT	219.11	236.59	249.47				.75	42%	136	30
	Failures		1.803	1							
2	MGT	219.11	255.19	274.64	294.73			.75	9%	115	40
	Failures		1	2.496	1.499						
3	MGT	231.99	255.19	274.64				.75	0%	115	40
	Failures		1.318	1.53							
4	MGT	231.99	268.07	287.52				.75	0%	115	40
	Failures		1	2.33							
5	MGT	231.99	249.47	268.07	287.52	307.61	324.77	.75	24%	115	40
	Failures		1	1.931	2.048	1.116	1.466				

interfailure times are not in fact Weibull-distributed as the name would appear to imply. Additional discussion of the NHPP, particularly the validity of the assumption of the NHPP for this type of failure process, is offered in Appendix A.

It is interesting to examine the data to determine whether our choice of the parametric form for the intensity function of the NHPP is reasonable. Most of the available procedures for validating the form of the intensity function are based on failure time data, unlike here where the data are left- and interval-censored. An empirical estimate of the intensity function, denoted by  $\hat{\lambda}_i(t)$  for the  $i$ th rail section, can be obtained by taking the failure count per mile and dividing by the length of the interval, specifically  $\hat{\lambda}_i(t_{i,k}) = \frac{N_i(t_{i,k}) - N_i(t_{i,k-1})}{t_{i,k} - t_{i,k-1}}$  for  $i = 1, \dots, 132$  and  $k = 2, \dots, r_i$ . Now consider the parametric form of the intensity function, which we assume to be  $\lambda(t) = L_i \alpha \gamma t^{\gamma-1}$ . Dividing by  $L_i$  and taking the natural logarithm of each side yields

$$\ln\left(\frac{\lambda(t)}{L_i}\right) = \ln(\alpha) + \ln(\gamma) + (\gamma - 1) \ln(t). \quad (2)$$

This implies that if we plot the natural logarithm of the empirical intensity function divided by the section length against the natural logarithm of MGT, then we should obtain an approximate straight line. Figure 1 shows such a plot for the CN data. Deviations from the line are to be expected, due to the natural variability in the data. However, performing a linear regression on this data does yield a significant fit ( $p < .001$ ). This implies that the parametric form for the intensity function is reasonable in this application.

The model of Campodonico and Singpurwalla can be extended to take into account the effect of covariates on failure intensity. In doing so, an obvious choice is the modulated Poisson process model proposed by Cox (1972a), specifically

$$(N_i(t)|\alpha, \gamma, \beta, L_i, \mathbf{Z}_i) \sim \text{NHPP}(L_i \alpha \gamma t^{\gamma-1} e^{\sum_{j=1}^4 \beta_j Z_{i,j}}) \quad (3)$$

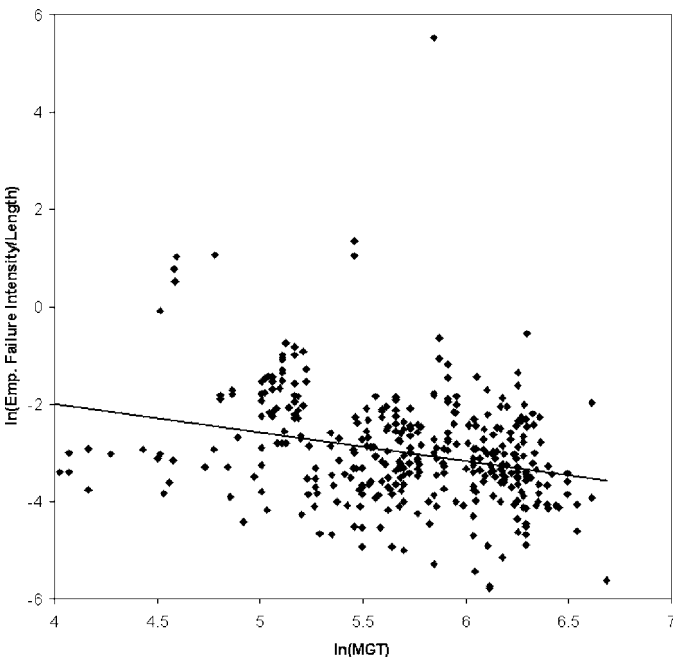


Figure 1. An Empirical Validation of the Parametric Form for the Intensity Function.

for  $i = 1, \dots, 132$ , where  $\beta = (\beta_1, \dots, \beta_4)$  and  $\mathbf{Z}_i = (Z_{i,1}, \dots, Z_{i,4})$ . Additional discussion of the modulated Poisson process is offered in Appendix A. Our empirical validation can be extended to consider the differences between the rail sections and to ascertain whether the assumption of a proportional effect for the covariates is reasonable. Taking the new parametric form for the intensity function, dividing by  $L_i$ , and taking the natural logarithm of each side yields

$$\ln\left(\frac{\lambda(t)}{L_i}\right) = \ln(\alpha) + \ln(\gamma) + (\gamma - 1) \ln(t) + \sum_{j=1}^4 \beta_j Z_{i,j}. \quad (4)$$

Note that in the new form, each rail section essentially has an added constant that depends on the level of the covariates. In fact, if we take the natural logarithm of the empirical intensity function and fit a model that includes a linear term for the natural logarithm of MGT and an additive term for each rail section, then we again obtain a significant model ( $p < .001$ ), with the additive rail effect term itself being significant ( $p < .001$ ). This implies that the covariates do have a proportional effect on the intensity function for each rail section.

However, this simple parametric model has disadvantages when there is missing relevant covariate information, as in our case. We can test for such additional heterogeneity using an empirical approach. If we take the natural logarithm of the empirical intensity function and fit the model given in (4), then we may examine the residuals of this model for heterogeneity across the rails. This procedure removes the covariate effects for the four known covariates and the baseline failure intensity, but leaves any additional section-to-section differences. An analysis of variance on these residuals across the different rail sections is significant ( $p = .0268$ ), indicating that there are additional additive differences between the failure intensities on the logarithmic scale. This confirms that section-to-section differences in failure intensity do exist beyond the effect of the known covariates and that there is indeed a proportional effect on the failure intensity.

A reasonable way to take into account this section-to-section variation is to index the scale parameter  $\alpha$  as  $\alpha_i$  and formulate the problem as a hierarchical Bayes problem in the sense of Lindley and Smith (1972). In this application, we may write this formulation as

$$(N_i(t)|\alpha_i, \gamma, \beta, L_i, \mathbf{Z}_i) \sim \text{NHPP}(L_i \alpha_i \gamma t^{\gamma-1} e^{\sum_{j=1}^4 \beta_j Z_{i,j}}), \quad (5)$$

$$\alpha_i \sim G_0,$$

for  $i = 1, \dots, m$ . The terms  $\alpha_i$  are assumed to be drawn from a mixing distribution  $G_0$ , whose form is known but whose parameters are not known. In this case examination of the posterior distribution of the scale parameters for each individual rail section, the  $\alpha_i$ 's, allows the analyst to determine which rail sections have significantly different failure intensities due to unavailable covariates. We should be cautious about this model form, though. The model in (5) implies that the values of missing covariates vary from one rail section to another, as reflected by the  $\alpha_i$ 's, and that the  $\alpha_i$ 's are conditionally independent given  $G_0$ . This is not appropriate in our case, because many rail sections may in fact share the same values for the unavailable covariates as for the known covariates (e.g., only three different

levels of grinding) and thus share the same value for  $\alpha_i$ . Thus the simple parametric model with a single scale parameter is insufficient for the problem at hand, because it cannot model the unavailable covariates, whereas the hierarchical Bayes model does not allow groups of rail sections to share the same scale parameter value.

A model that can incorporate differences between the rail sections beyond the available covariates but allows for similarity between groups of rails would be more appropriate. Such a model can be obtained using the mixture of Dirichlet processes (MDP) setup proposed by Antoniak (1974). In this manner, we are using the MDP setup in a similar manner to Escobar and West (1995), who examined the clustering of galaxies. In our case we are looking for the clustering of rail sections after known covariate effects have been removed. We may use the MDP setup by writing

$$(N_i(t)|\alpha_i, \gamma, \beta, L_i, \mathbf{Z}_i) \sim \text{NHPP}(L_i\alpha_i\gamma t^{\gamma-1}e^{\sum_{j=1}^4\beta_j Z_{i,j}}), \quad (6)$$

$$\alpha_i \sim G,$$

for  $i = 1, \dots, 132$ . The mixing distribution  $G$  is assumed to be unknown and thus is a parameter of the model. The prior distribution assumed for the mixing distribution is a Dirichlet process, hence the term “Dirichlet process mixed models” used by Mukhopadhyay and Gelfand (1997). Due to the discreteness of  $G$  (as a result of the Dirichlet process prior), there is a positive probability that some of the  $\alpha_i$ ’s will take the same values. This formation of groups of  $\alpha_i$ ’s allows grouping of similar rail sections. This represents the fact that some of the values of the unavailable covariates will be the same across multiple rail sections. The baseline distribution for the Dirichlet process prior is naturally assumed to be the same distribution chosen for the mixing distribution  $G_0$  in the hierarchical Bayes model with additional uncertainty reflected by the smoothing parameter of the Dirichlet process, denoted by  $M$ . The value of  $M$  effects the degree to which grouping occurs. Escobar and West (1995) added an additional level of hierarchy by assuming that the prior distribution of  $M$  is a gamma distribution. This assumption allows updating of the uncertainty concerning this critical smoothing parameter from the data and reveals the number of groups of rails that have similar values for the unavailable covariates. Examination of the distribution of the number of separate groups of rails allows us to discover whether the simple parametric model is appropriate (if there is strong support for only one group of rails) or whether the hierarchical form is appropriate (if there is strong support for 132 separate groups of rails). This latter case would indicate differences in the failure intensities of each rail section beyond that accounted for by the known covariates.

#### 4. ARE THE UNAVAILABLE COVARIATES IMPORTANT?

Although the experts felt a priori that important covariates were missing from the available data, our analysis allows us to evaluate the actual effect of missing covariates. We have shown using our ad hoc approach that there are additional proportional contributions to the failure intensities of the rail sections beyond the four available covariates. In this section we demonstrate that each of our model elaborations in the previous sections

improves the fit of the model. Next we examine the posterior distribution of the number of groups in the scale parameters of the MDP model to discover whether there are groupings in the remaining differences between the rail sections. Finally, we compare the marginal posterior distributions of the scale parameters to assess the size of the differences. To perform such an analysis, we must first determine appropriate prior distributions for the parameters of the model and update these priors with the available data.

In the MDP setup, the prior best guess for the mixing distribution  $G$  of the  $\alpha_i$ ’s is assumed to be a diffuse gamma distribution with mean 1 and variance 1,000. The prior distribution on the smoothing parameter  $M$  of the Dirichlet process prior is assumed to be a gamma distribution with parameters .132 and .001 that has a mean equal to the number of rail sections used in the analysis and a high variance. This choice was suggested by a knowledgeable referee to allow for values of  $M$  that would support from 1 to 132 groups of rail sections. A diffuse normal prior truncated at 0 with mean 1 and variance 10 is assumed for  $\gamma$ . The scientists at the AAR were uncertain about the effect of the covariates. Thus the prior distributions for the parameters in  $\beta$ , representing the effect of the covariates, are assumed to be independent normal distributions with mean 0 and variance 20, reflecting this uncertainty. In the analysis the covariate data are rescaled (by dividing each covariate by its maximum value), to avoid computational difficulties. Note that all prior specifications are diffuse, to allow the data to dominate in the analysis. This was a choice made by the AAR. Details of the Markov chain Monte Carlo (MCMC) methods for inference for our model are given in Appendix B. We ran single-chain Gibbs samplers to obtain 2,500 samples in each case, with a warm-up of 5,000 and a lag of 25 between successive samples. We denote the posterior sample that we obtained by  $(\alpha_1^k, \dots, \alpha_{132}^k, \beta_1^k, \dots, \beta_4^k, \gamma^k, M^k)$  for  $k = 1, \dots, 2,500$ . At the suggestion of the referees and editors, we varied the variance specifications and found no differences in the posterior results. The algorithm in Appendix B can be modified to accommodate the simpler models. The parametric model with no covariates in (1) can be obtained by forcing the number of groups to be one and leaving out the covariate terms. The parametric model with covariates in (3) can be obtained by forcing the number of groups to be one, but including the covariate terms. The hierarchical Bayes model in (5) can be obtained by forcing the number of groups to be 132. In each of these models, the prior on the single scale parameter is the gamma distribution assumed for the mixing distribution  $G$  of the  $\alpha_i$ ’s in the MDP setup. All other priors are the same if used.

The deviance information criterion (DIC) was proposed by Spiegelhalter et al. (2002) as a measure of fit for complex hierarchical models when the number of parameters is not clearly defined. This is the case with the MDP setup. The Bayesian deviance, denoted by  $DV(\alpha_1, \dots, \alpha_{132}, \gamma, \beta)$ , is defined as  $-2 \ln P(D|\alpha_1, \dots, \alpha_{132}, \gamma, \beta) + 2 \ln f(D)$ , where  $f(D)$  is some fully specified standardizing term that is a function of the data alone and thus does not effect model comparison. The model “fit” is represented by  $\overline{DV} = E[DV(\alpha_1, \dots, \alpha_{132}, \gamma, \beta)|D]$ , which can be approximated from the posterior sample. An estimate of the effective number of parameters, denoted by  $p_{DV}$ , is represented by  $E[DV(\alpha_1, \dots, \alpha_{132}, \gamma, \beta)|D] -$

Table 2. A Comparison of the Models

Model	$\overline{DV}$	$p_{DV}$	DIC
Parametric with no covariates	408.2	1.1	406.0
Parametric with covariates	251.3	6.6	257.9
Hierarchical Bayes	137.4	102.6	240.0
MDP setup	143.7	47.7	191.4

$DV(E[\alpha_1, \dots, \alpha_{132}, \gamma, \beta|D])$ , where the latter term is simply the function  $DV$  calculated at the posterior means of the parameters. The  $DIC$  is then defined as the sum  $DIC = \overline{DV} + p_{DV}$ , which is equivalent to the Akaike information criterion if the prior information is negligible (see Spiegelhalter et al. 2002 for details). The results of the  $DIC$  calculations are given in Table 2. Because lower values are preferable, the  $DIC$  results show that each elaboration that we made to model the failure process is warranted. Introducing the four known covariates yields an improvement in the  $DIC$ , as does modeling the unknown covariates through the MDP setup. The hierarchical Bayes model does in fact show a better fit statistic than the MDP setup, but the large numbers of parameters gives a worse overall  $DIC$  value. Note that the effective numbers of parameters for the parametric model with and without covariates are very close to the true number of parameters, six and two. For the MDP model, the effective number of parameters is about 48, whereas the model includes 4 covariate effect parameters, 1 shape parameter, and 132 scale parameters. This indicates that there are groups of similar rail sections in the data for which the scale parameters are clustered, reflecting groups of rails with the same values for the unavailable covariates. We should be careful when attempting to estimate the number of groups from the effective number of parameters, however. The value of  $p_{DV}$  appears to suggest  $48 - 4 - 1 = 43$  clusters of scale parameters, but because the hierarchical mixing distribution is handled in a nonparametric manner with a Dirichlet process prior, the true number of parameters and their interpretation are more complex than this simple calculation.

A better mechanism for examining the idea of clusters in the scale parameters, the posterior distribution of the number of groups in the  $\alpha_i$ 's is shown in Figure 2, confirming the existence of groups of rails with common values for the unavailable

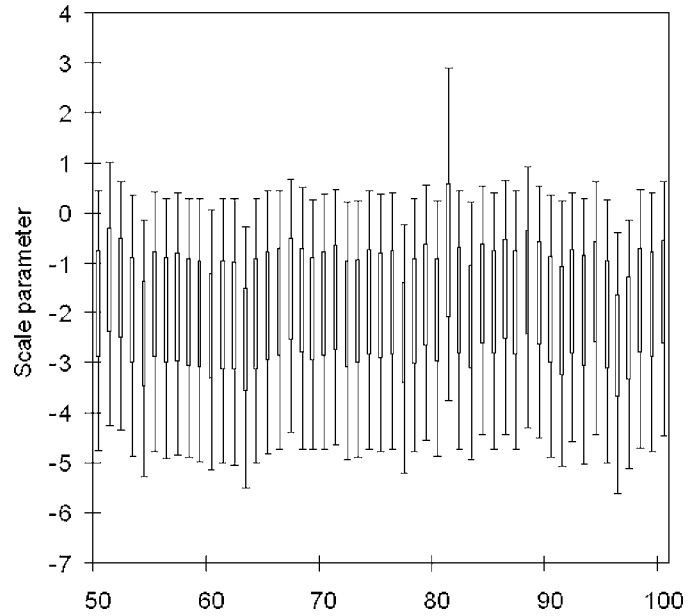
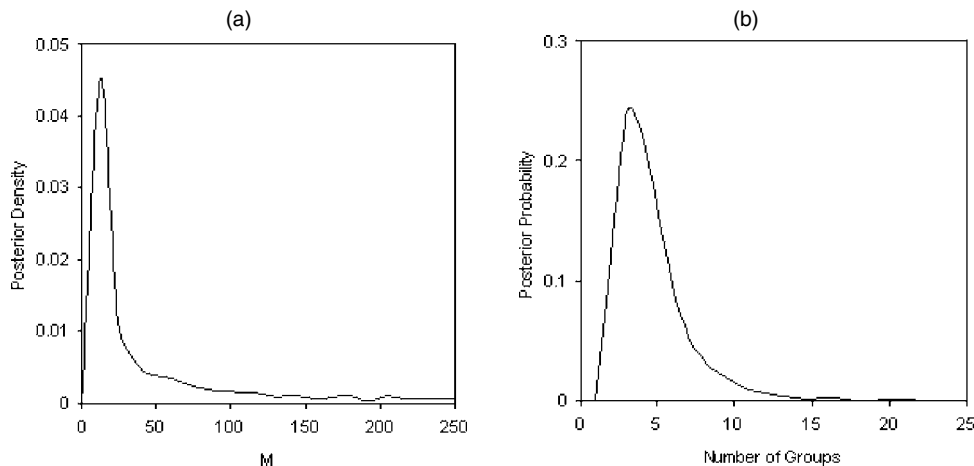


Figure 3. Boxplots of the Marginal Posterior Distributions of a Subset of the Scale Parameters on a Log Scale.

covariates. In fact, the estimated posterior probability of a single group is only .00041, evidence against the parametric model with a single scale parameter, and the estimated posterior probability of 132 groups is approximately 0, evidence against the hierarchical model with 132 scale parameters. This reinforces the conclusions made using the  $DIC$ . We note that the posterior distribution has its mass somewhere between 2 and 15, implying that whereas the data imply some differences between the failure intensities of the rail sections beyond the known covariate effects, there is in fact a relatively small number of groups. To demonstrate the magnitude of the effect of these differences in failure intensities, boxplots of the marginal posterior distributions of the log of a subset of the scale parameters,  $\log(\alpha_i)$ 's, are shown in Figure 3. The scale parameters' marginal posterior distributions show variation from rail section to rail section. The variation in the  $\alpha_i$ 's reflects differences between the rail sections that are not explained by the four covariates included in this analysis and then must be assigned to the unavailable co-

Figure 2. Marginal Posterior Distributions of the Parameter  $M$  (a) and the Number of Groups of  $\alpha_i$ 's (b) in the Dirichlet Process Mixing.

variates. However, in terms of the model, the scale parameters have accounted for these differences, allowing a meaningful examination of the effect of the known covariates.

### 5. IS GRINDING EFFECTIVE?

With the effect of missing covariates verified and accounted for, the important questions about grinding effectiveness can be answered by examining the marginal posterior distributions of the parameters of the model. Although the effect of grinding is the one most pertinent to this analysis, the AAR was also interested in the effect of the other available covariates and whether the effects of the unavailable covariates are significant. Figure 4 shows the marginal posterior distributions of the covariate effect parameters,  $\beta_1, \dots, \beta_4$ . The prior distributions are also shown, but they do not show up well, because they are extremely flat in the region containing the posterior mass. For each covariate effect parameter, positive (negative) values indicate that an increase in the covariate measurement increases (decreases) the failure intensity of the rail. We first turn to the grinding effect parameter, the primary interest of the present study. Most of the mass of this posterior distribution is between  $-2$  and  $0$ , implying that the practice of grinding is beneficial for increasing the life of a rail section. This verifies the engineering theory that led to the practice and was good news for the AAR and the railroad industry in general. We examine the end result of this effect in Section 6 by examining the optimal replacement interval for different levels of grinding.

For the posterior distribution of the rail weight effect parameter, the posterior mass is located mostly on the negative values, implying a decrease in failure intensity with the use of heavier rail. This is to be expected, because heavy rail will stand up to the wear of traffic load better than light rail. The results for the other two covariates at first appeared counterintuitive. Curved rails experience higher stresses and thus would be expected to fail more often. Moreover, trains traveling at higher speed exert more stress on the rail and should increase the failure intensity. However, the posterior distribution of the curvature effect parameter and the maximum speed effect parameter have a negative mean, implying that curved rails and rails with high speed limits fail less often. At first, this result created confusion, until one expert pointed out that the rail used in curved sections is generally of better quality than that used in straight sections and that rail sections on which trains are allowed to travel at high speed are also usually good-quality rails. The quality of steel used in the rail manufacturing process is rated on the Brinell hardness number, one of the covariates not available in the data. We also note that the results for these two parameters was not as conclusive as the other two, because there was significant posterior mass on either side of  $0$ . For completeness, the posterior distribution of the shape parameter,  $\gamma$ , is concentrated around values between  $1$  and  $1.4$ . This implies that the rail section degrades with higher traffic load. This finding is to be expected and is in agreement with previous studies, and so for the sake of brevity a figure illustrating this is not included.

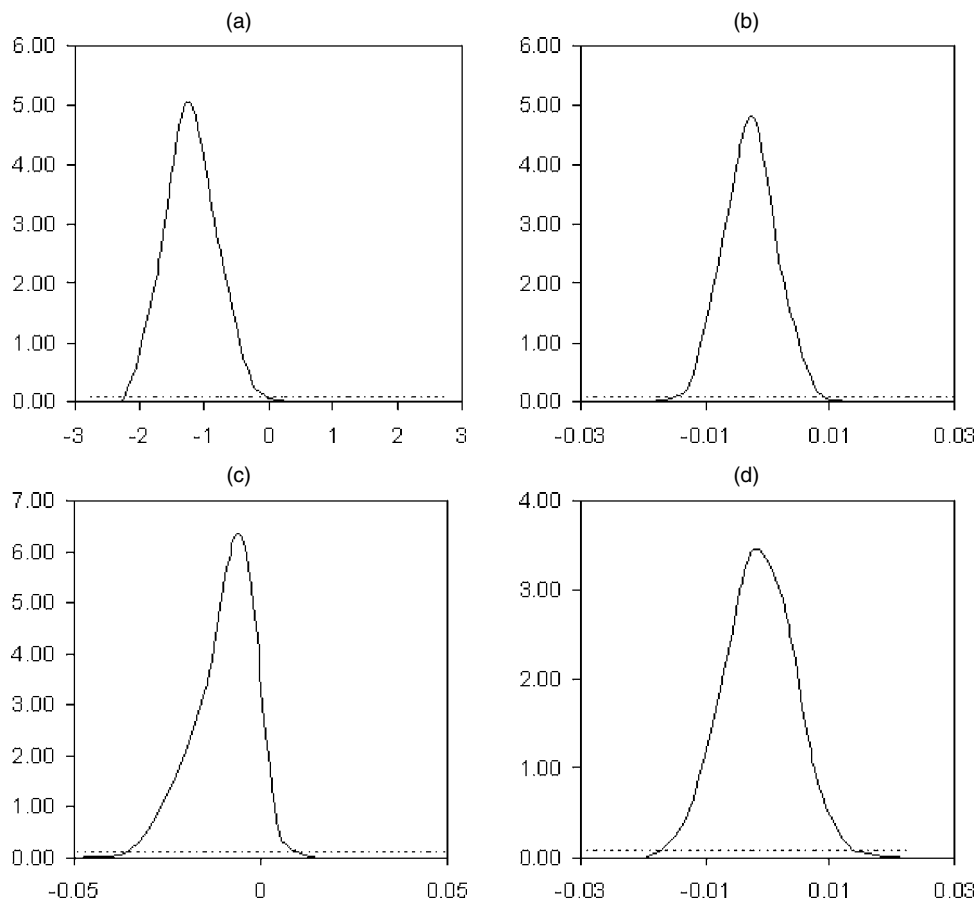


Figure 4. The Marginal Prior (.....) and Posterior (—) Distributions of the Covariate Effect Parameters. (a) Grinding effect parameter; (b) curvature effect parameter; (c) rail weight effect parameter; (d) maximum speed effect parameter.

This analysis demonstrates conclusively that the practice of grinding railroad tracks reduces the intensity of the failure process. It can also be concluded that the rails deteriorate with use. The variation in the  $\alpha_i$ 's indicates that some covariates that affect the failure behavior may be missing. The inclusion of such covariates as the Brinell hardness number, an indicator of the quality of the steel used, may be helpful in explaining the effect of rail curvature and maximum speed found in the analysis, but in our analysis this effect is modeled by the MDP setup. In a predictive sense, not knowing these additional covariates will add to the uncertainty because we will not know to which of these groups a new piece of rail may belong.

## 6. DOES GRINDING IMPROVE THE LIFE OF A RAIL?

In the previous section we illustrated the effectiveness of grinding as a preventive maintenance practice. It is also desirable to determine optimal replacement strategies and to examine the effect of grinding on these strategies. As pointed out in the Research Problem Statements report of the Transportation Research Board of the National Research Council, this is essential in determining an optimal rail-grinding schedule. In determining optimal strategies (see, e.g., Cho and Parlar 1991 for a review), the use of covariate information has not been considered. This is an important issue in our application, where a procedure is needed for determining the optimal block replacement period for a group of continuous rail sections and for assessing the effect of the grinding policy on their replacement interval. We follow the development of Mazzuchi and Soyer (1995, 1996), extending the approach to include covariate information and our MDP setup. The development of maintenance policies for railroad tracks, such as a grinding policy and replacement strategies, requires failure predictions for different rail sections at various MGT intervals. Once a posterior sample has been obtained under the MDP model, the expected number of failures occurring up to  $t$  MGT for an existing rail section  $i$  is approximated as

$$E[N_i(t)|D] \approx \frac{1}{2,500} \sum_{k=1}^{2,500} \alpha_i^k t^{\gamma^k} e^{\sum_{j=1}^4 \beta_j^k Z_{i,j}}. \quad (7)$$

For predictions for a new rail section  $i > 132$ , under the MDP approach, a sample must first be obtained from the new scale parameter of the rail section  $i$ , denoted by  $\alpha_i^k$ . A multinomial draw is performed, with probability  $\frac{M^k}{M^k + 132}$  that the draw is from the prior best-guess distribution  $G_0$  and probabilities  $\frac{1}{M^k + 132}$  for each of the sampled scale parameters ( $\alpha_1^k, \dots, \alpha_{132}^k$ ). Once the draw is made from the scale parameter, the expected number of failures can be approximated using (7).

One of the most commonly used replacement strategies for systems/items subject to wear is block replacement (see Cox 1962 for an early introduction). Under the block replacement protocol, all units are replaced at time points  $t_B, 2t_B, \dots$ , irrespective of their ages, and an in-service replacement or repair is made whenever failures occur. This replacement protocol is relevant to the rail track example because multiple continuous rail sections are replaced at one time, implying that the sections are grouped in to a block. In the case of rail replacement, the intervals are measured in terms of traffic usage instead of

time. As discussed previously, fatigue defects discovered between replacements are repaired either by corrective grinding or by welding. Under this scenario, in a planned replacement cycle of length  $t_B$  MGTs, the cost per MGT for the  $i$ th rail section,  $C(t_B, N_i(t_B))$ , is given by

$$\frac{c_P + c_F N_i(t_B)}{t_B}, \quad (8)$$

where  $N_i(t_B)$  represents the number of in-service failures for the  $i$ th rail section occurring in an interval of length  $t_B$ ,  $c_P$  is the cost of a planned replacement, and  $c_F$  is the cost of a repair consisting of corrective grinding or welding. Assuming that  $m$  rail sections will be replaced at  $t_B$  MGTs, the total cost per MGT,  $C(t_B, \mathbf{N}(t_B))$ , is given by

$$\sum_{i=1}^m C(t_B, N_i(t_B)), \quad (9)$$

where  $\mathbf{N}(t_B) = (N_1(t_B), \dots, N_m(t_B))$ . Following Mazzuchi and Soyer (1996), the optimal block replacement strategy  $t_B^*$  is determined by minimizing expected value of the cost,  $C(t_B, \mathbf{N}(t_B))$ . An approximate Bayesian optimal block replacement interval is determined by minimizing

$$E[C(t_B)|D] \approx \frac{1}{2,500} \sum_{k=1}^{2,500} \sum_{i=1}^m \frac{c_P + c_F \alpha_i^k t_B^{\gamma^k} e^{\sum_{j=1}^4 \beta_j^k Z_{i,j}}}{t_B}, \quad (10)$$

with respect to  $t_B$ .

To determine the effect of grinding on replacement intervals, we consider five sections of rail for which a common optimal block replacement period is required. The selected rail sections are the ones shown in Table 1 and form a continuous segment of a single rail track. Following the suggestion of experts, the cost of a planned replacement is assumed to be 20 times the cost of a minimal repair. The optimal block replacement period is approximately 620 MGTs at the current level of grinding, .75 mm per year. To illustrate the beneficial effect of grinding the optimal replacement interval for varying levels of grinding are shown in Figure 5. The figure demonstrates that the optimal replacement intervals significantly increases with increasing level of grinding, again demonstrating the effectiveness of grinding practices. This conclusion is good evidence in favor

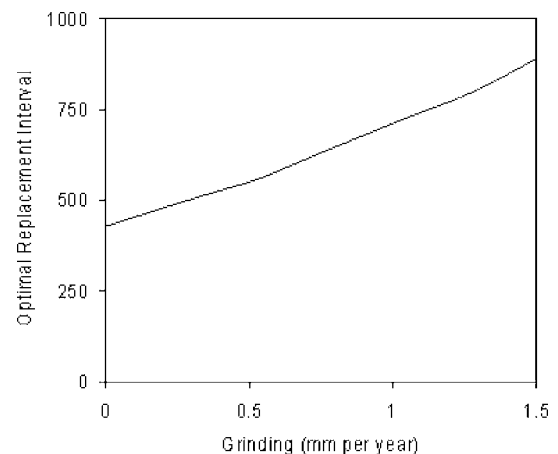


Figure 5. The Optimal Replacement Period for a Group of Rail Sections Under Varying Grinding Policies.



of the practice of grinding and thus is of great interest to the railroad industry.

## 7. CONCLUSIONS

In this analysis, the preventive maintenance practice known as grinding has been shown to have a beneficial effect on the rate of occurrence of fatigue defects in rail tracks. Several other covariates were also included in the analysis, whereas relevant covariates that were not available in the CN data were accounted for using the MDP setup for the scale parameters of the NHPP model. A decision-theoretic approach to determining optimal replacement intervals was used to illustrate the effect of grinding on the optimal replacement intervals, demonstrating that grinding will not only reduce the occurrence of fatigue failures, but also reduce their economic impact on the railroad industry.

## APPENDIX A: THE MODULATED POISSON PROCESS

Let  $N_i(t)$  denote the number of failures for the  $i$ th rail section in an interval of length  $t$  MGT. When a crack is found, the surface metal is ground away in a small area around the defect, leaving most of the rail section unaffected by the failure or the repair. Thus the failure rate of the rail section after repair is equal to the failure rate just before failure. This is known as minimal repair and implies that  $N_i(t)$  can be described by a NHPP with intensity function  $\lambda_i(t) = \frac{d}{dt}E[N_i(t)]$ , where  $E[\cdot]$  denotes the expectation. For a discussion of minimal repair and its association with the NHPP, see such standard texts as that by Ebeling (1997, p. 198).

Let  $\mathbf{Z}_i$  denote the  $p$ -dimensional vector of available covariates that describe the characteristics of the  $i$ th rail section and are not dependent on traffic usage. The modulated Poisson process was introduced by Cox (1972a) by assuming that in our case the intensity function of the  $i$ th rail section is related to the covariate vector via

$$\lambda_i(t; \mathbf{Z}_i) = \lambda_0(t)e^{\beta^T \mathbf{Z}_i}, \quad (\text{A.1})$$

where  $\lambda_0(t)$  is the baseline intensity function and  $\beta$  is a vector of  $p$  parameters. The modulated Poisson process defined by (A.1) is analogous to the proportional hazards model of Cox (1972b), where a similar form was used for the failure rate of a nonrepairable system. Under the modulated Poisson process, the ratio of the intensity functions of two rail sections at traffic usage  $t$  is given by  $\frac{\lambda_i(t; \mathbf{Z}_i)}{\lambda_j(t; \mathbf{Z}_j)} = e^{\beta^T (\mathbf{Z}_i - \mathbf{Z}_j)}$ , which does not depend on  $t$ . Hence the ratio of the hazard rates for two rails with different covariate effects is the same over time. In the modulated Poisson process, the effect of the wear on failure behavior of rail sections is described by the baseline intensity function  $\lambda_0(t)$ , and covariate effects are captured by the term  $e^{\beta^T \mathbf{Z}_i}$ .

## APPENDIX B: INFERENCE FOR THE MDP MODEL

Inference for the MDP model is an extension of the MCMC inference procedure for the parametric proportional hazards model developed by Dellaportas and Smith (1993) to include the likelihood for failure count data rather than actual failure times. However, inference for the MDP setup (Kuo 1986) requires efficient inference algorithms, such as those developed by Escobar and West (1995) and MacEachern (1994), with excellent reviews by MacEachern (1998) and MacEachern and Muller (2000). MacEachern and Muller also noted that the MDP setup is implicitly robust due to the wide range of possible prior distributions that can result from realizations of the Dirichlet process prior. In the specific application for each rail section  $i = 1, \dots, 132$ ,  $N_i(t)$  is observed at traffic usages  $t = t_{i,0}, \dots, t_{i,r_i}$  MGTs, where  $t_{i,0} < \dots < t_{i,r_i}$ . Thus the data

for the  $i$ th rail section are given by  $D_i = \{N_i(t_{ij}) = n_i(t_{ij}), j = 1, \dots, r_i, L_i, \mathbf{Z}_i\}$ . Using the independent increments property of the NHPP, the joint distribution of  $(N_i(t_{i,1}), \dots, N_i(t_{i,r_i})) | \alpha_i, \gamma, \beta, L_i, \mathbf{Z}_i$ , denoted by  $P(N_i(t) | \alpha_i, \gamma, \beta, L_i, \mathbf{Z}_i)$ , can be written as

$$\prod_{j=1}^{r_i} \frac{(L_i[\alpha_i t_{i,j}^\gamma - \alpha_i t_{i,j-1}^\gamma] e^{\sum_{k=1}^4 \beta_k Z_{i,k}})^{n_i(t_{i,j}) - n_i(t_{i,j-1})}}{[n_i(t_{i,j}) - n_i(t_{i,j-1})]!} \times \exp\{-L_i[\alpha_i t_{i,j}^\gamma - \alpha_i t_{i,j-1}^\gamma] e^{\sum_{k=1}^4 \beta_k Z_{i,k}}\}. \quad (\text{B.1})$$

Given the observed data from 132 rail sections,  $D = \bigcup_{i=1}^{132} D_i$ , the likelihood function of  $G$  and  $\beta$  is  $\prod_{i=1}^{132} \int P(N_i(t) | \alpha_i, \gamma, \beta, L_i, \mathbf{Z}_i) dG(\alpha_i)$ .

Given an arbitrary prior  $\pi(\beta, \gamma)$ , which are independent of the  $\alpha_i$ 's and  $G$ , the posterior distribution of  $G$  given  $\beta, \gamma$ , and  $D$  can be obtained as an MDP (see, e.g., Kuo and Mallick 1997). In implementing the Gibbs sampler, it is difficult to sample from the distribution  $G | \gamma, \beta, D$  (see, e.g., Kuo 1986). The approach introduced by Escobar and West (1995) can be adopted to the MDP setup given for the modulated Poisson process. The attractive feature of the Escobar and West approach is that in our case computation of the joint distribution  $\pi(\alpha_1, \dots, \alpha_{132}, \gamma, \beta | D)$  can be achieved without sampling from the posterior distribution of  $G | \gamma, \beta, D$ . It can be shown that  $\pi(\alpha_1, \dots, \alpha_{132} | \gamma, \beta, D)$  is equal to

$$\prod_{i=1}^{132} P(N_i(t) | \alpha_i, \gamma, \beta, L_i, \mathbf{Z}_i) \frac{MG_0(\alpha_i) + \sum_{j=1}^{i-1} \delta_{\alpha_j}(\alpha_i)}{M + i - 1}, \quad (\text{B.2})$$

where  $\delta_{\alpha_j}(\alpha_i)$  equals 1 if  $\alpha_i = \alpha_j$  and 0 otherwise. The  $\alpha_i$ 's will have a full conditional distribution  $(\alpha_i | \alpha^{(-i)}, \beta, D)$ , given by

$$q_{i,0} G_b(\alpha_i | N_i(t), \gamma, \beta, L_i, \mathbf{Z}_i) + q_{i,j} \sum_{j \neq i} \delta_{\alpha_j}(\alpha_i), \quad (\text{B.3})$$

where  $\alpha^{(-i)} = (\alpha_1, \dots, \alpha_{i-1}, \alpha_{i+1}, \dots, \alpha_{132})$  and  $G_b(\alpha_i | N_i(t), \gamma, \beta, L_i, \mathbf{Z}_i)$  is the baseline posterior distribution, which is proportional to  $P(N_i(t) | \alpha_i, \gamma, \beta, L_i, \mathbf{Z}_i) dG_0(\alpha_i)$ . The probabilities  $q_{i,0}$  and  $q_{i,j}$  are given by  $q_{i,0} \propto \int P(N_i(t) | \alpha_i, \gamma, \beta, L_i, \mathbf{Z}_i) dG_0(\alpha_i)$  and  $q_{i,j} \propto P(N_i(t) | \alpha_j, \gamma, \beta, L_i, \mathbf{Z}_i)$ , where  $P(N_i(t) | \alpha_j, \gamma, \beta, L_i, \mathbf{Z}_i)$  is the joint distribution for the  $i$ th rail section when  $\alpha_i = \alpha_j$  and  $q_{i,0} + \sum_{j \neq i} q_{i,j} = 1$ .

Because of the discreteness of  $G$  (as a result of the Dirichlet process prior), there is a positive probability that some of the  $\alpha_i$ 's will take the same values. An algorithm proposed by MacEachern (1994) exploits this fact to increase the efficiency of the algorithm and updates the  $\alpha_i$ 's in clusters. We use a further extension of the MDP setup incorporating  $M$  in to the Gibbs sampling analysis. Our approach follows that of Escobar and West (1995), assuming a priori that  $M$  follows an arbitrary prior  $\pi(M)$ . In their development, Escobar and West defined  $K$  to be the number of unique values of  $\alpha_1, \dots, \alpha_{132}$ , also referred to as the number of cliques by MacEachern (1998). Conditioned on  $K$ ,  $M$  is independent of all other parameters with density proportional to  $M^{K-1} (M + 132) B(M + 1, 132) \pi(M)$ , where  $B$  is the standard beta function. Escobar and West offered a simple two-step process for sampling from this distribution if  $\pi(M)$  is assumed to be a gamma distribution.  $K$  is also recorded in the Gibbs sample, because the distribution of the number of cliques is of interest in the reliability analysis in Section 5. The full conditional  $\pi(\beta_1 | \alpha_1, \dots, \alpha_{132}, \gamma, \beta_2, \dots, \beta_4, D)$  is proportional to  $\prod_{i=1}^{132} P(N_i(t) | \alpha_i, \gamma, \beta, L_i, \mathbf{Z}_i) \pi(\beta_1)$ , with similar expressions for  $\beta_2, \dots, \beta_4$ , and the full conditional density  $\pi(\gamma | \alpha_1, \dots, \alpha_{132}, \beta, D)$  is proportional to  $\prod_{i=1}^{132} P(N_i(t) | \alpha_i, \gamma, \beta, L_i, \mathbf{Z}_i) \pi(\gamma)$ . Obtaining samples from these distributions requires using the adaptive rejection sampling algorithm of Gilks and Wild (1993), because the densities are log-concave for appropriate  $\pi(\beta)$  and  $\pi(\gamma)$ . This is similar to the approach used by Dellaportas and Smith (1993) for inference on the

parametric proportional hazards model. For choices of  $\pi(\beta)$  and  $\pi(\gamma)$  that do not satisfy this condition, other algorithms, such as Metropolis sampling, can be applied.

[Received February 2000. Revised March 2004.]

## REFERENCES

- Antoniak, C. E. (1974), "Mixtures of Dirichlet Processes With Applications to Bayesian Non-Parametric Problems," *The Annals of Statistics*, 2, 1152–1174.
- Campodonico, S., and Singpurwalla, N. D. (1995), "Inference and Predictions for Poisson Point Processes Incorporating Expert Knowledge," *Journal of the American Statistical Association*, 90, 220–226.
- Cannon, D. F. (1999), "The Fight Against Rolling Contact Fatigue," *International Railway Journal*, 39, 27–28.
- Cho, D. I., and Parlar, M. (1991), "A Survey of Maintenance Models for Multi-Unit Systems," *European Journal of Operational Research*, 51, 1–23.
- Cox, D. R. (1962), *Renewal Theory*, London: Methuen.
- (1972a), "The Statistical Analysis of Dependencies in Point Processes," in *Stochastic Point Processes*, ed. P. A. W. Lewis, New York: Wiley, pp. 55–66.
- (1972b), "Regression Models and Life Tables," *Journal of the Royal Statistical Society, Ser. B*, 34, 187–220.
- Dellaportas, P., and Smith, A. F. M. (1993), "Bayesian Inference for Generalized Linear and Proportional Hazards Models via Gibbs Sampling," *Applied Statistics*, 42, 443–459.
- Ebeling, C. E. (1997), *An Introduction to Reliability and Maintainability Engineering*. New York: McGraw-Hill.
- Escobar, D. M., and West, M. (1995), "Bayesian Density Estimation and Inference Using Mixtures," *Journal of the American Statistical Association*, 90, 577–588.
- Gilks, W. R., and Wild, P. (1993), "Adaptive Rejection Sampling for Gibbs Sampling," *Applied Statistics*, 41, 337–348.
- Kuo, L. (1986), "Computation of Mixtures of Dirichlet Processes," *SIAM Journal of Scientific and Statistical Computing*, 7, 60–71.
- Kuo, L., and Mallick, B. K. (1997), "Bayesian Semi-Parametric Inference for the Accelerated Failure Time Model," *Canadian Journal of Statistics*, 25, 457–472.
- Lindley, D. V., and Smith, A. F. M. (1972), "Bayes Estimates for the Linear Model" (with discussion), *Journal of the Royal Statistical Society, Ser. B*, 34, 1–41.
- MacEachern, S. N. (1994), "Estimating Normal Means With a Conjugate-Style Dirichlet Process Prior," *Communications in Statistics, Part B—Simulation and Computation*, 23, 727–741.
- (1998), "Computational Methods for Mixture of Dirichlet Process Models," in *Practical Nonparametric and Semi-Parametric Bayesian Statistics*, eds. D. Dey, P. Muller, and D. Sinha, New York: Springer-Verlag, pp. 23–43.
- MacEachern, S. N., and Muller, P. (2000), "Efficient MCMC Schemes for Robust Model Extensions Using Encompassing Dirichlet Process Mixture Models," in *Robust Bayesian Analysis*, eds. D. R. Insua and F. Ruggeri, New York: Springer-Verlag, pp. 295–315.
- Mazzuchi, T. A., and Soyer, R. (1995), "A Bayesian Perspective on Some Replacement Strategies," *Reliability Engineering and System Safety*, 51, 295–303.
- (1996), "Adaptive Bayesian Replacement Strategies," in *Bayesian Statistics 5*, eds. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. Smith, Oxford, U.K.: Oxford University Press, pp. 667–674.
- Mukhopadhyay, S., and Gelfand, A. E. (1997), "Dirichlet Process Mixed Generalized Linear Models," *Journal of the American Statistical Association*, 92, 633–639.
- Short, T. H. (1993), "An Algorithm for the Detection and Measurement of Rail Surface Defects," *Journal of the American Statistical Association*, 88, 436–440.
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and van der Linde, A. (2002), "Bayesian Measures of Model Complexity and Fit," *Journal of the Royal Statistical Society, Ser. B*, 64, 583–639.
- Stanford, J., and Kalousek, J. (2000), "Preventative Gradual Rail Grinding Pays Dividends," *International Railway Journal*, 40, 25–26.
- Transportation Research Board (2000), Transportation Research Board A2M06 Committee Problem Statements, available at <http://www.nationalacademies.org/trb/publications/problems/a2m06ps.pdf>, last accessed March 18, 2002.