Modeling Spatial and Temporal Dependencies between Earthquakes

Lars Holden[1], Bent Natvig[1,2] Sigurd Sannan[1] and Hilmar Bungum[3]

[1]Norwegian Computing Center

P.O. Box 114 Blindern, N-0314 Oslo, Norway

[2] Department of Mathematics, University of Oslo, N-0316 Oslo, Norway

[3]NORSAR, P.O. Box 51, N-2027 Kjeller, Norway

Abstract

Two new different stochastic models for earthquake occurrence are discussed. Both models are focusing on the spatio-temporal interactions between earthquakes. The parameters of the models are estimated from a Bayesian updating of priors, using empirical data to derive posterior distributions. The first model is a marked point process model in which each earthquake is represented by its magnitude and coordinates in space and time. This model incorporates the occurrence of aftershocks as well as the build-up and subsequent release of strain. The second model is a hierarchical Bayesian space-time model in which the earthquakes are represented by potentials on a grid. The final ambition of the models is to make predictions on the occurrence of earthquakes.

1   INTRODUCTION

Earthquake forecasting in the strict sense with the exact prediction of the time, the location, and the magnitude of an earthquake has been a difficult area of research for several decades. One outcome of this research, however, is that we today know much more about why earthquake prediction is difficult (e.g., Kagan, 1997). This difficulty is in part tied to concepts such as self-similarity, criticality and nucleation processes: All earthquakes start small, and while we know much about the limits to growth, we do not know in sufficient details when and why a rupture stops before that.

In this paper we outline two new different stochastic models for earthquake occurrence which both are focusing on the spatio-temporal interactions between earthquakes. We believe that by including the increased knowledge of earthquake processes in more advanced stochastic models the prediction capabilities for earthquakes can be improved. Both models are based upon Bayesian approaches with user specified prior distributions for all parameters, while empirical data are used for deriving posterior distributions.

The first model is a marked point process model (e.g., Cressie, 1993) in which each earthquake is represented by its magnitude and coordinates in space and time. The marked point process model has the advantage that it represents each earthquake separately. This gives a fine resolution making it feasible to include known physical quantities connected to each single earthquake. At this point the model includes the effects of aftershocks as well as the build-up and subsequent release of strain.

The second model is a hierarchical Bayesian space-time model (see West and Harrison, 1997), in which the earthquakes are represented by potentials in a grid. The hierarchical model represents areal effects efficiently since the primary variables are related to properties of regions or grid blocks.

In both models there are many possible parametrizations. The principles behind the estimation and algorithms are independent of a particular parametrization. Another freedom of the models is the choice of prior distributions. If the choice of priors turns out to be controversial it is always possible to choose flat priors. This, however, gives less information and subsequently may lead to less precision in the predictions. The models are quite preliminary in that they solely are based upon data from an earthquake catalog and as such do not incorporate any known geophysical or geological information from the earthquake region.

It is a primary aim to develop the marked point process model further by including more detailed geological information such as faults and weakness zones, rupture characteristics and stress drops connected to the larger earthquakes in a catalog, tectonic processes in the region, etc. This model is a very flexible one and is well suited for the incorporation of such information and other kinds of changes. We believe that with such improvements the marked point process model will be a well suited tool for a better understanding of occurrence characteristics of earthquakes in various regions.

2   MARKED POINT PROCESS MODEL

Marked point processes are commonly used stochastic models for representing a finite number of events located in space and time. Earthquake occurrence can very well be fitted into a marked point process model. Each earthquake has, in addition to a location in space and time, parameters representing the magnitude and quite often also information about the earthquake fault lines. Point process models for earthquakes have previously been discussed by Vere-Jones (1995) and Ogata (1998). The model presented in this paper treats aftershocks in a similar fashion to the work by Ogata. In addition, the model takes into account the effect of strain build-up. The ultimate goal is to include as much as possible of known physical processes into the model.

2.1   THE MODEL

In our notation an earthquake is represented by $E = (x, M)$ and t, where $x = (x^1, x^2, x^3)$ is the hypocenter coordinates (x^1 = longitude, x^2 = latitude, x^3 = depth), M is the moment magnitude, and t is the time. An earthquake catalog $H_T = \{(E_i, t_i)\}_{t_i < T} = \{(x_i, M_i, t_i)\}_{t_i < T}$ consists of all observed earthquakes above a certain magnitude in a specified region, and in a given time period (T_0, T).

The two major assumptions made in the proposed model are:

- The intensity $\lambda_1(E, t | H_t, \beta)$ of earthquakes is a function of the parameters $(E, t) = (x, M, t)$, all previous earthquakes H_t in the region, and some parameters β to be determined by Bayesian updating. If additional data or physical information is available, this should be included in this intensity.

- The time averaged intensity $\lambda(E) = \lambda(x)\lambda(M|x)$ as a function of magnitude and location is known.

The time independent intensity $\lambda(x)$ represents the average number of earthquakes per unit time and unit volume. We have here estimated $\lambda(x)$ from a catalog. It may, however, be possible to estimate

$\lambda(x)$ from geological data of earthquakes, or a combination of such data and a catalog. This is believed to give a more stable estimate since a catalog may have completeness problems. In particular, it may not cover sufficiently many large earthquakes in each region to give the stability that is desired. This aside, it is a natural choice to let $\lambda(M|x)$ be determined by the well-known Gutenberg-Richter law for the distribution of magnitudes such that $\lambda(M|x) \propto 10^{-bM}$, where the value of the scaling parameter b usually is in the interval $(0.7, 1.2)$ (e.g., Vere-Jones, 1995).

We will assume that the intensity λ_1 is given by

$$\lambda_1(E, t | H_t, \beta) = \lambda_2(E|\beta) s(x, t | H_t, \beta) + \lambda_3(E, t | H_t, \beta), \tag{1}$$

where λ_2 is the background intensity, s represents the effect of strain build-up, and λ_3 represents the increase in the intensity after an earthquake used for modeling the aftershocks. The background intensity $\lambda_2(E|\beta)$ depends implicitly on the parameters β through the requirement that $\lambda_1(E, t | H_t, \beta)$ averaged over time equals $\lambda(E)$. If the strain build-up is omitted, $s = 1$, while if the aftershock treatment is omitted, $\lambda_3 = 0$. In the case that both $s = 1$ and $\lambda_3 = 0$ the model is just reduced to a simple Poisson model.

The intensity λ_3 is used to model the aftershocks. It can be used also to model foreshocks but we have not pursued this idea in the current paper. Let M_i be the magnitude of a shock in the catalog at the time t_i, and M the magnitude of a subsequent shock. Aftershocks M are then modeled by $\lambda_3(E, t | H_t, \beta) > 0$ for earthquakes $M < M_i$ for $t > t_i$. We will assume that λ_3 has the form

$$\lambda_3(E, t | H_t, \beta) = \sum_{(E_i, t_i) \in H_t} g(E, t, E_i, t_i, \beta), \tag{2}$$

where

$$g(E, t, E_i, t_i, \beta) = \beta_1 g_1(M, M_i, \beta_2) \, g_2(t, t_i, \beta_3, \beta_4) \, g_3(x, x_i, \beta_5). \tag{3}$$

The functions g_1, g_2, and g_3 represent magnitudial, temporal, and spatial effects, respectively. Note that the summation implies that if there is a large earthquake followed by a series of smaller earthquakes all of these earthquakes contribute to the intensity. A typical form of g_1 is $g_1(M, M_i, \beta_2) = \exp(\beta_2 M_i) \, 10^{-bM}$, in accordance with the Gutenberg-Richter law. This gives the same magnitude distribution as the time averaged intensity. For the temporal effect we assume that

$g_2(t, t_i, \beta_3, \beta_4) = 1/(t - t_i + \beta_4)^{\beta_3}$. The spatial effect can be represented by a function based on the distance between the hypocenters, *i.e.*, $g_3(x, x_i, \beta_5) = \exp(-\beta_5 ||x - x_i||^2)$.

It seems to be generally accepted that there is more regularity in the occurrence of earthquakes than can be accounted for in a Poisson model (Working Group on California Earthquakes Probabilities, 1995). The assumption is that in any particular region strain is slowly building up and then released due to earthquakes. This effect can be incorporated into a point process model. We first define a state variable S that can be connected to strain, or to stress for that matter. The interpretation of S may be different than the standard definition of strain (or stress), but its general nature will be such that it reflects the spatio-temporal release of strain/stress energy. For simplicity we refer to S as strain in the following. We define S by

$$S(x, t|H_t, \beta) = S_0(x, T_0) + \phi(x, \beta)(t - T_0) - \sum_{(E_i, t_i) \in H_t} h(x, E_i, \beta), \tag{4}$$

with

$$\phi(x, \beta) = \int h(x, E', \beta) \lambda(E') dE', \tag{5}$$

and

$$h(x, E', \beta) = \exp(\beta_7 M') \exp(-\beta_8 ||x - x'||^2). \tag{6}$$

Here ϕ represents the average strain build-up per unit time and h the release of strain for each earthquake. We have made the assumption that the mean strain release equals the mean strain build-up. The strain release h is factored into two terms related to the magnitude of the earthquake and a spatial effect, respectively. Thus, S represents the strain at any point (x, t) in space and time, given all the previous earthquakes contained in the catalog H_t. It is assumed that S builds up linearly and then decreases instantaneously with each earthquake. The strain has a variability that is independent of time, and s is an increasing function of S given by

$$s(x, t|H_t, \beta) = \exp(\beta_6 S). \tag{7}$$

It is implicitly given through this parametrization that the time-averaged intensity of earthquakes equals $\lambda(E)$. The effect of s is to reduce the variability in the time periods between very large earthquakes

compared to the simple Poisson model. The variability in time periods between large earthquakes becomes smaller and smaller with increasingly large values of β_6 and β_7. The parameter β_8 specifies the surrounding region of an earthquake in which strain is released.

2.2 POSTERIOR DISTRIBUTIONS FOR THE PARAMETERS

From the real catalog H_T of the period (T_0, T) it is possible to find the posterior distributions of the parameters β. These posterior distributions represent the best guesses for the parameters and should be used in all predictions. The posterior distributions for the parameters β, given the data in the catalog H_T, are defined by the equation

$$f(\beta|H_T) \propto f(\beta)f(H_T|\beta). \tag{8}$$

The likelihood $f(H_T|\beta)$ can be calculated from

$$
\begin{aligned}
f(H_T|\beta) &= \exp\left(-\int_{T_0}^{T}\int \lambda_1(E,t|H_t,\beta)dEdt\right)\prod_{i=1}^{n}\lambda_1(E_i,t_i|H_{t_i},\beta) \\
&\approx c(\beta)\prod_{i=1}^{n}\lambda_1(E_i,t_i|H_{t_i},\beta),
\end{aligned}
\tag{9}
$$

where the factor $\exp\left(-\int_{T_0}^{T}\int \lambda_1(E,t|H_t,\beta)dEdt\right)$ is approximated by a constant $c(\beta)$. This factor is due to periods (t_{i-1}, t_i) without earthquakes, while the factor $\prod_{i=1}^{n}\lambda_1(E_i,t_i|H_{t_i},\beta)$ represents the intensities for the actual earthquakes. The posterior distributions for the parameters are found by Markov chain Monte Carlo simulation from (8). In Markov chain Monte Carlo simulation one defines a chain of parameter sets that after a sufficient number of iterations satisfies the specified distributions. The maximum likelihood estimates for the parameters are the parameters that maximizes the expression (9).

2.3 SIMULATIONS AND PREDICTIONS

Simulation from the model is fairly straightforward. The simulated earthquakes are generated one at a time in chronological order, and the intensity for each new earthquake is given by (1). An example of a simulatin is given in the next section.

Predictions can be done by first sampling n values of the parameter set β from the distributions (8) using the most recent earthquake catalog H_t for the region. These parameter sets can be found by Markov chain Monte Carlo simulation. For each of these sets of parameters β new earthquakes are simulated based upon the intensity $\lambda_1(E, t|H_t, \beta)$, where H_t is continuosly updated during simulation. Probabilistic predictions are then obtained simply by counting the number of earthquakes in the different simulations. For short-term predictions and when the intensity is low it is also possible to use the intensities directly.

2.4   RESULTS

As a first step we have implemented an algorithm for estimating the maximum likelihoods of the parameters β. The empirical data we have used are based on an earthquake catalog over the time span $1932 - 1999$ compiled by the Southern California Earthquake Center, SCEC (2000), and limited to the region $31°–36°$N, $115°–120°$W. We have also limited the data to include only the epicenter coordinates x^1 = longitude and x^2 = latitude, thus neglecting the hypocenter coordinate x^3 = depth. In Fig. 1 (left) is shown the location of all quakes of magnitudes $M \geq 4.0$ for this region and the given time span. We have included all earthquakes of magnitude $M \geq 3.0$ in our data set, which comprises 15804 earthquakes over the time-span of 24837 days, giving an average intensity of 0.636 earthquakes per day. A simple time-magnitude plot in Fig. 1 (right) shows that the completeness seems to be reasonable back to 1942. The 5×5 degree area is divided into 1600 grid cells, with each cell corresponding to a size of about 14×12 km. The intensity $\lambda(E)$ for each grid cell is calculated from the empirical data. An average b-value of 0.93 for the Gutenberg-Richter relation is estimated from the data.

Using the 58-year period 1942–99 of the catalog, the estimation of the maximum likelihoods yields $\beta_1 = 0.4792, \beta_2 = 1.189, \beta_3 = 1.229, \beta_4 = 0.01867$, and $\beta_5 = 1938$ when the strain build-up is omitted and hence $s = 1$. In the calculation we have assumed that the contributions to $\lambda_3(E, t|H_t, \beta)$ for each quake go no more than 10 years back in time. Hence, the 10-year period 1932–41 of the empirical data is used indirectly in the calculation giving contributions to λ_3 for quakes in the period 1942–51. A detailed analysis of the parameters indicates that 60.3% of the quakes in the period 1942–99 are due to aftershock activity while the remaining 39.7% of the quakes are due to background
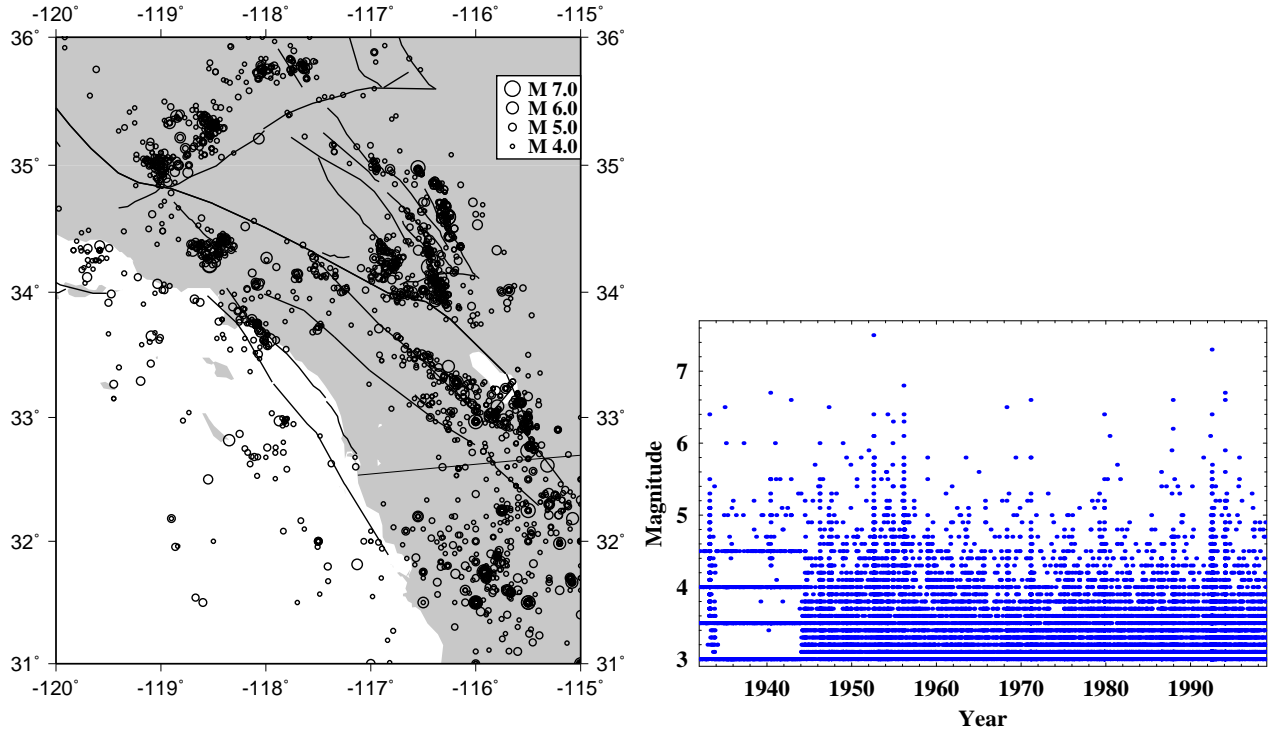
Figure 1: The location of earthquakes with $M \geq 4.0$ plotted together with the major faults (left), and the magnitude vs. time for earthquakes with $M \geq 3.0$ (right) in Southern California in the period $1932 - 1999$.

activity. An interpretation of the parameter β_2 is that the increased intensity due to an earthquake of magnitude $M = 5.6$ is twice that of a $M = 5.0$ earthquake. The values of β_3 and β_4 indicate that the increase of intensity is halved 20 minutes after an earthquake. Finally, the value of β_5 indicates that the increased intensity is halved at a distance of 1.9 km from the epicenter of the earthquake. Including the build-up of strain the calculation of the maximum likelihoods yields $\beta_1 = 0.4801, \beta_2 = 1.190, \beta_3 = 1.229, \beta_4 = 0.01867, \beta_5 = 1952, \beta_6 = 0.0004884, \beta_7 = 0$, and $\beta_8 = 0$. Turning on the strain build-up hence yields minor corrections to $\beta_1, \beta_2, \beta_3, \beta_4$, and β_5, while β_7 and β_8 both vanish. The direct interpretation of $\beta_7 = 0$ is that the build-up and release of strain (note that this is not strain energy) is independent of the magnitudes of the individual earthquakes, and that it is the number of earthquakes (the cumulative effect) that matters in this context. Indirectly, however, a large earthquake has a greater impact than a small one because it triggers a larger number of aftershocks.

Finding the parameter $\beta_8 = 0$ means that each earthquake in the catalog has a global effect on the build-up and release of strain. This is difficult to interpret physically but a plausible explanation of the result is that we so far in the model have used epicentral distances instead of the smallest distance to

8

the causative fault (the rupture plane). Another significant source of error may be related to our crude assumption that $S = S_0$ at the beginning and the end of the catalog. This assumption was made from simplicity and from the lack of knowledge about the actual state of strain in the various regions and different points of time. It is also clear that the earthquake catalog we have used in estimating the parameters of the model covers a too short time span (1932 – 1999) to carry sufficient information on the build-up and release of strain. Hopefully, in a refined model where these issues are taken better care of, the problem with the vanishing of β_8 will be resolved. The vanishing of β_7 suggests that this parameter, as given by the expression (6), is superfluous in the model. A better representation of the strain release might therefore be given by $h(x, E', \beta) = \exp(-\beta_7 \|x - x'\|^2 / 10^{M'})$, in accordance with the empirical fact that the spatial extension L of an earthquake usually behaves like $L \sim 10^{M/2}$ (implying self-similarity), except for larger earthquakes where $L \sim 10^{3M/4}$ (e.g., Stock and Smith, 2000). In this expression for $h(x, E', \beta)$ the old β_8 has been replaced by a new β_7, *i.e.*, advantageously there is one less parameter in the model to be estimated.
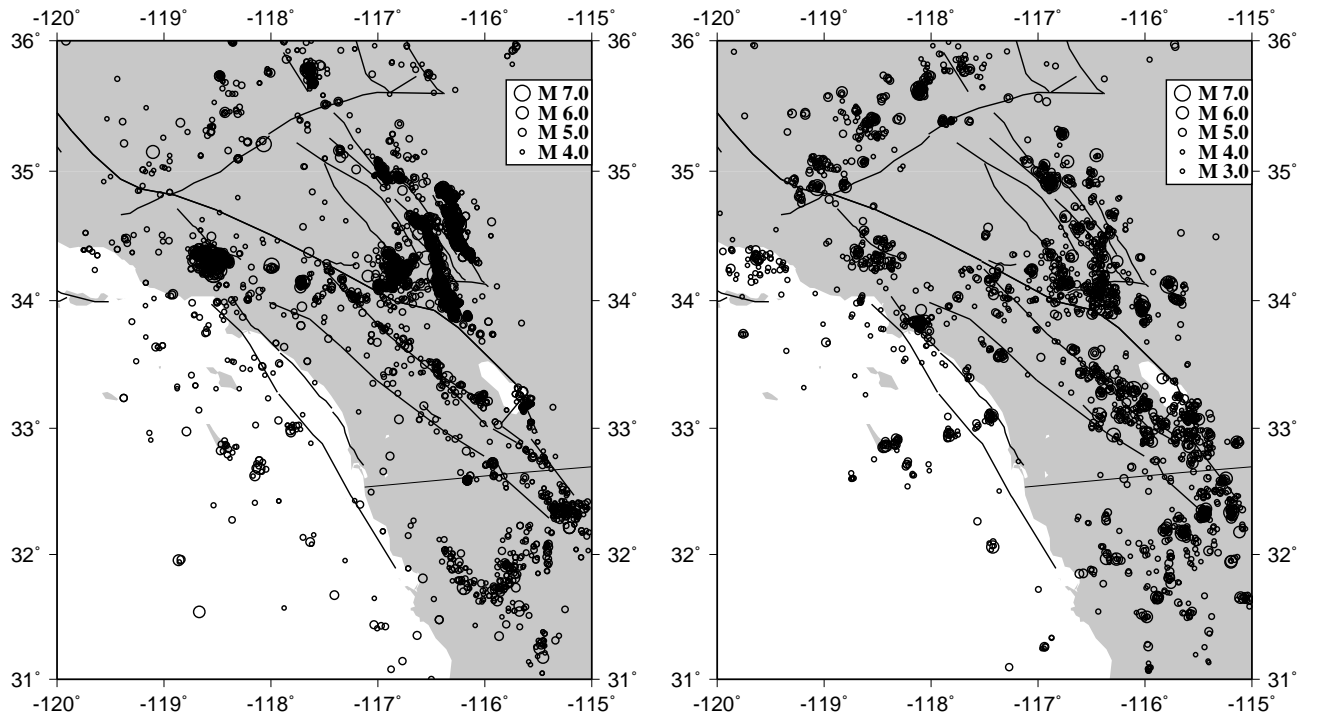

Figure 2: Earthquakes of $M \geq 3.0$ in the period 1990 – 1999 in Southern California (left), and simulated earthquakes over the same 10-year period using the marked point process model (right). The simulation is based on data from the time period 1932 – 1999.
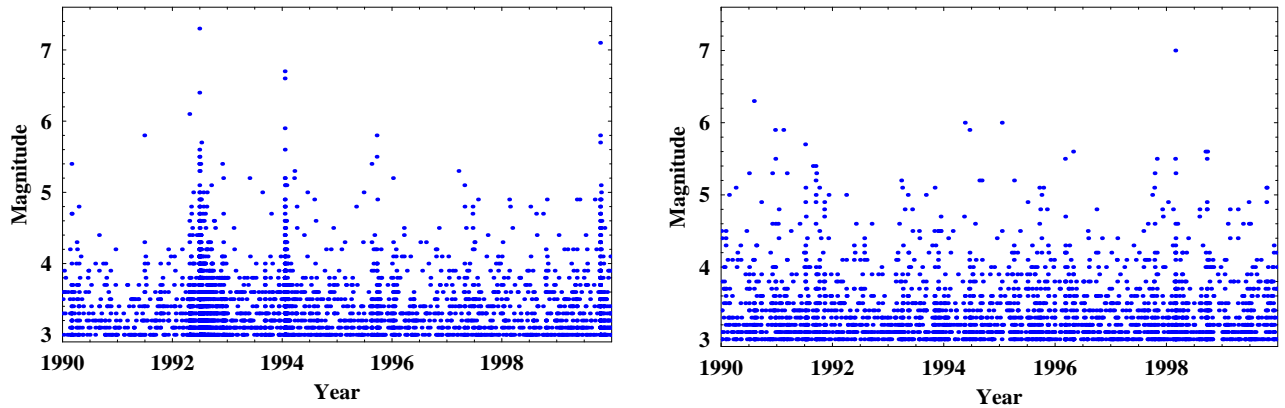
Figure 3: Magnitude vs. time of actual earthquakes in the period 1990 – 1999 (left), and simulated earthquakes over the same 10-year period (right) in Southern California.

At the next step we have implemented an algorithm for simulating earthquakes in the region of interest, given the estimated values of the β's. An example of a simulation over the 10-year period 1990–99 in the case of $s = 1$ is shown to the right in Figs. 2 and 3. This particular simulation thus does not take the build-up and release of strain into account but is a result of the background activity due to the 1932–99 catalog and aftershock activity due to the 1980–89 data. The simulation resulted in 2315 earthquakes over 3652 days, corresponding to 0.634 earthquakes per day. Of the 2315 earthquakes, 1419 of them, or 61.3%, were afterquakes. For a comparison with observed data, the corresponding data from the period 1990–99 are shown to the left in Figs. 2 and 3. Except for the fact that the period 1990–99 is a period of higher than average intensity, we see from Fig. 2 that the simulation produces earthquakes with a similar spatial distribution as the observed data. In addition it can be seen that the simulated catalog reflects features in the 1932–99 data that are not present in the 1990–99 data. The simulation results also display aftershock activity indicated by event sequences following the larger earthquakes as shown in Fig. 3 (right), although this effect is not as pronounced as for the observed data in Fig. 3 (left).

Simulations where the build-up and release of strain are taken into account produce results that are similar to those displayed in Figs. 2 and 3. This is not so surprising since we do not expect to see any real periodicity pattern on the occurrence of large earthquakes over the time-span of only 10 years. The return period for the largest earthquakes in this region is much longer than this. In Fig. 4, however, we have shown the typical behavior of the strain function $s(x, t|H_t, \beta)$ during a given simulation. The
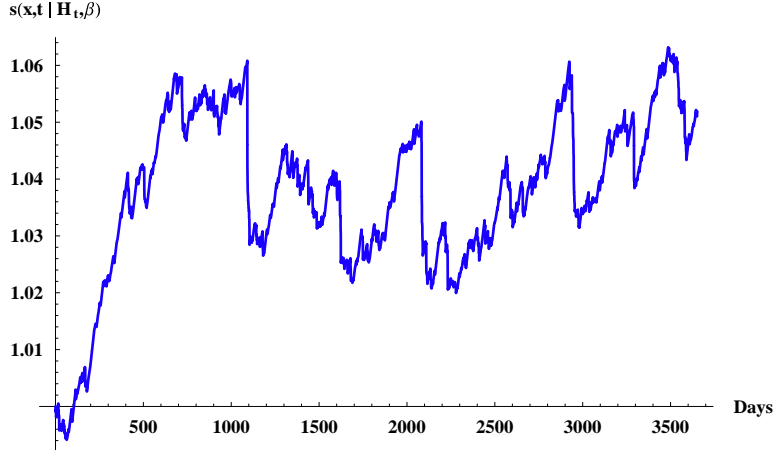
10

Figure 4: The strain function $s(x,t|H_t,\beta)$ for a typical simulation taking into account the build-up and release of strain. $s(x,t|H_t,\beta)$ increases when the activity is below the average background intensity and drops suddenly due to the aftershock activity following a large earthquake. The abrupt increase of $s(x,t|H_t,\beta)$ over the first 500 days is due to a low S_0 value.

sudden drops in the graph correspond to the release of strain due to the aftershocks following a large earthquake. To demonstrate the pure effect of the strain function $s(x,t|H_t,\beta)$ in our model, we have done simulations where the aftershocks treatment have been omitted by setting $\beta_1 = 0$, effectively giving $\lambda_3 = 0$. The other parameters are left unchanged, except that in the first simulation we set $\beta_6 = 0.0004708$ (unchanged), while in the second simulation we take β_6 to be a factor 1000 higher. The effect of varying β_6 in this manner is shown in Fig. 5 where the graphs show the frequency of earthquakes for each 30-day period of the simulation period. A greater value of β_6 yields a more even release of strain in each time interval, and hence a more even frequency of earthquakes in each 30-day period. The model is not particularly sensitive to the value of β_6 since it was necessary to multiply β_6 with a factor 1000 to achieve a significant difference. However, an improved estimate of the parameter β_8 may imply a greater sensitivity to β_6.
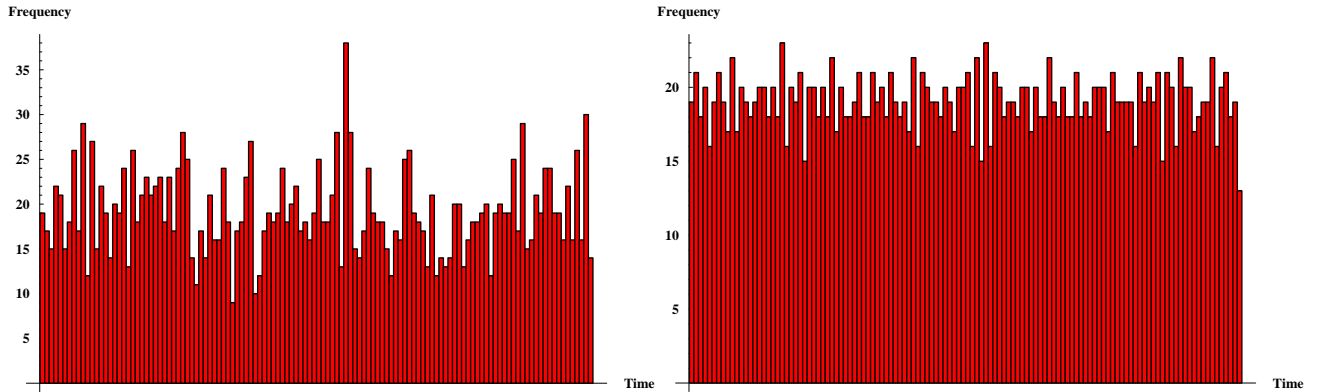
11

Figure 5: The frequency of earthquakes in intervals of 30 days for a given simulation. The left figure shows the frequency when the parameter $\beta_6 = 0.0004708$, while to the right is shown the frequency when $\beta_6 = 0.4708$.

2.5 DISCUSSION AND CONCLUSIONS

In conclusion we have by maximum likelihood optimization estimated the parameters β of the marked point process model, using earthquake data from Southern California from the period $1932 - 1999$. The first 5 parameters $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5$ are connected to the term λ_3 and represent the aftershock activity. The estimated values for these parameters seem reasonable. For instance, we may compare the results $\beta_3 = 1.229$ and $\beta_4 = 0.01867$ with the corresponding estimates found by Ogata (1998) in various extensions of his Epidemic Type Aftershock-Sequences model. Applied to data from two different districts of Japan, Ogata estimated a parameter p (corresponding to β_3 in our model) to be in the range 0.900–1.136 and a parameter c (corresponding to β_4) to be in the range 0.00172–0.0357. The last 3 parameters $\beta_6, \beta_7, \beta_8$ are connected to the factor s, which represents the build-up and release of strain in the model. Surprisingly, we found both β_7 and β_8 to vanish in the maximum likelihood estimation. The vanishing of β_7 suggests that the release of strain locally is independent of the magnitude of the earthquakes. The release of total strain energy, however, does depend on the size of the earthquake since a large earthquake causes release of strain over a much larger area than a small earthquake. Because of this, the parameter β_7 seems superfluous in the model and could be omitted by a slightly different representation of the strain release, as indicated earlier. The vanishing of β_8 in our estimation is difficult to interpret physically, however. This implies that all earthquakes in the catalog have a global

12

effect on the release of strain for the region considered. This in not what we would have expected and suggests that the given parametrization has been too crude to handle the spatial dependencies between the release of strain and the individual earthquakes. One explanation for this may simply be that we have used epicentral distances instead of the distances to the rupture plane. Clearly, more work needs to be done on this point.

At the current stage we have done simulations from the model which demonstrates that it is able to display both the spatio- and temporal interactions between earthquakes. In order for the model to generate as realistic earthquake patterns as possible, however, it is desirable to incorporate as much of the seismological, geological and geodetic knowledge as possible in the model. To this end it is important to keep in mind that the main aim of these efforts is to provide a probabilistic modeling of earthquakes based on present multidisciplinary understanding and knowledge, and not to replace this knowledge by a stochastic model. By use of Bayesian techniques it is possible to use empirical data to quantify the relationships between different physical measurements. Some of the most important potential extensions of the current marked point process model would be:

Extended earthquake source representations
By making a representation of the entire earthquake rupture plane and seismic moment release over the fault instead of just the hypocenter and the magnitude, it should be possible to delineate in more details the larger earthquakes and thereby also get a better representation of the spatial distributions of aftershocks and strain release connected to each earthquake.

Mapped faults and other structural geological features
Using mapped faults as an additional delineation of earthquake potentials will provide more realistic estimates of local intensities and orientations of earthquake rupture characteristics. The fault orientations may, in a preliminary model, be considered as having a static probability distribution depending simply on the location of the epicenter. It may also be possible to make it time dependent, depending also on the value and the orientation of the local strain along the mapped faults. A realistic strain build-up model should be related specifically to the mapped faults, incorporating both long-term and short-term geodetic information.

Strain release and build-up after earthquakes

An earthquake results in the release of strain. However, it may also increase the strain in certain regions depending on the character of the earthquake rupture and the local structural geology. The incorporation of a realistic strain release (and build-up) model should also consider potential regions for aseismic slip.

Improved triggering models and spatially dependent parameters

An improved modeling of potential triggering connections will be useful. The model can also be improved by using parameters that are spatially dependent, *i.e.*, both the βs and the b-value from the Gutenberg-Richter law may depend on the location, as well as vary with time.

Use data from before the beginning of a catalog

The average time independent intensities and perhaps also the strain build-up models may be improved by using in particular paleoseismic information, addressing in particular uncertainties related to potential 'earthquake deficit' situations. It would also be possible to extend the catalog back in time based on other types of simulations.

Use of other kinds of geo-information

One may let the intensities depend on other kinds of data of potential importance for earthquake occurrence. By using a Bayesian technique it is only necessary to provide the data and specify a prior distribution for the relationship between the measurement and the intensities. A posterior distribution is then derived from the data.

3   HIERARCHICAL BAYESIAN SPACE-TIME MODEL

The basic theory for Bayesian forecasting and dynamic models is presented in the book West and Harrison (1997). A part of this theory is known in the engineering community as Kalman filtering. In the excellent paper Wikle, Berliner and Cressie (1998) this basic theory is used to develop hierarchical Bayesian space-time models, and it is suggested that this gives more flexible models and methods for the analysis of environmental data distributed in space and time. The model is implemented in a Markov chain Monte Carlo simulation framework using Gibbs sampling and applied to an atmospheric

data set of monthly maximum temperatures.

Inspired by the work above we will in the following present a hierarchical Bayesian space-time model to be applied to an earthquake catalog. Although Vere-Jones (1995) is very much concerned with a marked point process approach to earthquake prediction he is stating at the end of the paper: " ... , and it is possible that something like a hidden Markov model (for the strain regime) could provide a more flexible modeling framework within which such precursory features could be explored." We feel that the present approach is just a full consequence of this line of thought.

3.1   THE MODEL

In the following, let an index x denote space location and t denote time. x and t will be discrete. Typically, space locations will be situated on a grid where the distance between nearest grid points is of the order 20 kilometers whereas the distance between points in time is of the order 10 days.

We now introduce the following unobserved (hidden) system state variable:

$\theta_t^x =$ earthquake potential at space location x and time t

It is natural to consider θ_t^x as an unknown, continuously distributed random quantity. If it were discretely distributed, a hidden Markov model approach might have been possible. Furthermore, introduce the observation corresponding to θ_t^x:

$M_t^x =$ observed earthquake magnitude at space location x and time t

We also introduce a corresponding noise (shock) term:

$\widehat{M}_t^x =$ noise (shock) at space location x and time t

In the following, all noise terms, with generic notation $\widehat{N}_t^x$, are independent in space and time, each having a normal distribution $N(0, \sigma_{\widehat{N}}^2)$, where $\sigma_{\widehat{N}}^2$ is a random quantity. These noise terms are also assumed to be totally independent of all other noise terms.

The basic link between the earthquake potential and the observed earthquake magnitude at space loca-

tion x and time t is given by the so-called *observation equation*:

$$M_t^x = (\theta_t^x + \widehat{M_t^x})I(\theta_t^x + \widehat{M_t^x} \geq k) , \qquad t = 1, 2, \ldots , \tag{10}$$

where for an event E

$$I(E) = \begin{cases} 1 & \text{if } E \text{ occurs} \\ 0 & \text{otherwise} \end{cases}$$

Here k is a detection limit, being for instance $M 4.0$. k may be replaced by k_t, which depends on time. The main reason for such a sophistication is a change in data quality with time. The basic idea behind the observation equation (10) is simple. If the sum of the potential and shock term at space location x and at time t exceeds the detection limit, we observe an earthquake at x and t with magnitude equal to this sum. If the sum does not exceed the detection limit, nothing is observed. Since M_t^x has a distribution involving truncation from below at k, standard Bayesian dynamic model theory cannot be applied and we have to rely on efficient Markov chain Monte Carlo simulation techniques.

The earthquake potential at x and t is now assumed to be decomposed in the following way:

$$\theta_t^x = \mu^x + \theta_{ST,t}^x + \theta_{S,t}^x + \theta_{LT,t}^x + \widehat{\theta_t^x} , \qquad t = 1, 2, \ldots \tag{11}$$

Here μ^x is a time independent contribution, $\theta_{ST,t}^x$ a short term contribution, $\theta_{S,t}^x$ a strain contribution and $\theta_{LT,t}^x$ a long term contribution. $\widehat{\theta_t^x}$ is a noise term linked to the system state θ_t^x.

We now give the so-called *system equations* for respectively the short term contribution, the strain contribution and the long term contribution. These equations link the system state variables at the time t to the corresponding variables at the previous point in time $t - 1$ and to information on observed earthquakes at $t - 1$.

We start by the short term contribution. Introduce the notation $\mathrm{N} = $ north, $\mathrm{E} = $ east, $\mathrm{S} = $ south, $\mathrm{W} = $ west, $\mathrm{A} = $ above, $\mathrm{B} = $ below. Furthermore, let $\mathrm{N}(x) = $ location north of x (if any) and define correspondingly $\mathrm{E}(x), \mathrm{S}(x), \mathrm{W}(x), \mathrm{A}(x), \mathrm{B}(x)$. Extending the corresponding contribution in the model of Wikle, Berliner and Cressie (1998) to 3-dimensional space, we have

$$\begin{aligned} \theta_{ST,t}^x = {}& a^x \theta_{ST,t-1}^x + a_\mathrm{N} \theta_{ST,t-1}^{\mathrm{N}(x)} + a_\mathrm{E} \theta_{ST,t-1}^{\mathrm{E}(x)} \\ & + a_\mathrm{S} \theta_{ST,t-1}^{\mathrm{S}(x)} + a_\mathrm{W} \theta_{ST,t-1}^{\mathrm{W}(x)} + a_\mathrm{A} \theta_{ST,t-1}^{\mathrm{A}(x)} + a_\mathrm{B} \theta_{ST,t-1}^{\mathrm{B}(x)} + \widehat{\theta_{ST,t}^x} , \qquad t = 1, 2, \ldots \end{aligned} \tag{12}$$

Hence the short term contribution to the system state at x and t is linked linearly to the corresponding contribution to the system state at time $t - 1$ and to the contributions to the system states of its nearest neighbors at time $t - 1$. We assume $(x = 1, \ldots, S)$:

$$\theta_{ST,0}^x \sim N[\widetilde{\theta}_{ST,0}^x, \widetilde{\sigma}_{\theta_{ST,0}^x}^2]$$

The $\theta_{ST,0}^x$s are assumed independent with expectations and variances to be considered as fixed parameters that must be specified.

We next consider the strain contribution given by

$$
\begin{aligned}
\theta_{S,t}^x &= (\theta_{S,t-1}^x + \beta_{t-1}^x)I(M_{t-1}^x = 0) + \widehat{\theta}_{S,t}^x , \qquad t = 1, 2, \ldots \\
\beta_t^x &= \beta^x + (\beta_{t-1}^x - \beta^x)I(M_{t-1}^x = 0) + \widehat{\beta}_t^x , \qquad t = 0, 1, \ldots ,
\end{aligned}
\tag{13}
$$

where we set $I(M_0^x = 0) = 1$ and $I(M_{-1}^x = 0) = 0$. β_{t-1}^x is an incremental growth term (positive or negative) added to the strain at x and $t-1$ to give the strain at x and t, if there is no observed earthquake at x and $t - 1$. If such an earthquake has occurred, we assume that all strain has been released, and hence the strain at x and t is assumed to be zero and a new build-up of strain is started. We model the distributions of the $\theta_{S,0}^x$s and β^xs completely parallel to the ones of the $\theta_{ST,0}^x$s given above.

Since $\theta_{S,t}^x$ and β_t^x depend on the observation M_{t-1}^x in addition to the dependence on $\theta_{S,t-1}^x$ and β_{t-1}^x again standard Bayesian dynamic model theory cannot be applied enforcing the reliance on efficient Markov chain Monte Carlo simulation techniques (see West and Harrison 1997, p. 565).

Finally, consider the long term contribution, which might be excluded from the model if it is not supported by the earthquake catalog. Let

$$p^x = \quad \text{the period of the long term cycle at space location } x$$

Then, assuming p^x to be odd, it follows from West and Harrison (1997), sections 8.6.3, 8.6.4, 8.6.6, that we have the following reduced Fourier form model:

$$
\begin{aligned}
\theta_{LT,t}^x &= \cos(2\pi/p^x)\theta_{LT,t-1}^x + \sin(2\pi/p^x)\alpha_{t-1}^x + \widehat{\theta}_{LT,t}^x \\
\alpha_t^x &= -\sin(2\pi/p^x)\theta_{LT,t-1}^x + \cos(2\pi/p^x)\alpha_{t-1}^x + \widehat{\alpha}_t^x , \qquad t = 1, 2, \ldots
\end{aligned}
\tag{14}
$$

$\theta_{LT,t}^x$ and α_t^x are interpreted as Fourier coefficients at x and t. The latter is only used to update the former. It should be noted that (14) is more general than the corresponding contribution in the model of Wikle, Berliner and Cressie (1998), where we would get

$$\theta_{LT,t}^x = \theta_{LT,0}^x \cos((2\pi/p)t) + \alpha_0^x \sin((2\pi/p)t) \,, \qquad t = 1, 2, \dots \,,$$

having $p^x = p$ for all x and no noise terms to ensure a random development in time.

Wikle, Berliner and Cressie (1998) introduce simple spatial trends in various parts of their modeling due to results of some exploratory data analyses. We will not make such restrictions in our model. We simply model the distributions of the $\theta_{LT,0}^x$s and α_0^xs completely parallel to the ones of the $\theta_{ST,0}^x$s given above.

We now consider the time-independent term, μ^x, in (11). Extending the corresponding contribution in the model of Wikle, Berliner and Cressie (1998) to 3-dimensional space, we have the following Markov Random Field model:

$$\mu^x \sim \mathrm{N}[\mu_0^x + b_{\mathrm{N,S}}\{\mu^{\mathrm{N}(x)} - \mu_0^{\mathrm{N}(x)} + \mu^{\mathrm{S}(x)} - \mu_0^{\mathrm{S}(x)}\}$$
$$b_{\mathrm{E,W}}\{\mu^{\mathrm{E}(x)} - \mu_0^{\mathrm{E}(x)} + \mu^{\mathrm{W}(x)} - \mu_0^{\mathrm{W}(x)}\} \qquad (15)$$
$$b_{\mathrm{A,B}}\{\mu^{\mathrm{A}(x)} - \mu_0^{\mathrm{A}(x)} + \mu^{\mathrm{B}(x)} - \mu_0^{\mathrm{B}(x)}\}, \sigma_\mu^2] \,,$$

where σ_μ^2 is a random quantity. Furthermore $(x = 1, \dots, S)$:

$$\mu_0^x \sim N(\widetilde{\mu}_0^x, \widetilde{\sigma}_{\mu_0^x}^2)$$

The μ_0^xs are assumed independent with expectations and variances to be considered as fixed parameters that must be specified. Finally, the spatial dependence parameters $b_{\mathrm{N,S}}, b_{E,W}$ and $b_{\mathrm{A,B}}$ are assumed independent and normally distributed with expectations respectively equal to $\widetilde{b}_{\mathrm{N,S}}, \widetilde{b}_{\mathrm{E,W}}$ and $\widetilde{b}_{\mathrm{A,B}}$ and variances respectively equal to $\widetilde{\sigma}_{b_{\mathrm{N,S}}}^2, \widetilde{\sigma}_{b_{\mathrm{E,W}}}^2$ and $\widetilde{\sigma}_{b_{\mathrm{A,B}}}^2$. These parameters are again fixed and must be specified.

Returning to (12), a^x is modelled completely parallel to μ^x in (15) as a Markov Random Field. Hence,

$$a^x \sim \mathrm{N}[a_0^x + a_{\mathrm{N,S}}\{a^{\mathrm{N}(x)} - a_0^{\mathrm{N}(x)} + a^{\mathrm{S}(x)} - a_0^{\mathrm{S}(x)}\}$$
$$+ a_{\mathrm{E,W}}\{a^{\mathrm{E}(x)} - a_0^{\mathrm{E}(x)} + a^{\mathrm{W}(x)} - a_0^{\mathrm{W}(x)}\} \qquad (16)$$
$$+ a_{\mathrm{A,B}}\{a^{\mathrm{A}(x)} - a_0^{\mathrm{A}(x)} + a^{\mathrm{B}(x)} - a_0^{\mathrm{B}(x)}\}, \sigma_a^2]$$

Having introduced a proper notation, the rest is straightforward. Furthermore, we assume that the nearest neighbor parameters in (12) $a_\mathrm{N}, a_\mathrm{E}, a_\mathrm{S}, a_\mathrm{W}, a_\mathrm{A}$ and a_B are independent and normally distributed with expectations respectively equal to $\widetilde{a}_\mathrm{N}, \widetilde{a}_\mathrm{E}, \widetilde{a}_\mathrm{S}, \widetilde{a}_\mathrm{W}, \widetilde{a}_\mathrm{A}$ and $\widetilde{a}_\mathrm{B}$ and variances respectively equal to $\widetilde{\sigma}^2_{a_\mathrm{N}}, \widetilde{\sigma}^2_{a_\mathrm{E}}, \widetilde{\sigma}^2_{a_\mathrm{S}}, \widetilde{\sigma}^2_{a_\mathrm{W}}, \widetilde{\sigma}^2_{a_\mathrm{A}}$ and $\widetilde{\sigma}^2_{a_\mathrm{B}}$. Again these parameters are fixed and must be specified.

Finally, following Wikle, Berliner and Cressie (1998) the random variances $\sigma^2_{\widehat{M}}, \sigma^2_{\widehat{\theta}}, \sigma^2_{\widehat{\theta}_{ST}}, \sigma^2_{\widehat{\theta}_S}, \sigma^2_{\widehat{\beta}}, \sigma^2_{\widehat{\theta}_{LT}}, \sigma^2_{\widehat{\alpha}}, \sigma^2_\mu$ and σ^2_a are assumed independent and inverse gamma (IG) distributed with parameters respectively equal to $(\widetilde{q}_{\sigma^2_{\widehat{M}}}, \widetilde{r}_{\sigma^2_{\widehat{M}}}), \ldots , (\widetilde{q}_{\sigma^2_a}, \widetilde{r}_{\sigma^2_a})$. These parameters are fixed and must be specified.

The basic approach now is to start out from observations:

$$M_t^x , \qquad t = 1, \ldots , T ; \qquad x = 1, \ldots , S .$$

Here S is the number of spatial locations on the grid. These observations are then used to find the corresponding uncertainty, measured by posterior probability distributions, in the system state variables:

$$\theta_t^x , \qquad t = 1, \ldots , T ; \qquad x = 1, \ldots , S .$$

Finally, we arrive at predictions on:

$$\theta_t^x , \qquad t = T + 1, T + 2, \ldots ; \qquad x = 1, \ldots , S$$
$$M_t^x , \qquad t = T + 1, T + 2, \ldots ; \qquad x = 1, \ldots , S$$

Details on the simulation approach are given in Section 3.3 and in the Appendix.

3.2   IMPLEMENTATION

As has already been pointed out our model and simulation approach is very parallel to the one of Wikle, Berliner and Cressie (1998). Their atmospheric data set of monthly maximum temperatures covered 20 years (1974–1993) measured at 131 stations. Hence, in our notation $T = 240$, $S = 131$. With this size of the data set they had no problems with the implementation of the Gibbs sampling. It should be mentioned that they used an exploratory data analysis to assess the fixed parameters that must be specified in the prior distributions. This means that the data set is used twice, both à priori and à posteriori, which is not in accordance with a strict Bayesian approach.

For our earthquake catalog from Southern California (see Fig. 1) we suggest starting out in 1932 with non-informative prior distributions and run the Gibbs sampling for data until and including 1970; i.e. covering 39 years. We then use the posterior distributions based on this run as the prior distributions in the next run starting out in 1971 covering 25 years until and including 1995. Then we use the posterior distributions from the second run to make earthquake predictions for 1996, 1997, 1998 and 1999. These predictions are then compared with the observed earthquake data for the same years.

To make the data set manageable we suggest as a start a distance between time points for the first run to be a month, giving $T_1 = 12 \cdot 39 = 468$, and half a month for the second run, giving $T_2 = 24 \cdot 25 = 600$. We also suggest starting out dropping the depth dimension and consider a grid of $S = 25 \times 25 = 625$ planar locations with distance between nearest grid points equal to 20 kilometer. Furthermore, as a start we suggest dropping the long term contribution $\theta_{LT,t}^x$ in (11) since this is the most speculative one due to the need of assessing, p^x, the period of the long term cycle at space location x. Finally, for simplicity we suggest starting out with a detection limit of $k = $ M4.0 for the first run and $k = $ M3.0 for the second run due to improved data quality.

3.3 SIMULATION

As in Wikle, Berliner and Cressie (1998) the Markov chain Monte Carlo simulation technique to be used is Gibbs sampling; see Casella and George (1992) and Gilks et al. (1996). With generic notation, let $\vec{N}$ be an $S \times 1$ vector of the random quantities N^x, $x = 1, \ldots , S$. Also following Wikle, Berliner and Cressie (1998) for random U and V, let $[U \mid V]$ represent the conditional density of U, given V. Furthermore, let D_0 be our initial prior information at $t = 0$, and let $D_t = \{\vec{M}_t, \vec{M}_{t-1}, \ldots , \vec{M}_1, D_0\}$. Our first aim is to get samples from the posterior distribution $[\vec{\theta}_1, \ldots , \vec{\theta}_T \mid D_T]$. Such samples follow immediately by taking out samples of $(\vec{\theta}_1, \ldots , \vec{\theta}_T)$ from samples from the corresponding posterior distribution of all random quantities involved in the model until $t = T$. To arrive at the latter samples by Gibbs sampling full conditional distributions of any random quantity, given all the other and D_T, are needed. In general, full conditional distributions are determined by writing the full joint distribution of all random quantities and the observations, divided by the appropriate normalizing constants. In hierarchical models, such as the one dealt with in the present paper, this process is typically simplified

due to the various conditional independence assumptions that are made. In particular, all components of the full joint distribution that do not depend on the specific random quantity, whose full conditional distribution is considered, "cancel" from the numerator and denominator of the latter distribution. The derivations in Wikle, Berliner and Cressie (1998) and in the present paper begin after such simplifications.

Having samples from the posterior distribution of all random quantities in the model until $t = T$, given D_T, samples of the additional random quantities at $t = T + 1$ are arrived at by applying (11)–(14). Especially, we get samples of $\vec{\theta}_{T+1}$, which can be used for prediction. Furthermore, having samples of $\vec{\theta}_{T+1}$ along with samples of $\sigma^2_{\widehat{M}}$, samples of $\vec{M}_{T+1}$ are arrived at by applying (10). Again these samples can be used for prediction. Proceeding like this samples of $\vec{\theta}_t$, $\vec{M}_t$, $t = T + 2, T + 3, \ldots$ are arrived at.

Since, as in Wikle, Berliner and Cressie (1998) we have chosen Gaussian distributions with conjugate priors, the derivation and implementation of the full conditional distributions needed for Gibbs sampling are mostly straightforward. One exception concerns the full conditional distribution for each of the Markov Random Field spatial dependence parameters $a_{N,S}, a_{E,W}, a_{A,B}, b_{N,S}, b_{E,W}, b_{A,B}$. Exactly, as in the latter paper we employ a Metropolis-Hastings step in the Gibbs sampling; see Chib and Greenberg (1995) and Gilks (1996). In the Appendix all these full conditional distributions are given and implemented. Furthermore, due to the Gaussian distributions, the antithetic coupling of two Gibbs sampling chains given in Frigessi, Gåsemyr and Rue (2000) could be applied to increase the speed of the algorithm.

To illustrate the derivations of the full conditional distributions, we will consider the ones of $\vec{\theta}_t$, $t = 1, \ldots, T$, $\vec{\theta}_{S,t}$, $t = 0, \ldots, T$ and $\sigma^2_{\widehat{M}}$. These are the most challenging ones due to the apparent complex relations (10) and (13), which are key relations in our earthquake prediction modeling. The generic notation $[U \mid \cdot]$ is used to represent the full conditional distribution of U given all the other random quantities until $t = T$, and D_T.

We start by establishing $[\vec{\theta}_t \mid \cdot]$, $t = 1, \ldots, T$. Let φ be the cumulative distribution function of the

$N(0,1)$ distribution and let

$$N_k(\theta_t^x, \sigma_{\widehat{M}}^2) = I(M_t^x = 0)\varphi((k - \theta_t^x)/\sigma_{\widehat{M}}) + I(M_t^x \geq k)N(\theta_t^x, \sigma_{\widehat{M}}^2)$$

be the mixture distribution having a probability mass $\varphi((k-\theta_t^x)/\sigma_{\widehat{M}})$ in 0 and a $N(\theta_t^x, \sigma_{\widehat{M}}^2)$ contribution for values not less than k. From (10) we have

$$[\vec{M}_t \mid \vec{\theta}_t, \sigma_{\widehat{M}}^2] \sim \prod_{x=1}^{S} N_k(\theta_t^x, \sigma_{\widehat{M}}^2) \,, \qquad t = 1, 2, \ldots$$

Furthermore, from (11) we have

$$[\vec{\theta}_t \mid \vec{\mu}, \vec{\theta}_{ST,t}, \vec{\theta}_{S,t}, \vec{\theta}_{LT,t}, \sigma_\theta^2] \sim \prod_{x=1}^{S} N(\mu^x + \theta_{ST,t}^x + \theta_{S,t}^x + \theta_{LT,t}^x, \sigma_\theta^2) \,, \qquad t = 1, 2, \ldots$$

By standard algebraic manipulations, we then get, $t = 1, \ldots, T$:

$$[\vec{\theta}_t \mid \cdot] \propto [\vec{M}_t \mid \vec{\theta}_t, \sigma_{\widehat{M}}^2][\vec{\theta}_t \mid \vec{\mu}, \vec{\theta}_{ST,t}, \vec{\theta}_{S,t}, \vec{\theta}_{LT,t}, \sigma_\theta^2]$$

$$\propto \prod_{x=1}^{S} [I(M_t^x = 0)\varphi((k - \theta_t^x)/\sigma_{\widehat{M}}) + I(M_t^x \geq k)\exp(-(M_t^x - \theta_t^x)^2/2\sigma_{\widehat{M}}^2)/(\sqrt{2\pi}\,\sigma_{\widehat{M}})]$$

$$\times [\exp(-(\theta_t^x - \mu^x - \theta_{ST,t}^x - \theta_{S,t}^x - \theta_{LT,t}^x)^2/2\sigma_\theta^2)]$$

$$\propto \prod_{x=1}^{S} [I(M_t^x = 0)\varphi((k - \theta_t^x)/\sigma_{\widehat{M}})\exp(-(\theta_t^x - \mu^x - \theta_{ST,t}^x - \theta_{S,t}^x - \theta_{LT,t}^x)^2/2\sigma_\theta^2)$$

$$+ I(M_t^x \geq k)\exp(-(\theta_t^x)^2(1/\sigma_{\widehat{M}}^2 + 1/\sigma_\theta^2)/2)\exp(\theta_t^x(M_t^x/\sigma_{\widehat{M}}^2 + (\mu^x + \theta_{ST,t}^x + \theta_{S,t}^x + \theta_{LT,t}^x)/\sigma_\theta^2)]$$

$$\propto \prod_{x=1}^{S} [I(M_t^x = 0)\varphi((k - \theta_t^x)/\sigma_{\widehat{M}})\exp(-(\theta_t^x - \mu^x - \theta_{ST,t}^x - \theta_{S,t}^x - \theta_{LT,t}^x)^2/2\sigma_\theta^2)$$

$$+ I(M_t^x \geq k)N((1/\sigma_{\widehat{M}}^2 + 1/\sigma_\theta^2)^{-1}(M_t^x/\sigma_{\widehat{M}}^2 + (\mu^x + \theta_{ST,t}^x + \theta_{S,t}^x + \theta_{LT,t}^x)/\sigma_\theta^2), (1/\sigma_{\widehat{M}}^2 + 1/\sigma_\theta^2)^{-1})]$$

Note that this result implies that θ_t^x, $x = 1, \ldots, S$ are conditionally independent given all random quantities except for $\vec{\theta}_t$, and D_T.

Since no explicit expression for $\varphi((k - \theta_t^x)/\sigma_{\widehat{M}})$ exists, a normalizing constant for the case $M_t^x = 0$ cannot be found. Hence, this is not a straightforward distribution to sample from. However, $\varphi((k - \theta_t^x)/\sigma_{\widehat{M}})$ can be computed numerically by a standard routine and hence we can employ a Metropolis-Hastings step in the Gibbs sampling. As a proposal distribution we use the one above corresponding to $k = -\infty$, i.e. with no truncation in (10), which is just a product of normal distributions.

A probably better method for each case $M_t^x = 0$ has been suggested by our colleague Jørund Gåsemyr. First draw a fictitious earthquake magnitude, Y_t^x, from the conditional distribution of such a magnitude less than the detection limit. Secondly, draw a θ_t^x from the same normal distribution used for the case $M_t^x \geq k$, but with M_t^x replaced by Y_t^x.

We next turn to $[\vec{\boldsymbol{\theta}}_{S,t} \mid \cdot]$, $t = 1, \ldots, T-1$. From (11) and (13) we have

$$[\vec{\boldsymbol{\theta}}_{S,t} \mid \cdot] \propto [\vec{\boldsymbol{\theta}}_t \mid \vec{\boldsymbol{\mu}}, \vec{\boldsymbol{\theta}}_{ST,t}, \vec{\boldsymbol{\theta}}_{S,t}, \vec{\boldsymbol{\theta}}_{LT,t}, \sigma_{\hat{\theta}}^2][\vec{\boldsymbol{\theta}}_{S,t+1} \mid \vec{\boldsymbol{\theta}}_{S,t}, \vec{\boldsymbol{\beta}}_t, \vec{\boldsymbol{M}}_t, \sigma_{\hat{\theta}_S}^2][\vec{\boldsymbol{\theta}}_{S,t} \mid \vec{\boldsymbol{\theta}}_{S,t-1}, \vec{\boldsymbol{\beta}}_{t-1}, \vec{\boldsymbol{M}}_{t-1}, \sigma_{\hat{\theta}_S}^2]$$

$$\propto \prod_{x=1}^{S}[\exp(-(\theta_t^x - \mu^x - \theta_{ST,t}^x - \theta_{S,t}^x - \theta_{LT,t}^x)^2/2\sigma_{\hat{\theta}}^2)$$

$$\times \exp(-(\theta_{S,t+1}^x - (\theta_{S,t}^x + \beta_t^x)I(M_t^x = 0))^2/2\sigma_{\hat{\theta}_S}^2)$$

$$\times \exp(-(\theta_{S,t}^x - (\theta_{S,t-1}^x + \beta_{t-1}^x)I(M_{t-1}^x = 0))^2/2\sigma_{\hat{\theta}_S}^2)]$$

$$\propto \prod_{x=1}^{S}\exp(-(\theta_{S,t}^x)^2(1/\sigma_{\hat{\theta}}^2 + (I(M_t^x = 0) + 1)/\sigma_{\hat{\theta}_S}^2)/2$$

$$\times \exp(\theta_{S,t}^x[(\theta_t^x - \mu^x - \theta_{ST,t}^x - \theta_{LT,t}^x)/\sigma_{\hat{\theta}}^2 + (I(M_t^x = 0)(\theta_{S,t+1}^x - \beta_t^x)$$

$$+ I(M_{t-1}^x = 0)(\theta_{S,t-1}^x + \beta_{t-1}^x))/\sigma_{\hat{\theta}_S}^2])$$

$$\sim \prod_{x=1}^{S} N(\bar{\theta}_{S,t}^x, \sigma_{\theta_{S,t}^x}^2) , \qquad \text{where}$$

$$\sigma_{\theta_{S,t}^x}^2 = (1/\sigma_{\hat{\theta}}^2 + (I(M_t^x = 0) + 1)/\sigma_{\hat{\theta}_S}^2)^{-1}$$

$$\bar{\theta}_{S,t}^x = \sigma_{\theta_{S,t}^x}^2[(\theta_t^x - \mu^x - \theta_{ST,t}^x - \theta_{LT,t}^x)/\sigma_{\hat{\theta}}^2 + (I(M_t^x = 0)(\theta_{S,t+1}^x - \beta_t^x)$$

$$+ I(M_{t-1}^x = 0)(\theta_{S,t-1}^x + \beta_{t-1}^x))/\sigma_{\hat{\theta}_S}^2]$$

Similarly, we get:

$$[\vec{\boldsymbol{\theta}}_{S,T} \mid \cdot] \sim \prod_{x=1}^{S} N(\bar{\theta}_{S,T}^x, \sigma_{\theta_{S,T}^x}^2) , \qquad \text{where}$$

$$\sigma_{\theta_{S,T}^x}^2 = (1/\sigma_{\hat{\theta}}^2 + 1/\sigma_{\hat{\theta}_S}^2)^{-1}$$

$$\bar{\theta}_{S,T}^x = \sigma_{\theta_{S,T}^x}^2[(\theta_T^x - \mu^x - \theta_{ST,T}^x - \theta_{LT,T}^x)/\sigma_{\hat{\theta}}^2 + I(M_{T-1}^x = 0)(\theta_{S,T-1}^x + \beta_{T-1}^x)/\sigma_{\hat{\theta}_S}^2]$$

$$[\vec{\boldsymbol{\theta}}_{S,0} \mid \cdot] \sim \prod_{x=1}^{S} N(\bar{\theta}_{S,0}^x, \sigma_{\theta_{S,0}^x}^2) , \qquad \text{where}$$

$$\sigma_{\theta_{S,0}^x}^2 = (1/\sigma_{\hat{\theta}_S}^2 + 1/\widetilde{\sigma}_{\theta_{S,0}^x}^2)^{-1}$$

$$\bar{\theta}_{S,0}^x = \sigma_{\theta_{S,0}^x}^2[(\theta_{S,1}^x - \beta_0^x)/\sigma_{\hat{\theta}_S}^2 + \widetilde{\theta}_{S,0}^x/\widetilde{\sigma}_{\theta_{S,0}^x}^2]$$

$[\vec{\boldsymbol{\theta}}_{S,t} \mid \cdot]$, $t = 0, \dots, T$ are straightforward to sample from. Again note that $\theta_{S,t}^x$, $x = 1, \dots, S$ are conditionally independent given all random quantities except for $\vec{\boldsymbol{\theta}}_{S,t}$, and D_T.

We finally consider $[\sigma_{\widehat{M}}^2 \mid \cdot]$. From (10) we have:

$$
\begin{aligned}
[\sigma_{\widehat{M}}^2 \mid \cdot] &\propto \prod_{t=1}^{T} [\vec{\boldsymbol{M}}_t \mid \vec{\boldsymbol{\theta}}_t, \sigma_{\widehat{M}}^2][\sigma_{\widehat{M}}^2 \mid \widetilde{q}_{\sigma_{\widehat{M}}^2}, \widetilde{r}_{\sigma_{\widehat{M}}^2}] \\
&\propto \prod_{t=1}^{T} \prod_{x=1}^{S} [I(M_t^x = 0)\varphi((k - \theta_t^x)/\sigma_{\widehat{M}}) + I(M_t^x \geq k)\exp(-(M_t^x - \theta_t^x)^2/2\sigma_{\widehat{M}}^2)/(\sqrt{2\pi}\,\sigma_{\widehat{M}})] \\
&\quad \times (1/\sigma_{\widehat{M}}^2)^{\widetilde{q}_{\sigma_{\widehat{M}}^2}+1} \exp(-1/(\widetilde{r}_{\sigma_{\widehat{M}}^2}\sigma_{\widehat{M}}^2))
\end{aligned}
$$

This is obviously not straightforward to sample from. Again we can employ a Metropolis-Hastings step in the Gibbs sampling with proposal distribution equal to the one above corresponding to $k = -\infty$, which in fact is $IG(ST/2 + \widetilde{q}_{\sigma_{\widehat{M}}^2}, [\sum_{t=1}^{T} \sum_{x=1}^{S}(M_t^x - \theta_t^x)^2/2 + 1/\widetilde{r}_{\sigma_{\widehat{M}}^2}]^{-1})$. However, here the suggestion of Jørund Gåsemyr seems just excellent. For each $M_t^x = 0$, just use the already drawn fictitious Y_t^x. Then draw a $\sigma_{\widehat{M}}^2$ from the inverse gamma distribution above, but with M_t^x replaced by Y_t^x everywhere where $M_t^x = 0$.

4 CONCLUDING REMARKS

We have in this report introduced two different stochastic models for earthquake occurrence, both focusing on the spatio-temporal interactions between earthquakes. Both models employ Bayesian updating of prior distributions for all parameters of the models, while empirical data are used for deriving posterior distributions. The first model is a marked point process model in which individual earthquakes are represented by their magnitudes and locations in space and time. In the second model the earthquakes are represented as potentials on a grid. At the current stage we have implemented an algorithm for simulating earthquakes only for the first model. This has been fairly straightforward because of the explicit representation of each earthquake in this model. An algorithm for simulations from the second model is planned and will be presented in a future report.

REFERENCES

Casella, G. and George, E.I. (1992), Explaining the Gibbs sampler, The American Statistician, 46, 167–174.

Chib, S. and Greenberg, E. (1995), Understanding the Metropolis-Hastings algorithm, The American Statistician, 49, 327–335.

Cressie, N. (1993), Statistics for Spatial Data, John Wiley & Sons Inc.

Frigessi, A., Gåsemyr, J. and Rue, H. (2000), Antithetic coupling of two Gibbs sampler chains, to appear in Annals Statistics.

Gilks, W.R., Richardson, S. and Spiegelhalter, D.J.(Eds.) (1996), Markov Chain Monte Carlo in Practice, Chapman and Hall, London.

Kagan, Y.Y. (1997), Are earthquakes predictable? Geophys. J. Int., 131, 505–525.

Ogata, Y. (1998), Space-time point-process models for earthquake occurrences, Ann. Inst. Statist. Math., 50, 379–402.

Rue, H. (1999), A fast and exact simulation algorithm for general Gaussian Markov Random Fields, Department of Mathematical Sciences, Norwegian University of Science and Technology, Trondheim, Norway.

Southern California Earthquake Center, SCEC (2000): Catalog of Southern California earthquakes publicly available at http://www.scecdc.scec.org/ftp/catalogs/SCSN/

Stock, C. and Smith, E.G.C. (2000), Evidence for different scaling of earthquake source parameters for large earthquakes depending on faulting mechanism, Geophys. J. Int., 143, 157–162.

Vere-Jones, D. (1995), Forecasting earthquakes and earthquake risk, Int. J. of Forecasting, 11, 503–538.

West, M. and Harrison, J. (1997), Bayesian Forecasting and Dynamic Models, second edition, Springer, New York.

Wikle, C.K., Berliner, L.M. and Cressie, N. A. (1998), Hierarchical Bayesian space-time models, Environmental and Ecological Statistics, 5, 117–154.

Working Group on California Earthquake Probabilities, Seismic Hazard in Southern California: Probable earthquakes, 1994 to 2024. Bulletin of the Seismological Society of America, 85, 379–439, April 1995.

APPENDIX: THE FULL CONDITIONAL DISTRIBUTIONS IN THE HIERARCHICAL BAYESIAN SPACE-TIME MODEL

In this Appendix we give the full conditional distributions needed for Gibbs sampling in the hierarchical Bayesian space-time model. Remember that $[\vec{\boldsymbol{\theta}}_t \mid \cdot]$, $t = 1, \ldots, T$, $[\vec{\boldsymbol{\theta}}_{S,t} \mid \cdot]$, $t = 0, \ldots, T$ and $[\sigma^2_{\widehat{M}} \mid \cdot]$ are given in Section 3.3. Completely parallel to the deduction of $[\vec{\boldsymbol{\theta}}_{S,t} \mid \cdot]$, $t = 0, \ldots, T$, we get

$$[\vec{\boldsymbol{\theta}}_{LT,t} \mid \cdot] \sim \prod_{x=1}^{S} N(\bar{\theta}^x_{LT,t}, \sigma^2_{\theta^x_{LT,t}}), \qquad t = 0, \ldots, T \,,$$

where

$$\sigma^2_{\theta^x_{LT,t}} = (1/\sigma^2_{\hat{\theta}} + (((\cos(2\pi/p^x))^2 + 1)/\sigma^2_{\hat{\theta}_{LT}} + (\sin(2\pi/p^x))^2/\sigma^2_{\hat{\alpha}})^{-1}, \qquad t = 1, \ldots, T-1$$

$$\bar{\theta}^x_{LT,t} = \sigma^2_{\theta^x_{LT,t}}[(\theta^x_t - \mu^x - \theta^x_{ST,t} - \theta^x_{S,t})/\sigma^2_{\hat{\theta}} + (\cos(2\pi/p^x)(\theta^x_{LT,t+1} - \sin(2\pi/p^x)\alpha^x_t)$$
$$+ \cos(2\pi/p^x)\theta^x_{LT,t-1} + \sin(2\pi/p^x)\alpha^x_{t-1})/\sigma^2_{\hat{\theta}_{LT}}$$
$$- \sin(2\pi/p^x)(\alpha^x_{t+1} - \cos(2\pi/p^x)\alpha^x_t)/\sigma^2_{\hat{\alpha}}], \qquad t = 1, \ldots, T-1$$

$$\sigma^2_{\theta^x_{LT,T}} = (1/\sigma^2_{\hat{\theta}} + 1/\sigma^2_{\hat{\theta}_{LT}})^{-1}$$

$$\bar{\theta}^x_{LT,T} = \sigma^2_{\theta^x_{LT,T}}[(\theta^x_T - \mu^x - \theta^x_{ST,T} - \theta^x_{S,T})/\sigma^2_{\hat{\theta}}$$
$$+ (\cos(2\pi/p^x)\theta^x_{LT,T-1} + \sin(2\pi/p^x)\alpha^x_{T-1})/\sigma^2_{\hat{\theta}_{LT}}]$$

$$\sigma^2_{\theta^x_{LT,0}} = (((\cos(2\pi/p^x))^2/\sigma^2_{\hat{\theta}_{LT}} + (\sin(2\pi/p^x))^2/\sigma^2_{\hat{\alpha}} + 1/\tilde{\sigma}^2_{\theta^x_{LT,0}})^{-1}$$

$$\bar{\theta}^x_{LT,0} = \sigma^2_{\theta^x_{LT,0}}[\cos(2\pi/p^x)(\theta^x_{LT,1} - \sin(2\pi/p^x)\alpha^x_0)/\sigma^2_{\hat{\theta}_{LT}}$$
$$- \sin(2\pi/p^x)(\alpha^x_1 - \cos(2\pi/p^x)\alpha^x_0)/\sigma^2_{\hat{\alpha}} + \tilde{\theta}^x_{LT,0}/\tilde{\sigma}^2_{\theta^x_{LT,0}}]$$

We now turn to $[\vec{\boldsymbol{\theta}}_{ST,t} \mid \cdot]$, and following Wikle, Berliner and Cressie (1998), write (12) in the form

$$\vec{\boldsymbol{\theta}}_{ST,t} = \boldsymbol{H}\,\vec{\boldsymbol{\theta}}_{ST,t-1} + \vec{\tilde{\boldsymbol{\theta}}}_{ST,t}$$

Here $\boldsymbol{H}$ is an $S \times S$ diagonal matrix with six off diagonals given by the nearest neighbor parameters $a_N, a_E, a_S, a_W, a_A, a_B$ and a main diagonal containing $\vec{a}$. For edge sites, some of the off-diagonals have corresponding zeros. Completely parallel to the deduction of (A.8), (A.10) and (A.12) in Wikle, Berliner and Cressie (1998), we get

$$[\vec{\boldsymbol{\theta}}_{ST,t} \mid \cdot] \sim N(\vec{\bar{\boldsymbol{\theta}}}_{ST,t}, \boldsymbol{\Sigma}_{\vec{\boldsymbol{\theta}}_{ST,t}}), \qquad t = 0, \ldots, T \,,$$

where

$$\boldsymbol{\Sigma}_{\vec{\boldsymbol{\theta}}_{ST,t}} = (\boldsymbol{I}/\sigma_{\hat{\theta}}^2 + (\boldsymbol{H}'\boldsymbol{H} + \boldsymbol{I})/\sigma_{\hat{\theta}_{ST}}^2)^{-1}, \qquad t = 1,\dots,T-1$$

$$\vec{\boldsymbol{\theta}}_{ST,t} = \boldsymbol{\Sigma}_{\vec{\boldsymbol{\theta}}_{ST,t}}[(\vec{\boldsymbol{\theta}}_t - \vec{\boldsymbol{\mu}} - \vec{\boldsymbol{\theta}}_{S,t} - \vec{\boldsymbol{\theta}}_{LT,t})/\sigma_{\hat{\theta}}^2 + (\boldsymbol{H}'\vec{\boldsymbol{\theta}}_{ST,t+1} + \boldsymbol{H}\vec{\boldsymbol{\theta}}_{ST,t-1})/\sigma_{\hat{\theta}_{ST}}^2]$$

$$\boldsymbol{\Sigma}_{\vec{\boldsymbol{\theta}}_{ST,T}} = (1/\sigma_{\hat{\theta}}^2 + 1/\sigma_{\hat{\theta}_{ST}}^2)^{-1}\boldsymbol{I}$$

$$\vec{\boldsymbol{\theta}}_{ST,T} = \boldsymbol{\Sigma}_{\vec{\boldsymbol{\theta}}_{ST,T}}[(\vec{\boldsymbol{\theta}}_T - \vec{\boldsymbol{\mu}} - \vec{\boldsymbol{\theta}}_{S,T} - \vec{\boldsymbol{\theta}}_{LT,T})/\sigma_{\hat{\theta}}^2 + (\boldsymbol{H}\vec{\boldsymbol{\theta}}_{ST,T-1})/\sigma_{\hat{\theta}_{ST}}^2]$$

Let $\widetilde{\boldsymbol{\Sigma}}_{\vec{\boldsymbol{\theta}}_{ST,0}}$ be the diagonal matrix containing $\vec{\boldsymbol{\sigma}}_{\hat{\theta}_{ST,0}}^2$. Then

$$\boldsymbol{\Sigma}_{\vec{\boldsymbol{\theta}}_{ST,0}} = (\boldsymbol{H}'\boldsymbol{H}/\sigma_{\hat{\theta}_{ST}}^2 + \widetilde{\boldsymbol{\Sigma}}_{\vec{\boldsymbol{\theta}}_{ST,0}}^{-1})^{-1}$$

$$\vec{\boldsymbol{\theta}}_{ST,0} = \boldsymbol{\Sigma}_{\vec{\boldsymbol{\theta}}_{ST,0}}(\boldsymbol{H}'\vec{\boldsymbol{\theta}}_{ST,1}/\sigma_{\hat{\theta}_{ST}}^2 + \widetilde{\boldsymbol{\Sigma}}_{\vec{\boldsymbol{\theta}}_{ST,0}}^{-1}\widetilde{\vec{\boldsymbol{\theta}}}_{ST,0})$$

We next turn to $[\vec{\boldsymbol{\alpha}}_t \mid \cdot]$, $[\vec{\boldsymbol{\beta}}_t \mid \cdot]$ and $[\vec{\boldsymbol{\beta}} \mid \cdot]$:

$$[\vec{\boldsymbol{\alpha}}_t \mid \cdot] \sim \prod_{x=1}^{S} N(\bar{\alpha}_t^x, \sigma_{\alpha_t^x}^2), \qquad t = 0,\dots,T ,$$

where for $t = 1,\dots,T-1$

$$\sigma_{\alpha_t^x}^2 = ((\sin(2\pi/p^x))^2/\sigma_{\hat{\theta}_{LT}}^2 + ((\cos(2\pi/p^x))^2 + 1)/\sigma_{\hat{\alpha}}^2)^{-1}$$

$$\bar{\alpha}_t^x = \sigma_{\alpha_t^x}^2[\sin(2\pi/p^x)(\theta_{LT,t+1}^x - \cos(2\pi/p^x)\theta_{LT,t}^x)/\sigma_{\hat{\theta}_{LT}}^2$$

$$+ (\cos(2\pi/p^x)(\alpha_{t+1}^x + \sin(2\pi/p^x)\theta_{LT,t}^x) - \sin(2\pi/p^x)\theta_{LT,t-1}^x + \cos(2\pi/p^x)\alpha_{t-1}^x)/\sigma_{\hat{\alpha}}^2]$$

$$\sigma_{\alpha_T^x}^2 = \sigma_{\hat{\alpha}}^2$$

$$\bar{\alpha}_T^x = -\sin(2\pi/p^x)\theta_{LT,T-1}^x + \cos(2\pi/p^x)\alpha_{T-1}^x$$

$$\sigma_{\alpha_0^x}^2 = ((\sin(2\pi/p^x))^2/\sigma_{\hat{\theta}_{LT}}^2 + (\cos(2\pi/p^x))^2/\sigma_{\hat{\alpha}}^2 + 1/\widetilde{\sigma}_{\alpha_0^x}^2)^{-1}$$

$$\bar{\alpha}_0^x = \sigma_{\alpha_0^x}^2[\sin(2\pi/p^x)(\theta_{LT,1}^x - \cos(2\pi/p^x)\theta_{LT,0}^x)/\sigma_{\hat{\theta}_{LT}}^2$$

$$+ \cos(2\pi/p^x)(\alpha_1^x + \sin(2\pi/p^x)\theta_{LT,0}^x)/\sigma_{\hat{\alpha}}^2 + \widetilde{\alpha}_0^x/\widetilde{\sigma}_{\alpha_0^x}^2],$$

$$[\vec{\boldsymbol{\beta}}_t \mid \cdot] \sim \prod_{x=1}^{S} N(\bar{\beta}_t^x, \sigma_{\beta_t^x}^2), \qquad t = 0,\dots,T ,$$

where for $t = 1, \ldots, T - 1$

$$\sigma^2_{\beta^x_t} = (I(M^x_t = 0)/\sigma^2_{\hat{\theta}_S} + (I(M^x_t = 0) + 1)/\sigma^2_{\hat{\beta}})^{-1}$$

$$\bar{\beta}^x_t = \sigma^2_{\beta^x_t}[I(M^x_t = 0)(\theta^x_{S,t+1} - \theta^x_{S,t})/\sigma^2_{\hat{\theta}_S} + (I(M^x_t = 0)\beta^x_{t+1} + \beta^x + (\beta^x_{t-1} - \beta^x)I(M^x_{t-1} = 0))/\sigma^2_{\hat{\beta}}]$$

$$\sigma^2_{\beta^x_T} = \sigma^2_{\hat{\beta}}$$

$$\bar{\beta}^x_T = (\beta^x + (\beta^x_{T-1} - \beta^x)I(M^x_{T-1} = 0))$$

$$\sigma^2_{\beta^x_0} = (1/\sigma^2_{\hat{\theta}_S} + 2/\sigma^2_{\hat{\beta}})^{-1}$$

$$\bar{\beta}^x_0 = \sigma^2_{\beta^x_0}[(\theta^x_{S,1} - \theta^x_{S,0})/\sigma^2_{\hat{\theta}_S} + (\beta^x_1 + \beta^x)/\sigma^2_{\hat{\beta}}],$$

$$[\vec{\boldsymbol{\beta}} \mid \cdot] \sim \prod_{x=1}^{S} N(\bar{\beta}^x, \sigma^2_{\beta^x}), \qquad \text{where}$$

$$\sigma^2_{\beta^x} = (\sum_{t=0}^{T} I(M^x_{t-1} \geq k)/\sigma^2_{\hat{\beta}} + 1/\tilde{\sigma}^2_{\beta^x})^{-1}$$

$$\bar{\beta}^x = \sigma^2_{\beta^x}[\sum_{t=0}^{T} I(M^x_{t-1} \geq k)\beta^x_t/\sigma^2_{\hat{\beta}} + \tilde{\beta}^x/\tilde{\sigma}^2_{\beta^x}]$$

Next consider $[\vec{\boldsymbol{\mu}} \mid \cdot]$ and $[\vec{\boldsymbol{\mu}}_0 \mid \cdot]$. Parallel to (16) of Wikle, Berliner and Cressie (1998) we have from (15)

$$[\vec{\boldsymbol{\mu}} \mid \vec{\boldsymbol{\mu}}_0, b_{N,S}, b_{E,W}, b_{A,B}, \sigma^2_\mu] \sim N(\vec{\boldsymbol{\mu}}_0, (\boldsymbol{I} - \boldsymbol{C}_b)^{-1}\sigma^2_\mu) ,$$

where $\boldsymbol{C}_b$ is a symmetric $S \times S$ matrix with six off-diagonals given by $b_{N,S}, b_{E,W}, b_{A,B}$ (see Cressie 1993, p. 434). Completely parallel to the deduction of (A.14) in Wikle, Berliner and Cressie (1998) we get

$$[\vec{\boldsymbol{\mu}} \mid \cdot] \sim N(\vec{\boldsymbol{\mu}}, \boldsymbol{\Sigma}_{\vec{\boldsymbol{\mu}}}) ,$$

where

$$\boldsymbol{\Sigma}_{\vec{\boldsymbol{\mu}}} = ((\boldsymbol{I} - \boldsymbol{C}_b)/\sigma^2_\mu + \boldsymbol{I}T/\sigma^2_{\hat{\theta}})^{-1}$$

$$\vec{\boldsymbol{\mu}} = \boldsymbol{\Sigma}_{\vec{\boldsymbol{\mu}}}[(\boldsymbol{I} - \boldsymbol{C}_b)\vec{\boldsymbol{\mu}}_0/\sigma^2_\mu + \sum_{t=1}^{T}(\vec{\boldsymbol{\theta}}_t - \vec{\boldsymbol{\theta}}_{ST,t} - \vec{\boldsymbol{\theta}}_{S,t} - \vec{\boldsymbol{\theta}}_{LT,t})/\sigma^2_{\hat{\theta}}]$$

As an alternative, the fast and exact simulation algorithm for general Gaussian Markov Random Fields given in Rue (1999) could be applied to arrive at $[\vec{\boldsymbol{\mu}} \mid \cdot]$.

Let $\widetilde{\Sigma}_{\vec{\mu}_0}$ be the diagonal matrix containing $\vec{\sigma}_{\mu_0}^2$. Then

$$[\vec{\mu}_0 \mid \cdot] \sim N(\vec{\bar{\mu}}_0, \Sigma_{\vec{\mu}_0}) \ ,$$

where

$$\Sigma_{\vec{\mu}_0} = ((\boldsymbol{I} - \boldsymbol{C}_b)/\sigma_\mu^2 + \widetilde{\Sigma}_{\vec{\mu}_0}^{-1})^{-1}$$

$$\vec{\bar{\mu}}_0 = \Sigma_{\vec{\mu}_0}[(\boldsymbol{I} - \boldsymbol{C}_b)\vec{\mu}/\sigma_\mu^2 + \widetilde{\Sigma}_{\vec{\mu}_0}^{-1}\vec{\tilde{\mu}}_0]$$

We now consider $[\vec{a} \mid \cdot]$ and $[\vec{a}_0 \mid \cdot]$. As in Wikle, Berliner and Cressie (1998) let $\boldsymbol{H}_a$ be the $\boldsymbol{H}$ matrix with zeros on the main diagonal instead of $\vec{a}$ and let $\boldsymbol{\theta}_{ST,t-1}$ be the diagonal matrix containing $\vec{\boldsymbol{\theta}}_{ST,t-1}$. Also define $\boldsymbol{C}_a$ completely parallel to $\boldsymbol{C}_b$ with $b_{N,S}, b_{E,W}, b_{A,B}$ replaced by $a_{N,S}, a_{E,W}$ and $a_{A,B}$. Then completely parallel to (A.28) of the latter paper, correcting for a misprint, we get:

$$[\vec{a} \mid \cdot] \sim N(\vec{\bar{a}}, \Sigma_{\vec{a}}) \ ,$$

where

$$\Sigma_{\vec{a}} = ((\boldsymbol{I} - \boldsymbol{C}_a)/\sigma_a^2 + \sum_{t=1}^{T} \boldsymbol{\theta}_{ST,t-1}\boldsymbol{\theta}_{ST,t-1}/\sigma_{\theta_{ST}}^2)^{-1}$$

$$\vec{\bar{a}} = \Sigma_{\vec{a}}[(\boldsymbol{I} - \boldsymbol{C}_a)\vec{a}_0/\sigma_a^2 + \sum_{t=1}^{T} (\vec{\boldsymbol{\theta}}_{ST,t} - \boldsymbol{H}_a\vec{\boldsymbol{\theta}}_{ST,t-1})'\boldsymbol{\theta}_{ST,t-1}/\sigma_{\theta_{ST}}^2]$$

Again, the algorithm of Rue (1999) could be applied to arrive at $[\vec{a} \mid \cdot]$. $[\vec{a}_0 \mid \cdot]$ is identical to $[\boldsymbol{\mu}_0 \mid \cdot]$ by everywhere replacing μ and b by a.

Turning to $[a_N \mid \cdot], [a_E \mid \cdot], [a_S \mid \cdot], [a_W \mid \cdot], [a_A \mid \cdot]$ and $[a_B \mid \cdot]$, we consider the first one since the results for the rest are completely parallel. As in Wikle, Berliner and Cressie (1998) let $\boldsymbol{H}_{a_N}$ be the $\boldsymbol{H}$ matrix with zeros on the off-diagonal of a_N's, let $\boldsymbol{J}_{a_N}$ be an $S \times S$ matrix with ones on the off-diagonal of a_N's and zeros elsewhere and let $\vec{\boldsymbol{\theta}}_{ST,t-1}^{a_N} = \boldsymbol{J}_{a_N}\vec{\boldsymbol{\theta}}_{ST,t-1}$. Then completely parallel to the deduction of (A.31) of the latter paper, again correcting for a misprint, we get

$$[a_N \mid \cdot] \sim N(\bar{a}_N, \sigma_{a_N}^2) \ , \qquad \text{where}$$

$$\sigma_{a_N}^2 = \left(\sum_{t=1}^{T} (\vec{\boldsymbol{\theta}}_{ST,t-1}^{a_N})' \vec{\boldsymbol{\theta}}_{ST,t-1}^{a_N} / \sigma_{\hat{\theta}_{ST}}^2 + 1/\widetilde{\sigma}_{a_N}^2 \right)^{-1}$$

$$\bar{a}_N = \sigma_{a_N}^2 \left[\sum_{t=1}^{T} (\vec{\boldsymbol{\theta}}_{ST,t} - \boldsymbol{H}_{a_N} \vec{\boldsymbol{\theta}}_{ST,t-1})' \vec{\boldsymbol{\theta}}_{ST,t-1}^{a_N} / \sigma_{\hat{\theta}_{ST}}^2 + \widetilde{a}_N / \widetilde{\sigma}_{a_N}^2 \right]$$

The Markov Random Field spatial dependence parameters $a_{N,S}$, $a_{E,W}$, $a_{A,B}$, $b_{N,S}$, $b_{E,W}$, $b_{A,B}$ have already been commented on in Section 3.3. We consider the first one since again the results for the rest are completely similar. Parallel to (A.32) of Wikle, Berliner and Cressie (1998), we have

$$[a_{N,S} \mid \cdot] \propto |(\boldsymbol{I} - \boldsymbol{C}_a)^{-1}|^{-1/2} \exp(-(\vec{a} - \vec{a}_0)'(\boldsymbol{I} - \boldsymbol{C}_a)(\vec{a} - \vec{a}_0)/2\sigma_a^2)$$

$$\times \exp(-(a_{N,S} - \widetilde{a}_{N,S})^2 / 2\widetilde{\sigma}_{a_{N,S}}^2) \,,$$

where $a_{N,S}$ is on two of the six off-diagonals of $\boldsymbol{C}_a$. The presence of $a_{N,S}$ in the determinant term makes this a difficult distribution to sample from. Instead as in the latter paper we employ a Metropolis-Hastings step in the Gibbs sampling with the following proposal distribution:

$$[a_{N,S}^{\text{Pseudo}} \mid \cdot] \sim N(\bar{a}_{N,S}^{\text{Pseudo}}, \sigma_{a_{N,S}^{\text{Pseudo}}}^2) \,, \qquad \text{where}$$

$$\sigma_{a_{N,S}^{\text{Pseudo}}}^2 = \left(\sum_{x=1}^{S} (a^{N(x)} - a_0^{N(x)} + a^{S(x)} - a_0^{S(x)})^2 / \sigma_a^2 + 1/\widetilde{\sigma}_{a_{N,S}}^2 \right)^{-1}$$

$$\bar{a}_{N,S}^{\text{Pseudo}} = \sigma_{a_{N,S}^{\text{Pseudo}}}^2 \left[\sum_{x=1}^{S} (a^{N(x)} - a_0^{N(x)} + a^{S(x)} - a_0^{S(x)}) \right.$$

$$\times (a^x - a_0^x - a_{E,W}(a^{E(x)} - a_0^{E(x)} + a^{W(x)} - a_0^{W(x)})$$

$$\left. -a_{A,B}(a^{A(x)} - a_0^{A(x)} + a^{B(x)} - a_0^{B(x)})) + \widetilde{a}_{N,S}/\widetilde{\sigma}_{a_{N,S}}^2 \right]$$

Note that when considering $b_{N,S}, b_{E,W}$ and $b_{A,B}$ these random quantities are replacing $a_{N,S}$, $a_{E,W}$ and $a_{A,B}$, whereas otherwise a is replaced by μ.

What is left are the remaining variances and we get completely parallel to the deductions in Wikle,

Berliner and Cressie (1998):

$$[\sigma_{\widehat{\theta}}^2 \mid \cdot] \sim IG\left(ST/2 + \widetilde{q}_{\sigma_{\widehat{\theta}}^2}, \left[\sum_{t=1}^{T}\sum_{x=1}^{S}(\theta_t^x - \mu^x - \theta_{ST,t}^x - \theta_{S,t}^x - \theta_{LT,t}^x)^2/2 + 1/\widetilde{r}_{\sigma_{\widehat{\theta}}^2}\right]^{-1}\right)$$

$$[\sigma_{\widehat{\theta}_{ST}}^2 \mid \cdot] \sim IG\left(ST/2 + \widetilde{q}_{\sigma_{\widehat{\theta}_{ST}}^2}, \left[\sum_{t=1}^{T}\sum_{x=1}^{S}(\theta_{ST,t}^x - a^x\theta_{ST,t-1}^x - a_N\theta_{ST,t-1}^{N(x)} - a_E\theta_{ST,t-1}^{E(x)}\right.\right.$$

$$\left.\left. -a_S\theta_{ST,t-1}^{S(x)} - a_W\theta_{ST,t-1}^{W(x)} - a_A\theta_{ST,t-1}^{A(x)} - a_B\theta_{ST,t-1}^{B(x)})^2/2 + 1/\widetilde{r}_{\sigma_{\widehat{\theta}_{ST}}^2}\right]^{-1}\right)$$

$$[\sigma_{\widehat{\theta}_S}^2 \mid \cdot] \sim IG\left(ST/2 + \widetilde{q}_{\sigma_{\widehat{\theta}_S}^2}, \left[\sum_{t=1}^{T}\sum_{x=1}^{S}(\theta_{S,t}^x - (\theta_{S,t-1}^x + \beta_{t-1}^x)I(M_{t-1}^x = 0))^2/2 + 1/\widetilde{r}_{\sigma_{\widehat{\theta}_S}^2}\right]^{-1}\right)$$

$$[\sigma_{\widehat{\beta}}^2 \mid \cdot] \sim IG\left(ST/2 + \widetilde{q}_{\sigma_{\widehat{\beta}}^2}, \left[\sum_{t=1}^{T}\sum_{x=1}^{S}(\beta_t^x - \beta^x - (\beta_{t-1}^x - \beta^x)I(M_{t-1}^x = 0))^2/2 + 1/\widetilde{r}_{\sigma_{\widehat{\beta}}^2}\right]^{-1}\right)$$

$$[\sigma_{\widehat{\theta}_{LT}}^2 \mid \cdot] \sim IG\left(ST/2 + \widetilde{q}_{\sigma_{\widehat{\theta}_{LT}}^2}, \left[\sum_{t=1}^{T}\sum_{x=1}^{S}(\theta_{LT,t}^x - \cos(2\pi/p^x)\theta_{LT,t-1}^x\right.\right.$$

$$\left.\left. - \sin(2\pi/p^x)\alpha_{t-1}^x)^2/2 + 1/\widetilde{r}_{\sigma_{\widehat{\theta}_{LT}}^2}\right]^{-1}\right)$$

$$[\sigma_{\widehat{\alpha}}^2 \mid \cdot] \sim IG\left(ST/2 + \widetilde{q}_{\sigma_{\widehat{\alpha}}^2}, \left[\sum_{t=1}^{T}\sum_{x=1}^{S}(\alpha_t^x + \sin(2\pi/p^x)\theta_{LT,t-1}^x - \cos(2\pi/p^x)\alpha_{t-1}^x)^2/2 + 1/\widetilde{r}_{\sigma_{\widehat{\alpha}}^2}\right]^{-1}\right)$$

$$[\sigma_\mu^2 \mid \cdot] \sim IG(S/2 + \widetilde{q}_{\sigma_\mu^2}, [(\vec{\boldsymbol{\mu}} - \vec{\boldsymbol{\mu}}_0)'(\boldsymbol{I} - \boldsymbol{C}_b)(\vec{\boldsymbol{\mu}} - \vec{\boldsymbol{\mu}}_0)/2 + 1/\widetilde{r}_{\sigma_\mu^2}]^{-1})$$

$$[\sigma_a^2 \mid \cdot] \sim IG(S/2 + \widetilde{q}_{\sigma_a^2}, [(\vec{\boldsymbol{a}} - \vec{\boldsymbol{a}}_0)'(\boldsymbol{I} - \boldsymbol{C}_a)(\vec{\boldsymbol{a}} - \vec{\boldsymbol{a}}_0)/2 + 1/\widetilde{r}_{\sigma_a^2}]^{-1})$$