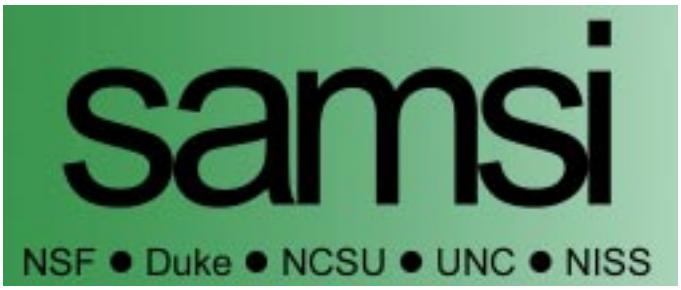

A Comparison of Bayesian Hierarchical Space-Time Models for Earthquake Data

Bent Natvig and Ingunn Fride Tvete

Technical Report #2004-18
June 28, 2004

This material was based upon work supported by the National Science Foundation under Agreement No. DMS-0112069. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Statistical and Applied Mathematical Sciences Institute
PO Box 14006
Research Triangle Park, NC 27709-4006
www.samsi.info

A Comparison of Bayesian Hierarchical Space-Time Models for Earthquake Data

Bent NATVIG and Ingunn Fride TVETE

Stochastic earthquake models are often based on a marked point process approach as for instance presented in Vere-Jones (1995). This gives a fine resolution both in space and time making it possible to represent each earthquake with corresponding foreshocks and aftershocks separately. However, it is not obvious that this microscopic approach is advantageous when aiming at earthquake predictions. In the present paper we take a macroscopic point of view considering grid cells of $0.5^\circ \times 0.5^\circ$, or about 50 km $\times$ 50 km, and time periods of 4 months, which seems suitable for predictions. Hereby, also the effects of foreshocks and aftershocks are circumvented. More specifically, we will discuss different alternative Bayesian hierarchical space-time models in the spirit of Wikle et al. (1998). For each time period the observations are the magnitudes of the largest observed earthquake within each grid cell. In our models these largest observed earthquakes are represented by hidden system state variables called potentials. The potentials at each time period and grid point are decomposed into a time independent term and various alternative time dependent terms with spatial description. As data we apply parts of an earthquake catalog provided by The Northern California Earthquake Data Center where we limit ourselves to the area $32^\circ - 37^\circ N$, $115^\circ - 120^\circ W$ for the years 1981 through 1999 containing the Landers and Hector Mine earthquakes of magnitudes respectively 7.3 and 7.1 on the Richter scale. Based on the various space-time models earthquake predictions for the next two time periods for all grid cells are arrived at. The models are implemented within an MCMC framework in Matlab. The model that gives the overall best predictions is claimed to give valuable and new knowledge on the spatial and time related dependencies between earthquakes.

KEY WORDS: Dynamical systems; Geostatistics; Gibbs sampling; MCMC; Model comparison; Seismology.

Bent Natvig is Professor, Department of Mathematics, University of Oslo, $N - 0316$, Norway (E-mail: *bent@math.uio.no*). Ingunn Fride Tvete is Research Fellow, Department of Mathematics, University of Oslo, $N - 0316$, Norway, (E-mail: *ift@math.uio.no*). This material was based upon work supported by the National Science Foundation under Agreement No. $DMS - 0112069$. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. This joint work was started in February 2003 while the authors visited the Statistical and Applied Mathematical Sciences Institute (SAMSI) in North Carolina. We are indebted to its director James O. Berger and to Richard L. Smith, the program leader for Large-Scale Computer Models for Environmental Systems, for an excellent research environment. The work has also benefited from the "Evaluation of Bayesian Hierarchical Models"

project, supported by the Norwegian Research Council, under grant $154911/V30$, at the Department of Mathematics, University of Oslo. A series of persons have provided very useful suggestions during this research. Especially, Jørund Gåsemyr has been extremely helpful. Otherwise we thank in alphabetical order Hilmar Bungum, Fredrik Dahl, Alan Gelfand, Ingrid Glad, Geir Storvik, Christopher Wikle and Robert Wolpert.

1 INTRODUCTION

Vere-Jones (1995) argues that because of the complexity of an earthquake occurrence there is a need to consider probabilistic models. Several papers are based on a marked point process approach to such modeling, e.g. Ogata (1998), the first part of Holden et al. (2000) and recently Schoenberg (2003), where each earthquake is represented by its magnitude and coordinates in space and time. The marked point process approach has the advantage that it represents each earthquake with corresponding foreshocks and aftershocks separately. This gives a fine resolution making it feasible also to include known physical quantities connected to each single earthquake. However, it is not obvious that this microscopic approach is advantageous when aiming at earthquake predictions. In the present paper we take a macroscopic point of view considering grid cells of $0.5^\circ \times 0.5^\circ$, or about 50 km $\times$ 50 km, and time periods of 4 months.
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 by Wikle et al. (1998) this basic theory is used to develop a hierarchical Bayesian space-time model, and it is suggested that this gives more flexible models and methods for the analysis of environmental data distributed in space and time. Their model is implemented in an MCMC simulation framework using Gibbs sampling and applied to an atmospheric data set of monthly maximum temperatures. Another excellent paper in this area is Wikle et al. (2001) considering tropical surface winds. Inspired by their work we will in the following present different alternative Bayesian hierarchical space-time earthquake models, following up the second part of Holden et al. (2000). For each time period the observations are the magnitudes of the largest observed earthquake within each grid cell. In our models these largest observed earthquakes are represented by hidden system state variables called potentials. The potentials at each time period and grid point are decomposed into a time independent term and various alternative time dependent terms with spatial description.
Although Vere-Jones (1995) is very much concerned with a marked point process approach to earthquake prediction he is stating at the end of his 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. Earthquake prediction has been a challenge for a long time. Geller et al.

(1997) ask: "Is prediction inherently impossible or just fiendishly difficult? In practice, it doesn't matter. Scientifically, the question can be addressed using a Bayesian approach".
Based on the various space-time models short time earthquake predictions for the next two time periods for all grid cells are arrived at in the present paper. The models are implemented within an MCMC framework in Matlab. The model that gives the overall best predictions is claimed to give valuable and new knowledge on the spatial and time related dependencies between earthquakes. One additional advantage of our Bayesian hierarchical modeling is its flexibility. A module can easily be removed and replaced by another. This makes it easy to include expert knowledge in different ways and to try out different model varieties.

2 DATA

We consider parts of an earthquake catalog provided by the *The Northern California Earthquake Data Center* for the years 1981 through 1999, covering the region $32^\circ - 37^\circ$N and $115^\circ - 120^\circ$W. The reason for starting in 1981 is that previous to this year there were not many registered earthquakes below 3 on the Richter scale ($3M$) in the catalog. The basic question we ask is how large the maximal earthquake for a given time period and grid cell going to be. We therefore only consider the largest earthquake for each time period in each grid cell. We could of course also have looked at the average earthquake magnitude or the aggregated earthquake magnitude, but then we would really be answering a different kind of question.
There could be several ways to compare models. We have found it most natural and demanding to use prediction results to decide if a model is *good* or not. Schoenberg (2003) instead applies a multidimensional residual analysis approach. Within our chosen area and time span two major earthquakes have taken place. The first is the Landers earthquake. This $7.3M$ earthquake took place on the 28th of June 1992. The second is the Hector Mine earthquake on the 16th of October 1999, being of magnitude $7.1M$. The great challenge for any earthquake model is the ability to predict such large and rare earthquakes within a reasonable time period and grid cell size. Hence, it is of interest to see how well our models predict these two extreme events. First we will run the models for the years 1981 through 1991, and then make predictions for the next two periods, the last of these containing the Landers earthquake. According to Price and Sandwell (1998) this earthquake caused some regional deformations. Hence, we will restart the simulations in 1993 and run through April 1999. We then again make predictions for the next two periods, the last containing the Hector Mine earthquake. This way all the data will only be used once.
Taking the average maximal magnitudes over the periods $1981-1991$ and $1993-$ April 1999 respectively, we obtain the two top left plots in Figure 1. These plots display the earthquake occurrences as a "diagonal belt" in the NW-SE direction. This feature is

present in all years, but there seems to be more seismic activity in the top right corner of our area in the years 1993– April 1999 following the Landers earthquake. This is in accordance with the mentioned belief that this earthquake caused some regional deformations. We also see this in the bottom two left plots, where the standard deviations are plotted. Looking at the top right plot in Figure 1 we see that there were several other large earthquakes in our area in the period that the Landers earthquake took place. From the bottom right plot it is seen that this is different to the period when the Hector Mine earthquake occurred where there was only one additional large earthquake. It must also be mentioned that in the period prior to the Landers earthquake there was more seismic activity compared to the period prior to the Hector Mine earthquake (this is displayed in Figures 4 and 6).

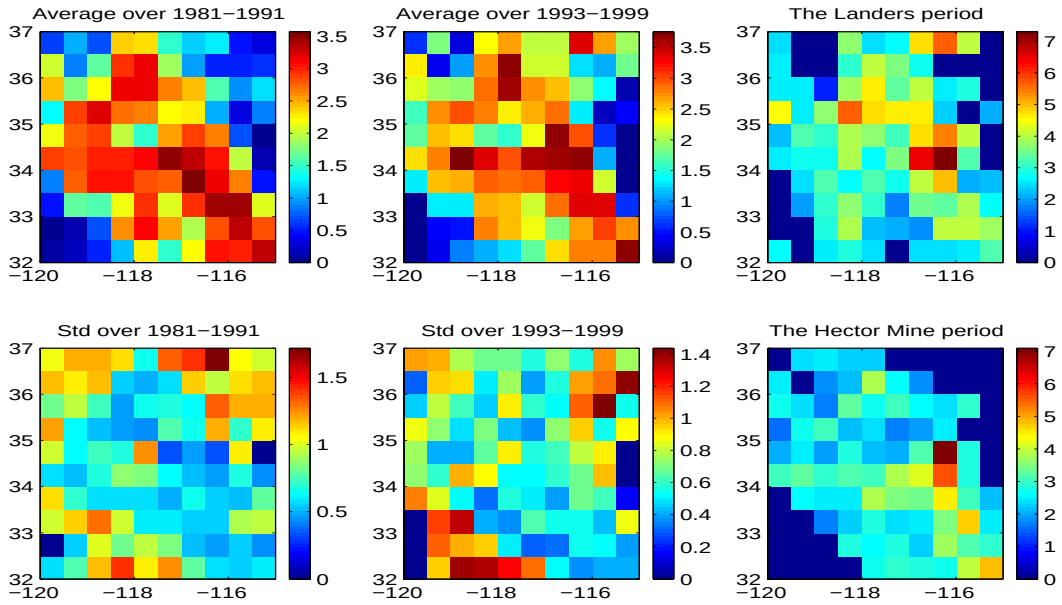

Figure 1: Top left: average maximal magnitudes in 1981 – 1991. Top middle: average maximal magnitudes in 1993– April 1999. Top right: magnitudes during the period containing the Landers earthquake. Bottom left: the standard deviations in 1981 – 1991. Bottom middle: the standard deviations in 1993– April 1999. Bottom right: magnitudes during the period containing the Hector Mine earthquake. Keep in mind the scales.

3 THE MODELS

We will first present one main model, before considering three simpler versions of it. We will then compare the different model alternatives. In the following let an index

x denote space location and t denote time period, $x = 1, ..., X$ and $t = 1, ..., T$. As already mentioned space locations will be situated on a grid with grid cells of 50 km $\times$ 50 km, whereas the distance between points in time is 4 months. If we choose the grid cells too small, we would end up having many grid cells without observations. In addition it would take much longer time to do the simulations. We have also tried grid cells of about 20 km $\times$ 20 km, and our main conclusions remain the same. We have also tried a smaller time resolution, and also here the main features remain the same.
We imagine that in every grid cell and time period there is an underlying potential for a maximal magnitude earthquake. We therefore introduce the following hidden system state variables:

$\theta_t^x =$ maximal magnitude earthquake potential at space location x and time period t.

It is natural to consider this as an unknown, continuously distributed random quantity.
We introduce the observed earthquake magnitudes corresponding to θ_t^x:

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

We also introduce the following corresponding noise (shock) term:

$\widehat{M}_t^x =$ noise (shock) at space location x and time period 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 $N(0, \sigma^2_{\widehat{N}})$, where $\sigma^2_{\widehat{N}}$ is a random quantity. These noise terms are also assumed to be totally independent of all other noise terms.
The link between the earthquake potential and the observed maximal magnitude earthquake at space location x and time period t is given by the observation equation:

$$M_t^x = \theta_t^x + \widehat{M}_t^x, \qquad t = 1, 2, ..., T, \quad x = 1, 2, ..., X. \tag{1}$$

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

$$\theta_t^x = \mu^x + \theta_{S,t}^x + \widehat{\theta}_t^x \ , \qquad t = 1, 2, ..., T, \quad x = 1, 2, ..., X. \tag{2}$$

Here μ^x is a time independent contribution and $\theta_{S,t}^x$ a time dependent contribution with a spatial description. $\widehat{\theta}_t^x$ is a noise term linked to the system state θ_t^x.
We introduce the notation $N =$ north, $E =$ east, $S =$ south and $W =$ west.
Let $N(x) =$ grid cell north of x (if any) and define correspondingly $E(x), S(x)$ and $W(x)$. We model the time independent term as a conditionally specified Gaussian

Markov Random Field (MRF) with first-order spatial dependence, i.e.:

$$\begin{aligned}\mu^x \sim N[&\mu_0^x + c_{NS}\{\mu^{N(x)} - \mu_0^{N(x)} + \mu^{S(x)} - \mu_0^{S(x)}\} \\ &+c_{EW}\{\mu^{E(x)} - \mu_0^{\mathrm{E}(x)} + \mu^{W(x)} - \mu_0^{W(x)}\}, \sigma_\mu^2] \,, \qquad x = 1, 2, ..., X. \end{aligned} \tag{3}$$

μ_0^x is the grid point specific MRF mean, c_{NS} and c_{EW} are respectively north-south and east-west MRF spatial dependence parameters, and σ_μ^2 is the homogeneous MRF variance. μ_0^x is modeled as having a quadratic form, hopefully describing the diagonal feature we see summarized in the two top left plots of Figure 1. Hence, we let m^x and n^x denote longitude and latitude corresponding to the grid point x and assume:

$$\mu_0^x = \mu_0[1] + \mu_0[2]m^x + \mu_0[3]n^x + \mu_0[4](m^x)^2 + \mu_0[5](n^x)^2 + \mu_0[6]m^x n^x, \qquad x = 1, 2, ..., X. \tag{4}$$

A space-time dynamic term can be modeled as a "nearest neighbor" vector autoregressive (VAR) model:

$$\begin{aligned}\theta_{\mathrm{S},t}^x = a^x\theta_{\mathrm{S},t-1}^x + a_N\theta_{\mathrm{S},t-1}^{N(x)} + a_E\theta_{\mathrm{S},t-1}^{E(x)} + a_S\theta_{\mathrm{S},t-1}^{S(x)} + a_W\theta_{\mathrm{S},t-1}^{W(x)} + \widehat{\theta}_{S,t}^x, \\ t = 1, 2, ..., T, \quad x = 1, 2, ..., X,\end{aligned} \tag{5}$$

which is a form of the STARMA models described by Cressie (1993). The autoregressive parameter for the grid cell under consideration, a^x, is allowed to vary spatially, but the nearest neighbor parameters a_N, a_S, a_E and a_W are not. We imagine that the latter parameters have interpretations attached to the underlying earthquake dynamics. In addition we can add a grid point dependent, but time-independent, incremental growth term, β^x, since different grid cells could have different growth potentials. For instance, a grid cell covering a region along a fault would be expected to have a higher growth potential than one in an area far from any fault. Within this space-time dynamic term we also wish to take into account a build-up during a period of earthquakes below $3M$, and smaller earthquakes above $3M$, and let the term drop when a major earthquake appears and tension is released. This adjustment is linked to the linear transformation $a^x\theta_{\mathrm{S},t-1}^x + \beta^x$ of the term at the previous time period at the same grid cell. This leads to:

$$\begin{aligned}\theta_{\mathrm{S},t}^x = &(a^x\theta_{\mathrm{S},t-1}^x + \beta^x)e^{-\kappa(M_{t-1}^x - 3)^2 I(M_{t-1}^x > 3)} \\ &+a_N\theta_{\mathrm{S},t-1}^{N(x)} + a_E\theta_{\mathrm{S},t-1}^{E(x)} + a_S\theta_{\mathrm{S},t-1}^{S(x)} + a_W\theta_{\mathrm{S},t-1}^{W(x)} + \widehat{\theta}_{S,t}^x, \\ &t = 1, 2, ..., T, \quad x = 1, 2, ..., X,\end{aligned} \tag{6}$$

where we set $e^{-\kappa(M_0^x - 3)^2 I(M_0^x > 3)} = 1$. This term can then be interpreted as a combined short time term and strain term, where we intend to model the features of both these contributions. (Originally, we modeled these two features in two separate terms, a short time term and a strain term, but we ran into identifiability problems.) The κ

parameter regulates the reduction in the strain according to the earthquake at the previous time period at the same grid cell. We decided on a threshold of $3M$ in the strain build up part, i.e. there is only a reduction in the strain if the earthquake at the previous period was greater than 3M. The reason for this is that earthquakes below $3M$ are not causing any damage. We have also tried thresholds of $2M$ and $4M$ and the main conclusions remain the same.

a^x is modeled completely parallel to μ^x. Hence,

$$\begin{aligned} a^x \sim \mathrm{N}[a_0^x + a_{NS}\{a^{N(x)} - a_0^{N(x)} + a^{S(x)} - a_0^{S(x)}\} \\ +a_{EW}\{a^{E(x)} - a_0^{E(x)} + a^{W(x)} - a_0^{W(x)}\}, \sigma_a^2], \qquad x = 1, 2, ..., X, \end{aligned} \tag{7}$$

where a_0^x is the grid point specific MRF mean modeled as having a spatial trend:

$$a_0^x = a_0[1] + a_0[2]m^x + a_0[3]n^x, \qquad x = 1, 2, ..., X. \tag{8}$$

β^x is also modeled completely parallel to μ^x:

$$\begin{aligned} \beta^x \sim N[\beta_0^x + b_{NS}\{\beta^{N(x)} - \beta_0^{N(x)} + \beta^{S(x)} - \beta_0^{S(x)}\} \\ +b_{EW}\{\beta^{E(x)} - \beta_0^{E(x)} + \beta^{W(x)} - \beta_0^{W(x)}\}, \sigma_\beta^2], \qquad x = 1, 2, ..., X, \end{aligned} \tag{9}$$

where β_0^x is the grid point specific MRF mean modeled completely parallel to a_0^x.
All prior distributions are assumed to be independent and given as follows. We let $\kappa \sim N(\widetilde{\kappa}, \widetilde{\sigma}_\kappa^2)$, where $\widetilde{\kappa}$ and $\widetilde{\sigma}_\kappa^2$ are fixed parameters that must be specified. Similarly, $\theta_{S,0}^x \sim N(\widetilde{\theta}_{S,0}^x, \widetilde{\sigma}_{\theta_{S,0}^x}^2)$. Without constraints we assume $a_{NS} \sim N(\widetilde{a}_{NS}, \widetilde{\sigma}_{a_{NS}}^2)$ and correspondingly for a_{EW}, b_{NS}, b_{EW}, c_{NS} and c_{EW}. However, to ensure positive definiteness of the marginal covariance matrices, as in Wikle et al. (1998), we must constrain these parameters. Details are given in the appendix. For the trend coefficients we assume that:

$$\begin{aligned} \mu_0[i] &\sim N(\widetilde{\mu}_0[i], \widetilde{\sigma}_{\mu_0}^2[i]), \quad i = 1, ..., 6 \\ a_0[i] &\sim N(\widetilde{a}_0[i], \widetilde{\sigma}_{a_0}^2[i]), \quad i = 1, ..., 3 \\ \beta_0[i] &\sim N(\widetilde{\beta}_0[i], \widetilde{\sigma}_{\beta_0}^2[i]), \quad i = 1, ..., 3, \end{aligned} \tag{10}$$

while $a_N \sim N(\widetilde{a}_N, \widetilde{\sigma}_{a_N}^2)$ and similarly for the other nearest neighbor autoregressive parameters a_E, a_S and a_E. For the noise terms we use inverse gamma conjugate priors, i.e. $\sigma_{\widehat{M}}^2 \sim IG(\widetilde{q}_{\sigma_{\widehat{M}}^2}, \widetilde{r}_{\sigma_{\widehat{M}}^2})$ with corresponding distributions for $\sigma_{\hat{\theta}}^2$, $\sigma_{\hat{\theta}_S}^2$, σ_μ^2, σ_a^2 and σ_β^2.

We now specify the different model alternatives:

Model Alternative 1:

By setting the second term on the right hand side in (2) equal to zero we get a system equation containing only a time independent, conditionally specified Gaussian

MRF, μ^x, in addition to a noise term.

Model Alternative 2:

Here we extend model alternative 1 to also include the short time term (5).

Model Alternative 3:

Here we extend model alternative 2 by just setting β^x equal to 0 in (6). We then get a time dependent term with a spatial description where the magnitude of the earthquake at the previous time period at the same grid cell is taken into account.

Hence, model alternative 1 is nested within model alternative 2, which is again nested within model alternative 3. Finally, the latter is nested within the main model.

4 RESULTS AND DISCUSSIONS

We start out with vague priors on the parameters when considering the first period $1981 - 1991$. For the second period $1993-$ April 1999 we use the estimated posterior values from the first period, since these are then our best prior guesses. We use the MCMC simulation technique Gibbs sampling, which requires the full conditional distributions for all parameters involved. Because we use Gaussian distributions with conjugate priors the derivation of the full conditional distributions are fairly straightforward, except for the a_{NS}, a_{EW}, b_{NS}, b_{EW}, c_{NS}, c_{EW} and κ parameters. For all these parameters we employ a Metropolis-Hastings step in the Gibbs sampler. See the appendix for details. In order to hopefully achieve convergence of the Markov Chain we take 50 000 iterations as the *burn in* iterations. We then run the chain another 150 000 iterations. Since MCMC samples are correlated we only use every 1000th sample from the Markov Chain to arrive at posterior estimates and for predictions. This gives averages of 150 simulated values as our predictions. For most of the parameters the convergence seems fairly satisfactory. Different starting values on the parameters have been tried. Most of the parameters do not seem to be very sensitive to these.
We compared the predicted values with the observed ones by computing the root mean residual sum of squares (RMS):

$$RMS = [\frac{1}{S}\sum_{x=1}^{S}(y_{pred_x} - y_{obs_x})^2]^{\frac{1}{2}}, y_{pred_x} = \frac{1}{N}\sum_{j=1}^{N} y_{pred_{x,j}}, \tag{11}$$

where $N = 150$ and runs through every 1000th simulated prediction.
This measure based on averages does not take into account the uncertainty in the predictions. An alternative that reflects this uncertainty is:

$$RMS^* = [\frac{1}{S}\sum_{x=1}^{S}\frac{1}{N}\sum_{j=1}^{N}(y_{pred_{x,j}} - y_{obs_x})^2]^{\frac{1}{2}}. \quad (12)$$

	1981-1992				*1993-1999*			
model	RMS_1	RMS_2	RMS_1^*	RMS_2^*	RMS_1	RMS_2	RMS_1^*	RMS_2^*
1	1.019	1.477	1.151	1.569	0.838	1.056	0.986	1.177
2	1.017	1.449	1.122	1.522	0.889	1.044	1.028	1.159
3	1.000	1.424	1.110	1.501	0.875	1.033	1.008	1.146
$main\ model$	1.001	1.442	1.119	1.518	0.869	1.051	0.998	1.154

Table 1: Root mean residual sums of squares for the first prediction period, RMS_1 and RMS_1^*, and RMS_2 and RMS_2^* for the second, both for $1981-1992$ and for $1993-1999$.

The prediction results are presented in Table 1. We start by commenting on these based on the years $1981-1992$. We see for all four model alternatives that the RMS_1 value is close to 1 and the RMS_2 value somewhat above 1.4. This difference is due to the extra uncertainty associated with predicting an additional period ahead, but also to the Landers earthquake and some other large earthquakes in nearby grid cells in the second period. We see that the third model alternative has the lowest RMS for both periods, the main model being the second best. From this it seems reasonable to conclude that the inclusion of the β^x term in the main model does not improve the predictions, but rather increases the uncertainty in the modeling. This estimate is for almost all grid cells between 0 and 1, but due to large standard deviations not significantly different from 0. For the main model the estimate of κ is 0.1398, having 0.1731 as standard deviation. This may reflect a small κ value not giving much reduction in the strain when large earthquakes occur. The strain build-up process is not captured too well by this model, and because it contains more parameters than the other alternatives there is only increased uncertainty, and one might just as well use a simpler model.
The overall differences in RMS values between the first model (worst) and the third one (best) are not very large, implying that most of the earthquake dynamics is explained through the spatial structure. We also see that it is sensible to include the data from the previous period in the same grid cell. In the third model κ was estimated to be 0.7653. This gives $e^{-0.7653*(4-3)^2} = 0.4652$ and $e^{-0.7653*(5-3)^2} = 0.0468$. Hence, approximately we will have a factor $a^x \times 0.47$ to be multiplied by $\theta^x_{S,t-1}$ if we had an earthquake of magnitude 4 in the previous period, i.e. a significant amount of tension is then released, and $a^x \times 0.05$ if we had an earthquake of magnitude 5 in the previous period. Considering the RMS^* values, the conclusions on the ranking of the models remain the same.
We next comment on the prediction results based on the years $1993-1999$. For these

predictions the RMS_1 values are somewhat above 0.8 and the RMS_2 values just above 1. A reason for this difference is of course the Hector Mine earthquake in the second period. We see that the RMS_1 value for the first model is the smallest. For the second prediction period the third model alternative has the smallest RMS_2 value. Considering the RMS^* values, again the conclusions on the ranking of the models remain the same. For the main model we now also get a small estimate of κ, 0.2150, having 0.2342 as standard deviation, again indicating that this model alternative is not working the way intended. For the third model alternative we get a κ of 0.9886, giving $e^{-0.9886*(4-3)^2} = 0.3721$ for an earthquake of magnitude 4 in the previous period, and $e^{-0.9886*(5-3)^2} = 0.0192$ for an earthquake of magnitude 5 in the previous period.
The RMS_2 and RMS_2^* values presented in Table 1 are based on predictions made two periods ahead. If one instead made predictions for the periods respectively containing the Landers and Hector Mine earthquake only one period ahead, one would expect better predictions. Doing this for the third model, RMS = 1.364 and RMS^* = 1.450 for the period containing the Landers earthquake and RMS = 1.020 and RMS^* = 1.136 for the period containing the Hector Mine earthquake. There is a better prediction improvement for the period containing the Landers earthquake. The reason for this is that in the period prior to the Landers earthquake there are greater signs of increased earthquake activity than in the period prior to the Hector Mine earthquake. This change is captured by our model.
For the third model alternative for the period $1981 - 1991$, we see by comparing the top left plots in Figures 1 and 2 that the μ^x term basically captures the spatial structure in the data for this period well. A plot of μ_0^x displays similar results. a^x enters the model through the combined short time and strain term. In the top right plot in Figure 2 we see how this is distributed in our area. The estimate is for most grid cells positive, but due to large standard deviations mostly not significantly different from 0. The two bottom plots in Figure 2 show the differences between observed and predicted earthquakes for the two prediction periods in 1992. Also taking Figures 4 and 5 into account we conclude that for the small earthquakes the model seems to do satisfactorily, but for larger earthquakes the predictions are poor. It is clear that the model fails to capture the larger earthquakes, i.e. the model is too conservative.

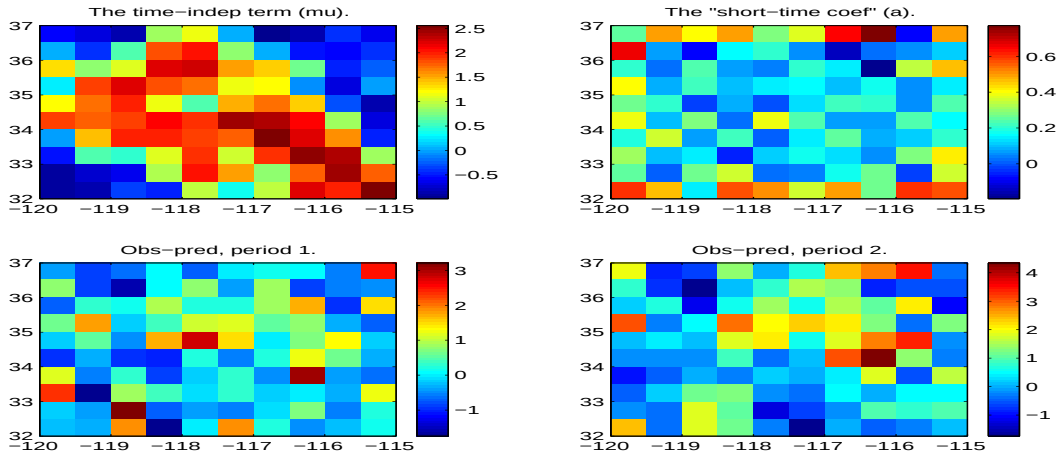

Figure 2: Top left: μ^x. Top right: a^x. Bottom left(right): differences between observed and predicted earthquakes for prediction period 1 (2) for the third model, based on the years $1981 - 1992$. Keep in mind the scales.

When we consider the parameters and predictions based on the years $1993-$ April 1999, we see some of the same features. From the top left plot of Figure 3 we notice how the μ^x term now looks different, having positive values in the top right corner of our area, and by this capturing the change in the spatial structure in the data compared to the modeling based on the years 1981-1991. Again we see from the two bottom plots in Figure 3, also taking Figures 6 and 7 into account, that the model does well for the smaller earthquakes, but fails to capture the larger ones. An interesting exception is the top right corner of the bottom right plot of Figure 3 covering prediction period 2. Here, from Figure 7 there are no observed earthquakes, but predictions are as high as 3.2703. This is due to larger earthquakes in this area previous to the two prediction periods.

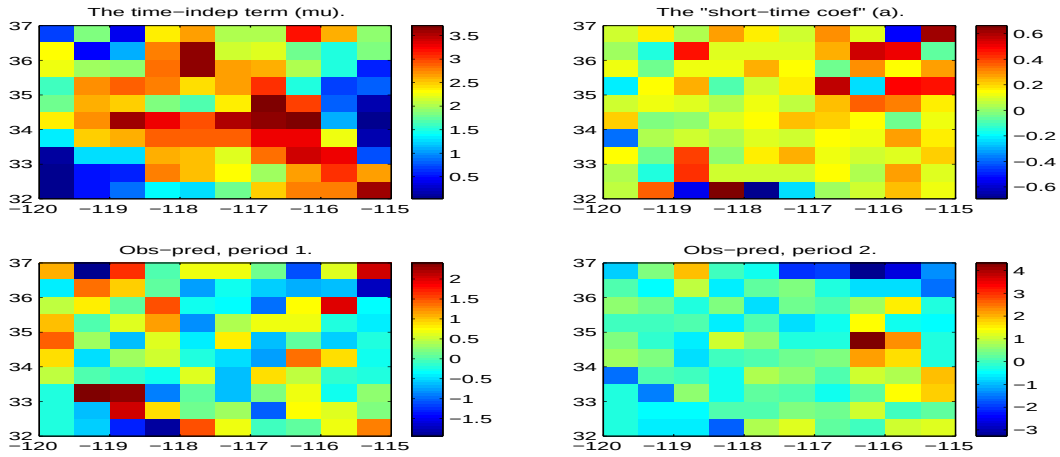

Figure 3: Top left: μ^x. Top right: a^x. Bottom left(right): differences between observed and predicted earthquakes for prediction period 1 (2) for the third model, based on the years $1993 - 1999$. Keep in mind the scales.

As already mentioned the predicted and the observed values are displayed in Figures $4, 5, 6$ and 7. For all the largest earthquakes in each of the prediction periods we see that our model clearly underpredicts the earthquakes. The highest prediction is 3.6739. Considering the cells where the Landers earthquake and the Hector Mine earthquake occurred, we see that the predictions are higher in some of the neighbor cells. The reason for this is that in periods prior to the two prediction periods there were larger earthquakes in the neighbor cells.

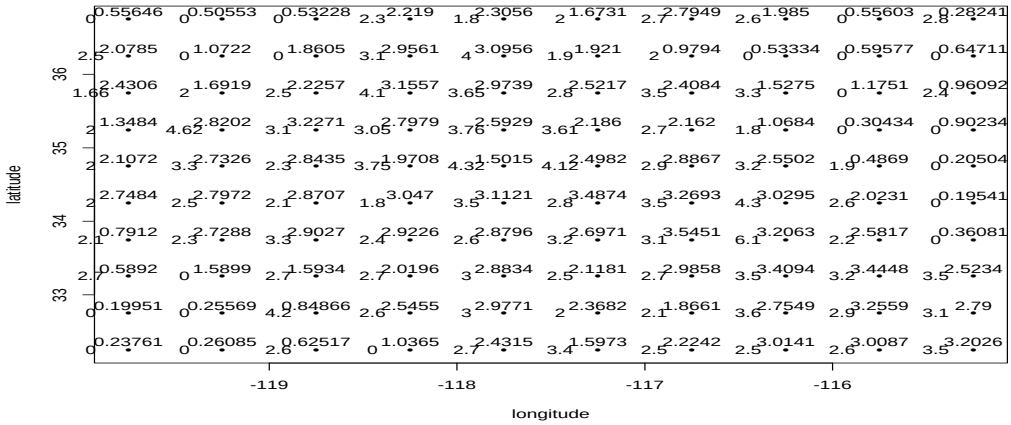

Figure 4: Comparison of predicted (above right) and observed (below left) earthquakes for prediction period 1 for the third model, based on the years 1981-1992.

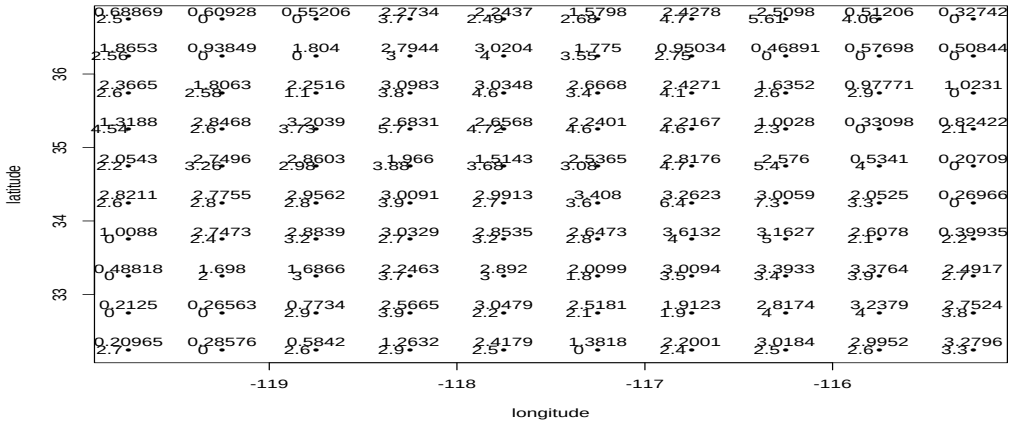

Figure 5: Comparison of predicted (above right) and observed (below left) earthquakes for prediction period 2 for the third model, based on the years 1981-1992.

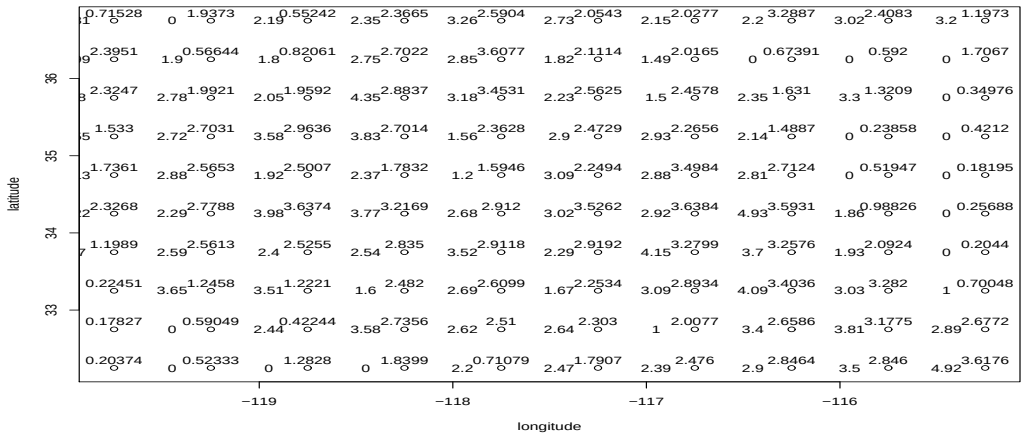

Figure 6: Comparison of predicted (above right) and observed (below left) earthquakes for prediction period 1 for the third model, based on the years 1993-1999.

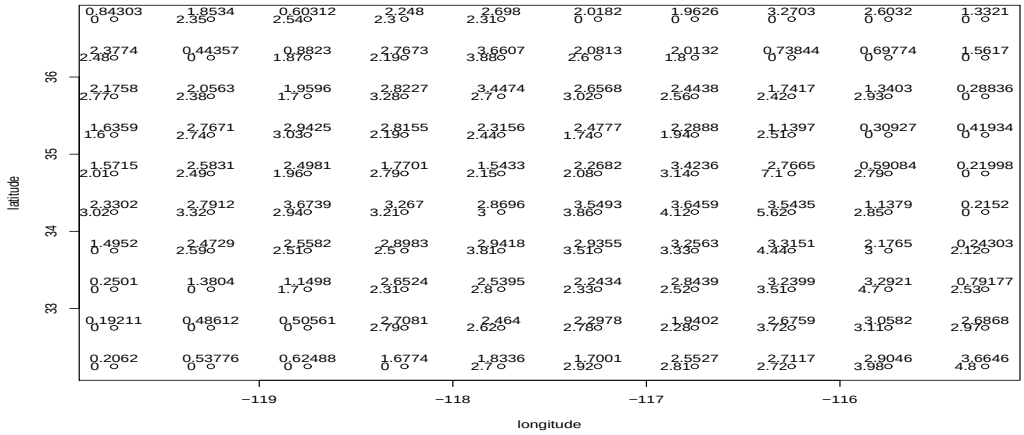

Figure 7: Comparison of predicted (above right) and observed (below left) earthquakes for prediction period 2 for the third model, based on the years 1993-1999.

Table 2 displays the prior and posterior means and standard deviations for the third model alternative for both runs. The first column presents the prior values for the parameters for the first period. The second column displays the posterior means and standard deviations after the first run covering the years 1981 through 1991. These

posterior estimates are then used as prior estimates for the second run for the years 1993 until April 1999. Finally, in the third column are the posterior estimates from the second run displayed.

Having little prior knowledge on the mean μ_0^x of the time independent contribution μ^x, we assume prior to the first run that the first coefficient $\mu_0[1]$ has a mean of 1 and the other coefficients to have a mean of 0, all with the same variance 2. We see that for both runs we obtain about the same posterior mean estimates, $\mu_0[1]$ being close to 1 and the others small, but negative. We also notice that several of the parameters are not significantly different from 0. We must keep in mind that the explanatory variables, longitude and latitude, in (4), enter the equation in their original form, i.e. the longitude has large negative values and the latitude has positive values. Therefore it is natural that these parameter estimates, $\mu_0[2]$ - $\mu_0[6]$, are small. We see much of the same features for the $a_0[1]$ - $a_0[3]$ parameters, except that the $a_0[1]$ parameter has an estimated mean much lower for the second period than the first, resulting in somewhat smaller a^x values for many of the grid cells in the second run. This can be seen in Figures 2 and 3.

For a^x we see that the spatial dependencies parameters, a_{NS} and a_{EW}, seem to be equally important. From the estimates of the spatial dependencies parameters in the time independent μ^x term, c_{NS} and c_{EW}, we see that the c_{NS} parameter is estimated to be larger than c_{EW} . This implies a stronger north - south influence on the spatial variability of the μ^x process, as expected. The posterior estimates of the nearest neighbor autoregressive parameters a_N, a_E, a_S and a_W change a bit from the first run to the second, but are mostly, somewhat surprisingly, not significantly different from 0.

The κ parameter increases from the first to the second run. As already mentioned, in the first run κ was estimated to have a mean of 0.7653, while in the second run it has increased to 0.9886. We also notice that the estimated standard deviations for κ are large. Anyway, we think that the establishing of these estimates, with their corresponding tension release interpretations, is the most interesting contribution of the present paper. Looking at the variances, we see that they all have a lower posterior mean and standard deviation then assumed prior to 1981. We also notice that the posterior estimates of the variances do not change much from the first to the second run, except for σ_a^2. It is interesting to notice that all the variances have a fairly low posterior mean and standard deviation. In addition to the contribution from the κ parameter, this might explain why the model fails to capture the larger earthquakes. There seems not to be enough variation to allow for such *big jumps* in earthquake predictions.

Par.	*pri. mean (sd) for* $1981-1991$	*MCMC post. mean (sd) for* $1981-1991$ *and pri. mean (sd) for* $1993-$ April 1999	*MCMC post. mean (sd) for* $1993-$ April 1999
$\mu_0[1]$	1(1.4142)	1.0159 (1.0228)	0.9650 (0.9878)
$\mu_0[2]$	0(1.4142)	-0.3068 (0.4533)	-0.4459 (0.3240)
$\mu_0[3]$	0(1.4142)	-0.2940 (1.5173)	-0.3711 (1.0618)
$\mu_0[4]$	0(1.4142)	-0.0247 (0.0066)	-0.0137 (0.0057)
$\mu_0[5]$	0(1.4142)	-0.2625 (0.0681)	-0.1141 (0.0598)
$\mu_0[6]$	0(1.4142)	-0.1551 (0.0404)	-0.0706 (0.0355)
$a_0[1]$	1(1.4142)	1.0111 (1.0196)	0.1236 (0.4355)
$a_0[2]$	0(1.4142)	0.0161 (0.0223)	0.0079 (0.0197)
$a_0[3]$	0(1.4142)	0.0278 (0.0713)	0.0265 (0.0677)
a_{NS}	0.2(0.5)	0.1097 (0.0716)	0.0774 (0.0608)
a_{EW}	0.2(0.5)	0.0911 (0.0602)	0.0821 (0.06349
c_{NS}	0.2(0.5)	0.3899 (0.0468)	0.4221 (0.0421)
c_{EW}	0.2(0.5)	0.0754 (0.0463)	0.0579 (0.0386)
a_N	0.2(0.5)	-0.0436 (0.0385)	0.0440 (0.0554)
a_E	0.2(0.5)	0.0269 (0.0381)	0.0507 (0.0564)
a_S	0.2(0.5)	0.0513 (0.0432)	0.0395 (0.0525)
a_W	0.2(0.5)	0.0107 (0.0363)	0.0870 (0.0467)
κ	0.05(0.3162)	0.7653 (0.2415)	0.9886 (0.5123)
σ_a^2	1(0.1428)	0.5895 (0.0642)	0.4641 (0.0387)
σ_μ^2	1(0.1428)	0.6951 (0.0740)	0.6116 (0.0567)
$\sigma^2_{\widehat{\theta}_S}$	0.5(0.0.0721)	0.2754 (0.0240)	0.2609 (0.0070)
$\sigma^2_{\widehat{\theta}}$	0.5(0.5477)	0.1865 (0.0332)	0.1812 (0.0060)
$\sigma^2_{\widehat{M}}$	0.3(0.0490)	0.2548 (0.0320)	0.2437 (0.0056)

Table 2: Prior and posterior estimates for the third model.

5 CONCLUSIONS AND FINAL REMARKS

We have in this study applied Bayesian hierarchical models to study the development of earthquakes in space and time, and to make short time predictions. We have concluded that most of the earthquake dynamics is explained through the spatial structure. In periods with little change in earthquake intensity a model containing only a spatial Gaussian Markov random field in addition to a noise term (the first model alternative) seems to describe the development of earthquakes almost as well as a more complex time dynamic model. But when there is a change in earthquake intensity from one period to another a more complex model, containing a time dependent term with a spatial description where the magnitude of the earthquake at the previous time period at the same grid cell is taken into account (the third model alternative), is preferable. Geo-

physicists and other scientists doing research on earthquakes should find our approach interesting, and by including their expert knowledge this path to modeling should give increased insight in the space-time interaction structure of earthquake occurrences and hopefully a step in the direction of improved predictions.

APPENDIX: DERIVATION OF THE FULL CONDITIONAL DISTRIBUTIONS

With generic notation, let $\overrightarrow{N}$ be an $X \times 1$ vector of the random quantities N^x, $x = 1, 2, ..., X$. We present the full conditional distributions for the main model, the ones for the three alternatives are just special cases.

1. $\underline{[\overrightarrow{\theta}_t|\cdot] \quad t = 1, ..., T}$

   $[\overrightarrow{\theta}_t|\cdot] \propto [\overrightarrow{M}_t|\overrightarrow{\theta}_t, \sigma^2_{\widehat{M}}][\overrightarrow{\theta}_t|\overrightarrow{\mu}, \overrightarrow{\theta}_{S,t}, \sigma^2_{\widehat{\theta}}]$

   $\propto exp\{-\frac{1}{2}[\frac{1}{\sigma^2_{\widehat{M}}}(\overrightarrow{M}_t - \overrightarrow{\theta}_t)'(\overrightarrow{M}_t - \overrightarrow{\theta}_t)]\}$

   $\times exp\{-\frac{1}{2}[\frac{1}{\sigma^2_{\widehat{\theta}}}(\overrightarrow{\theta}_t - (\overrightarrow{\mu} + \overrightarrow{\theta}_{S,t}))'(\overrightarrow{\theta}_t - (\overrightarrow{\mu} + \overrightarrow{\theta}_{S,t}))]\}$

   $\propto exp\{-\frac{1}{2}[(\overrightarrow{\theta}'_t(\frac{1}{\sigma^2_{\widehat{M}}} + \frac{1}{\sigma^2_{\widehat{\theta}}})\overrightarrow{\theta}_t) - 2\overrightarrow{\theta}'_t(\frac{1}{\sigma^2_{\widehat{M}}}\overrightarrow{M}_t + \frac{1}{\sigma_{\widehat{\theta}}}(\overrightarrow{\mu} + \overrightarrow{\theta}_{S,t}))]\}$

   This gives

   $[\overrightarrow{\theta}_t|\cdot] \sim N(\overrightarrow{\theta}_t, \Sigma_{\overrightarrow{\theta}_t}),$

   where

   $\Sigma_{\overrightarrow{\theta}_t} = (\frac{1}{\sigma^2_{\widehat{M}}} + \frac{1}{\sigma^2_{\widehat{\theta}}})^{-1} I$

   and

   $\overrightarrow{\theta}_t = \Sigma_{\overrightarrow{\theta}_t}(\frac{1}{\sigma^2_{\widehat{M}}}\overrightarrow{M}_t + \frac{1}{\sigma^2_{\widehat{\theta}}}(\overrightarrow{\mu} + \overrightarrow{\theta}_{S,t}))$

2. Let diag($\overrightarrow{a}$) be the $X \times X$ matrix with $\overrightarrow{a}$ on the main diagonal, D^κ_t the $X \times X$ matrix with $e^{-\kappa(M^x_{t-1}-3)^2 I(M^x_{t-1}>3)}$, $x = 1, ...X$ on the main diagonal and H_0 the $X \times X$ matrix with zeros on the main diagonal and a_N, a_S, a_E and a_W on the

proper four off-diagonals.

$\underline{[\overrightarrow{\theta}_{S,t}|\cdot] \quad t = 1, ..., T-1}$

$$[\overrightarrow{\theta}_{S,t}|\cdot] \propto$$
$$[\overrightarrow{\theta}_t|\overrightarrow{\mu}, \overrightarrow{\theta}_{S,t}, \sigma^2_{\hat{\theta}}][\overrightarrow{\theta}_{S,t+1}|\overrightarrow{\theta}_{S,t}, \overrightarrow{a}, \overrightarrow{\beta}, a_N, a_S, a_W, a_E, \overrightarrow{M}_t, \kappa, \sigma^2_{\hat{\theta}_S}]$$
$$\times[\overrightarrow{\theta}_{S,t}|\overrightarrow{\theta}_{S,t-1}, \overrightarrow{a}, \overrightarrow{\beta}, a_N, a_S, a_W, a_E, \overrightarrow{M}_{t-1}, \kappa, \sigma^2_{\hat{\theta}_S}]$$
$$\propto exp\{-\tfrac{1}{2}[\tfrac{1}{\sigma^2_{\hat{\theta}}}(\overrightarrow{\theta}_t - (\overrightarrow{\mu} + \overrightarrow{\theta}_{S,t}))'(\overrightarrow{\theta}_t - (\overrightarrow{\mu} + \overrightarrow{\theta}_{S,t}))$$
$$+\tfrac{1}{\sigma^2_{\hat{\theta}_S}}(\overrightarrow{\theta}_{S,t+1} - (diag(\overrightarrow{a})D^\kappa_t + H_0)\overrightarrow{\theta}_{S,t} - D^\kappa_t\overrightarrow{\beta})'$$
$$(\overrightarrow{\theta}_{S,t+1} - (diag(\overrightarrow{a})D^\kappa_t + H_0)\overrightarrow{\theta}_{S,t} - D^\kappa_t\overrightarrow{\beta})$$
$$+\tfrac{1}{\sigma^2_{\hat{\theta}_S}}(\overrightarrow{\theta}_{S,t} - (diag(\overrightarrow{a})D^\kappa_{t-1} + H_0)\overrightarrow{\theta}_{S,t-1} - D^\kappa_{t-1}\overrightarrow{\beta})'$$
$$(\overrightarrow{\theta}_{S,t} - (diag(\overrightarrow{a})D^\kappa_{t-1} + H_0)\overrightarrow{\theta}_{S,t-1} - D^\kappa_{t-1}\overrightarrow{\beta})]\}$$

$$\propto exp\{-\tfrac{1}{2}[\overrightarrow{\theta}'_{S,t}(\tfrac{I}{\sigma^2_{\hat{\theta}}} + ((diag(\overrightarrow{a})D^\kappa_t + H_0)'(diag(\overrightarrow{a})D^\kappa_t + H_0) + I)\tfrac{1}{\sigma^2_{\hat{\theta}_S}})\overrightarrow{\theta}_{S,t}$$
$$-2\overrightarrow{\theta}'_{S,t}((\overrightarrow{\theta}_t - \overrightarrow{\mu})\tfrac{1}{\sigma^2_{\hat{\theta}}} +$$
$$((diag(\overrightarrow{a})D^\kappa_t + H'_0)\overrightarrow{\theta}_{S,t+1} - (diag(\overrightarrow{a})D^\kappa_t + H'_0)D^\kappa_t\overrightarrow{\beta} + (diag(\overrightarrow{a})D^\kappa_{t-1} +$$
$$H_0)\overrightarrow{\theta}_{S,t-1} + D^\kappa_{t-1}\overrightarrow{\beta})\tfrac{1}{\sigma^2_{\hat{\theta}_S}})]\}$$

This gives

$$[\overrightarrow{\theta}_{S,t}|\cdot] \sim N(\overrightarrow{\overline{\theta}}_{S,t}, \Sigma_{\overrightarrow{\theta}_{S,t}}),$$

where

$$\Sigma_{\overrightarrow{\theta}_{S,t}} = \tfrac{I}{\sigma^2_{\hat{\theta}}} + ((diag(\overrightarrow{a})D^\kappa_t + H'_0)(diag(\overrightarrow{a})D^\kappa_t + H_0) + I)\tfrac{1}{\sigma^2_{\hat{\theta}_S}}$$

and

$$\overrightarrow{\overline{\theta}}_{S,t} = \Sigma_{\overrightarrow{\theta}_{S,t}}((\overrightarrow{\theta}_t - \overrightarrow{\mu})\tfrac{1}{\sigma^2_{\hat{\theta}}} + ((diag(\overrightarrow{a})D^\kappa_t + H'_0)\overrightarrow{\theta}_{S,t+1}$$
$$- (diag(\overrightarrow{a})D^\kappa_t + H'_0)D^\kappa_t\overrightarrow{\beta} + (diag(\overrightarrow{a})D^\kappa_{t-1} + H_0)\overrightarrow{\theta}_{S,t-1} + D^\kappa_{t-1}\overrightarrow{\beta})\tfrac{1}{\sigma^2_{\hat{\theta}_S}}).$$

$\underline{[\overrightarrow{\theta}_{S,T}|\cdot]}$

Completely parallel we get

$$[\overrightarrow{\theta}_{S,T}|\cdot] \sim N(\overrightarrow{\overline{\theta}}_{S,T}, \Sigma_{\overrightarrow{\theta}_{S,T}}),$$

where

$$\Sigma_{\overrightarrow{\theta}_{S,T}} = I(\tfrac{1}{\sigma^2_{\widehat{\theta}}} + \tfrac{1}{\sigma^2_{\widehat{\theta}_S}})^{-1}$$

and

$$\overrightarrow{\overline{\theta}}_{S,T} = \Sigma_{\overrightarrow{\theta}_{S,T}}((\overrightarrow{\theta}_T - \overrightarrow{\mu})\tfrac{1}{\sigma^2_{\widehat{\theta}}} + ((diag(\overrightarrow{a})D^{\kappa}_{T-1} + H_0)\overrightarrow{\theta}_{S,T-1} + D^{\kappa}_{T-1}\overrightarrow{\beta})\tfrac{1}{\sigma^2_{\widehat{\theta}_S}}).$$

$\underline{[\overrightarrow{\theta}_{S,0}|\cdot]}$

$$[\overrightarrow{\theta}_{S,0}|\cdot] \propto [\overrightarrow{\theta}_{S,1}|\overrightarrow{\theta}_{S,0}, \overrightarrow{a}, \overrightarrow{\beta}, a_N, a_S, a_W, a_E, \overrightarrow{M}_0, \kappa, \sigma^2_{\widehat{\theta}_S}][\overrightarrow{\theta}_{S,0}|\overrightarrow{\widetilde{\theta}}_{S,0}, \overrightarrow{\widetilde{\sigma}}^2_{S,0}]\}$$

$$\propto exp\{-\tfrac{1}{2}[\tfrac{1}{\sigma^2_{\widehat{\theta}_S}}(\overrightarrow{\theta}_{S,1} - (diag(\overrightarrow{a})D^{\kappa}_0 + H_0)\overrightarrow{\theta}_{S,0} - D^{\kappa}_0\overrightarrow{\beta})'$$
$$(\overrightarrow{\theta}_{S,1} - (diag(\overrightarrow{a})D^{\kappa}_0 + H_0)\overrightarrow{\theta}_{S,0} - D^{\kappa}_0\overrightarrow{\beta})$$
$$+(\overrightarrow{\theta}_{S,0} - \overrightarrow{\widetilde{\theta}}_{S,0})'diag(\overrightarrow{\widetilde{\sigma}}^2_{S,0}{}^{-1})(\overrightarrow{\theta}_{S,0} - \overrightarrow{\widetilde{\theta}}_{S,0})]$$

$$\propto exp\{-\tfrac{1}{2}[\overrightarrow{\theta}'_{S,0}((diag(\overrightarrow{a})D^{\kappa}_0 + H_0)'(diag(\overrightarrow{a})D^{\kappa}_0 + H_0)\tfrac{1}{\sigma^2_{\widehat{\theta}_S}} + diag(\overrightarrow{\widetilde{\sigma}}^2_{S,0}{}^{-1}))\overrightarrow{\theta}_{S,0}$$
$$-2\overrightarrow{\theta}'_{S,0}(((diag(\overrightarrow{a})D^{\kappa}_0 + H'_0)\overrightarrow{\theta}_{S,1} - (diag(\overrightarrow{a})D^{\kappa}_0 + H'_0)D^{\kappa}_0\overrightarrow{\beta})\tfrac{1}{\sigma^2_{\widehat{\theta}_S}} + diag(\overrightarrow{\widetilde{\sigma}}^2_{S,0}{}^{-1})\overrightarrow{\widetilde{\theta}}_{S,0})]\}$$

This gives, with $D^{\kappa}_0 = I$ and $diag(\overrightarrow{a})D^{\kappa}_0 + H_0 = H_a$,

$$[\overrightarrow{\theta}_{S,0}|\cdot] \sim N(\overrightarrow{\overline{\theta}}_{S,0}, \Sigma_{\overrightarrow{\theta}_{S,0}}),$$

where

$$\Sigma_{\overrightarrow{\theta}_{S,0}} = (H'_a H_a \tfrac{1}{\sigma^2_{\widetilde{\sigma}_S}} + diag(\overrightarrow{\widetilde{\sigma}}^2_{S,0}{}^{-1}))^{-1}$$

and

$$\overrightarrow{\overline{\theta}}_{S,0} = \Sigma_{\overrightarrow{\theta}_{S,0}}((H'_a\overrightarrow{\theta}_{S,1} - H'_a\overrightarrow{\beta})\tfrac{1}{\sigma^2_{\widehat{\theta}_S}} + diag(\overrightarrow{\widetilde{\sigma}}^2_{S,0}{}^{-1})\overrightarrow{\widetilde{\theta}}_{S,0}).$$

3. $[\overrightarrow{\mu_0}|\cdot]$, $[\overrightarrow{a_0}|\cdot]$ and $[\overrightarrow{\beta_0}|\cdot]$

Just as in Wikle et al. (1998) we get from (4):

$\overrightarrow{\mu_0} = (\overrightarrow{1} \quad \overrightarrow{m} \quad \overrightarrow{n} \quad \overrightarrow{m^2} \quad \overrightarrow{n^2} \quad \overrightarrow{mn})\,(\mu_0[1] \quad \mu_0[2] \quad \mu_0[3] \quad \mu_0[4] \quad \mu_0[5] \quad \mu_0[6])'$
$= P\mu_{0L}$, where P is the $X \times 6$ trend design matrix,
and $\mu_{0L} = (\mu_0[1] \quad \mu_0[2] \quad \mu_0[3] \quad \mu_0[4] \quad \mu_0[5] \quad \mu_0[6])'$ is the trend parameter vector. We write using (10) $(\mu_{0L}|\tilde{\mu}_{0L}, \tilde{\Sigma}_{\mu_0}) \sim N(\tilde{\mu}_{0L}, \tilde{\Sigma}_{\mu_0})$, where $\tilde{\mu}_{0L}$ is the 6×1 vector of the $\tilde{\mu}_0[i]$'s and $\tilde{\Sigma}_{\mu_0}$ is a 6×6 diagonal matrix with $\tilde{\sigma}^2_{\mu_0[1]}, \tilde{\sigma}^2_{\mu_0[2]}, ..., \tilde{\sigma}^2_{\mu_0[6]}$ on the main diagonal.

Let C_c be a symmetric $X \times X$ matrix where c_{NS} and c_{EW} are each on two of the four off-diagonals (see *Cressie (1993), p. 434*). The full conditional distribution of μ_{0L} is, as in *Wikle et al. (1998)*, given by

$$[\mu_{0L}|\cdot] \sim N(\overrightarrow{\mu}_{0L}, \Sigma_{\overrightarrow{\mu}_{0L}}),$$

where

$$\Sigma_{\overrightarrow{\mu}_{0L}} = (\tfrac{1}{\sigma^2_\mu} P'(I - C_c)P + \widetilde{\Sigma}^{-1}_{\mu_0})^{-1}$$

and

$$\overrightarrow{\mu}_{0L} = \Sigma_{\overrightarrow{\mu_{0L}}}(\tfrac{1}{\sigma^2_\mu} P'(I - C_c)\overrightarrow{\mu} + \widetilde{\Sigma}^{-1}_{\mu_0}\widetilde{\mu}_{0L}).$$

The full conditional distributions for $\overrightarrow{a_0}$ and $\overrightarrow{\beta_0}$ are derived in a similar way.

4. $[\overrightarrow{\mu}|\cdot]$

Parallel to *Wikle et al. (1998)* we get:

$$[\overrightarrow{\mu}|\cdot] \sim N(\overrightarrow{\mu}, \Sigma_{\overrightarrow{\mu}}),$$

where

$$\Sigma_{\overrightarrow{\mu}} = ((I - C_c)\tfrac{1}{\sigma^2_\mu} + \tfrac{IT}{\sigma^2_{\hat{\theta}}})^{-1}$$

and

$\overrightarrow{\mu} = \Sigma_{\overrightarrow{\mu}}((I - C_c)\overrightarrow{\mu_0}\frac{1}{\sigma^2_{\mu}} + \sum_{t=1}^{T}(\overrightarrow{\theta}_t - \overrightarrow{\theta}_{S,t})\frac{1}{\sigma^2_{\hat{\theta}}})$.

5. $\underline{[\overrightarrow{a}|\cdot]}$

Let C_a be defined parallel to C_c in terms of a_{NS} and a_{EW}.

$[\overrightarrow{a}|\cdot] \propto [\overrightarrow{a}|\overrightarrow{a_0}, a_{NS}, a_{EW}, \sigma^2_a]\Pi_{t=1}^{T}[\overrightarrow{\theta}_{S,t}|\overrightarrow{\theta}_{S,t-1}, \overrightarrow{a}, \overrightarrow{\beta}, a_N, a_S, a_E, a_W, \overrightarrow{M}_{t-1}, \kappa, \sigma^2_{\hat{\theta}_S}]$

$\propto exp\{-\frac{1}{2}[\frac{1}{\sigma^2_a}(\overrightarrow{a} - \overrightarrow{a_0})'(I - C_a)(\overrightarrow{a} - \overrightarrow{a_0})]\}$
$\times exp\{-\frac{1}{2}[\frac{1}{\sigma^2_{\hat{\theta}_S}}\sum_{t=1}^{T}(\overrightarrow{\theta}_{S,t} - (diag(\overrightarrow{a})D^{\kappa}_{t-1} + H_0)\overrightarrow{\theta}_{S,t-1} - D^{\kappa}_{t-1}\overrightarrow{\beta})'$
$(\overrightarrow{\theta}_{S,t} - (diag(\overrightarrow{a})D^{\kappa}_{t-1} + H_0)\overrightarrow{\theta}_{S,t-1} - D^{\kappa}_{t-1}\overrightarrow{\beta})]\}$

$\propto exp\{-\frac{1}{2}[\overrightarrow{a}'((I-C_a)\frac{1}{\sigma^2_a}+\frac{1}{\sigma^2_{\hat{\theta}_S}}\sum_{t=1}^{T}(D^{\kappa}_{t-1}diag(\overrightarrow{\theta}_{S,t-1})D^{\kappa}_{t-1}diag(\overrightarrow{\theta}_{S,t-1})))\overrightarrow{a}$
$-2\overrightarrow{a}'((I-C_a)\overrightarrow{a_0}\frac{1}{\sigma^2_a}+\frac{1}{\sigma^2_{\hat{\theta}_S}}\sum_{t=1}^{T}D^{\kappa}_{t-1}diag(\overrightarrow{\theta}_{S,t-1})(\overrightarrow{\theta}_{S,t}-H_0\overrightarrow{\theta}_{S,t-1}-D^{\kappa}_{t-1}\overrightarrow{\beta}))]\}$

This gives

$[\overrightarrow{a}|\cdot] \sim N(\overrightarrow{\overline{a}}, \Sigma_{\overrightarrow{a}})$,

where

$\Sigma_{\overrightarrow{a}} = ((I - C_a)\frac{1}{\sigma^2_a} + \frac{1}{\sigma^2_{\hat{\theta}_S}}\sum_{t=1}^{T}D^{\kappa}_{t-1}diag(\overrightarrow{\theta}_{S,t-1})D^{\kappa}_{t-1}diag(\overrightarrow{\theta}_{S,t-1}))^{-1}$

and

$\overrightarrow{\overline{a}} = \Sigma_{\overrightarrow{a}}((I - C_a)\overrightarrow{a_0}\frac{1}{\sigma^2_a}$
$+\sum_{t=1}^{T}(D^{\kappa}_{t-1}diag(\overrightarrow{\theta}_{S,t-1})(\overrightarrow{\theta}_{S,t} - H_0\overrightarrow{\theta}_{S,t-1} - D^{\kappa}_{t-1}\overrightarrow{\beta}))\frac{1}{\sigma^2_{\hat{\theta}_S}})$.

6. $\underline{[\overrightarrow{\beta}|\cdot]}$

Let C_b be defined parallel to C_c in terms of b_{NS} and b_{EW}.
$[\overrightarrow{\beta}|\cdot] \propto [\overrightarrow{\beta}|\overrightarrow{\beta}_0, b_{NS}, b_{EW}, \sigma^2_{\beta}]\Pi_{t=1}^{T}[\overrightarrow{\theta}_{S,t}|\overrightarrow{\theta}_{S,t-1}, \overrightarrow{a}, \overrightarrow{\beta}, a_N, a_S, a_E, a_W, \overrightarrow{M}_{t-1}, \kappa, \sigma^2_{\hat{\theta}_S}]$

$\propto exp\{-\frac{1}{2}[\frac{1}{\sigma^2_{\beta}}(\overrightarrow{\beta} - \overrightarrow{\beta}_0)'(I - C_b)(\overrightarrow{\beta} - \overrightarrow{\beta}_0)]\}$
$\times exp\{-\frac{1}{2}[\frac{1}{\sigma^2_{\hat{\theta}_S}}\sum_{t=1}^{T}(\overrightarrow{\theta}_{S,t} - (diag(\overrightarrow{a})D^{\kappa}_{t-1} + H_0)\overrightarrow{\theta}_{S,t-1} - D^{\kappa}_{t-1}\overrightarrow{\beta})'$

$(\overrightarrow{\theta}_{S,t} - (diag(\overrightarrow{a})D^{\kappa}_{t-1} + H_0)\overrightarrow{\theta}_{S,t-1} - D^{\kappa}_{t-1}\overrightarrow{\beta})]\}$

$\propto exp\{-\frac{1}{2}[\overrightarrow{\beta}'((I - C_b)\frac{1}{\sigma^2_{\beta}} + \frac{1}{\sigma^2_{\hat{\theta}_S}}\sum_{t=1}^{T} D^{\kappa}_{t-1}D^{\kappa}_{t-1})\overrightarrow{\beta}$
$- 2\overrightarrow{\beta}'((I - C_b)\overrightarrow{\beta_0}\frac{1}{\sigma^2_{\beta}} + \frac{1}{\sigma^2_{\hat{\theta}_S}}\sum_{t=1}^{T} D^{\kappa}_{t-1}(\overrightarrow{\theta}_{S,t} - (diag(\overrightarrow{a})D^{\kappa}_{t-1} + H_0)\overrightarrow{\theta}_{S,t-1}))]\}.$

This gives

$[\overrightarrow{\beta}|\cdot] \propto\sim N(\overrightarrow{\overline{\beta}}, \Sigma_{\overrightarrow{\beta}})$

where,

$\Sigma_{\overrightarrow{\beta}} = ((I - C_b)\frac{1}{\sigma^2_{\beta}} + \frac{1}{\sigma^2_{\hat{\theta}_S}}\sum_{t=1}^{T} D^{\kappa}_{t-1}D^{\kappa}_{t-1})^{-1}$

and

$\overrightarrow{\overline{\beta}} = \Sigma_{\overrightarrow{\beta}}((I - C_b)\overrightarrow{\beta_0}\frac{1}{\sigma^2_{\beta}} + \frac{1}{\sigma^2_{\hat{\theta}_S}}\sum_{t=1}^{T} D^{\kappa}_{t-1}(\overrightarrow{\theta}_{S,t} - (diag(\overrightarrow{a})D^{\kappa}_{t-1} + H_0)\overrightarrow{\theta}_{S,t-1})).$

7. $[a_N|\cdot]$

Parallel to *Wikle et al. (1998)* let H_{0a_N} be the H_0 matrix with the diagonal corresponding to the a_N coefficient replaced by zeros, and $\overrightarrow{\theta}^{a_N}_{S,t-1} = J_{a_N}\overrightarrow{\theta}_{S,t-1}$, where J_{a_N} is the $X \times X$ matrix with ones along the diagonal corresponding to the a_N coefficient in H_0, and with zeros elsewhere.

$[a_N|\cdot] \propto \Pi^{T}_{t=1}[\overrightarrow{\theta}_{S,t}|\overrightarrow{\theta}_{S,t-1}, \overrightarrow{a}, \overrightarrow{\beta}, a_N, a_S, a_W, a_E, \overrightarrow{M}_{t-1}, \kappa, \sigma^2_{\hat{\theta}_S}][a_N|\widetilde{a_N}, \widetilde{\sigma}^2_{a_N}]$

$\propto\ exp\{-\frac{1}{2}[\frac{1}{\sigma^2_{\hat{\theta}_S}}\sum_{t=1}^{T}(\overrightarrow{\theta}_{S,t} - (diag(\overrightarrow{a})D^{\kappa}_{t-1} + H_{0a_N})\overrightarrow{\theta}_{S,t-1} - a_N\overrightarrow{\theta}^{a_N}_{S,t-1} -$
$D^{\kappa}_{t-1}\overrightarrow{\beta})'(\overrightarrow{\theta}_{S,t} - (diag(\overrightarrow{a})D^{\kappa}_{t-1} + H_{0a_N})\overrightarrow{\theta}_{S,t-1} - a_N\overrightarrow{\theta}^{a_N}_{S,t-1} - D^{\kappa}_{t-1}\overrightarrow{\beta})]\}$
$\times exp\{-\frac{1}{2}\frac{1}{\widetilde{\sigma}^2_{a_N}}(a_N - \widetilde{a}_N)^2\}$

This gives

$[a_N|\cdot] \sim N(\overline{a}_N, \sigma^2_{a_N})$

where

$\sigma^2_{a_N} = (\sum_{t=1}^{T} \overrightarrow{\theta}^{a_N}_{S,t-1}{}' \overrightarrow{\theta}^{a_N}_{S,t-1} \frac{1}{\sigma^2_{\hat{\theta}_S}} + \frac{1}{\widetilde{\sigma}^2_{a_N}})^{-1}$

and

$\overline{a}_N = \sigma^2_{a_N}(\sum_{t=1}^{T} \overrightarrow{\theta}^{a_N}_{S,t-1}{}'(\overrightarrow{\theta}_{S,t} - (diag(\overrightarrow{\alpha})D^{\kappa}_{t-1} + H_{0a_N})\overrightarrow{\theta}_{S,t-1} - D^{\kappa}_{t-1}\overrightarrow{\beta}) \frac{1}{\sigma^2_{\hat{\theta}_S}} + \frac{\widetilde{a}_N}{\widetilde{\sigma}^2_{a_N}})$.

Distributions for a_E, a_W and a_S are obtained in the same way.

8. $\underline{[c_{NS}|\cdot]}$ and $\underline{[c_{EW}|\cdot]}$,

As opposed to *Wikle et al. (1998)* we start out considering the full joint conditional distribution of c_{NS} and c_{EW}.

$[c_{NS}, c_{EW}|\cdot] \propto [\overrightarrow{\mu}\,|\,\overrightarrow{\mu}_0, c_{NS}, c_{EW}, \sigma^2_{\mu}][c_{NS}|\widetilde{c}_{NS}, \widetilde{\sigma}^2_{c_{NS}}][c_{EW}|\widetilde{c}_{EW}, \widetilde{\sigma}^2_{c_{EW}}]$

$\propto \frac{1}{|(I-C_c)^{-1}|^{1/2}} exp\{-\frac{1}{2\sigma^2_{\mu}}[(\overrightarrow{\mu} - \overrightarrow{\mu}_0)'(I - C_c)(\overrightarrow{\mu} - \overrightarrow{\mu}_0)]\}$
$\times exp\{-\frac{1}{2\widetilde{\sigma}^2_{c_{NS}}}(c_{NS} - \widetilde{c}_{NS})^2\} \times exp\{-\frac{1}{2\widetilde{\sigma}^2_{c_{EW}}}(c_{EW} - \widetilde{c}_{EW})^2\}$

Remember that c_{NS} and c_{EW} are each on two of the four off-diagonals of C_c. To ensure positive definiteness of the matrix $(I - C_c)$ we have the constraints that both c_{NS} and c_{EW} are positive and that $c_{NS} + c_{EW} \leq 0.5$, see Besag and Kooperberg (1995). The presence of c_{NS} and c_{EW} in the determinant term makes this a difficult distribution to sample from. We employ a Metropolis-Hastings step in the Gibbs sampling with a proposal distribution for c_{NS} given as in *Wikle et al. (1998)* by

$[c_{NS^{Pseudo}}|\cdot] \sim N(\overline{c}^{Pseudo}_{NS}, \sigma^2_{c^{Pseudo}_{NS}})$,

where
$\sigma^2_{c^{Pseudo}_{NS}} = (\sum_{x=1}^{X}(\mu^{N(x)} - \mu_0^{N(x)} + \mu^{S(x)} - \mu_0^{S(x)})^2 \frac{1}{\sigma^2_{\mu}} + \frac{1}{\widetilde{\sigma}^2_{c_{NS}}})^{-1}$

and

$\overline{c}^{Pseudo}_{NS} = \sigma^2_{c^{Pseudo}_{NS}}[\sum_{x=1}^{X}(\mu^{N(x)} - \mu_0^{N(x)} + \mu^{S(x)} - \mu_0^{S(x)})$
$\times (\mu^x - \mu_0^x - c_{EW}(\mu^{E(x)} - \mu_0^{E(x)} + \mu^{W(x)} - \mu_0^{W(x)})) \frac{1}{\sigma^2_{\mu}} + \frac{\widetilde{c}_{NS}}{\widetilde{\sigma}^2_{c_{NS}}}]$.

The proposal distribution for c_{EW} is analogous. We then sample from $[c_{NS^{Pseudo}}|\cdot]$ and $[c_{EW^{Pseudo}}|\cdot]$. If either of the drawn samples is less than 0 and/or if their sum

is greater than 0.5 both the drawn samples are rejected, since they are not valid proposals. If not, a Metropolis-Hastings step is performed to decide if the valid proposals are to be accepted or rejected. We consider the full constrained joint conditional distribution as target distribution. An advantage of our approach is that the determinant only has to be computed once in each iteration.

Distributions for (a_{NS}, a_{EW}) and (b_{NS}, b_{EW}) are obtained in the same way.

9. $\underline{[\kappa|\cdot]}$

$$[\kappa|\cdot] \propto \Pi_{t=1}^{T}[\overrightarrow{\theta}_{S,t}|\overrightarrow{\theta}_{S,t-1}, \overrightarrow{a}, \overrightarrow{\beta}, a_N, a_S, a_W, a_E, \overrightarrow{M}_{t-1}, \kappa, \sigma^2_{\hat{\theta}_S}] \times [\kappa|\widetilde{\kappa}, \widetilde{\sigma}^2_{\kappa}]$$

$$\propto \Pi_{t=1}^{T} exp\{-\tfrac{1}{2}[\tfrac{1}{\sigma^2_{\hat{\theta}_S}}(\overrightarrow{\theta}_{S,t} - (diag(\overrightarrow{a})D^{\kappa}_{t-1} + H_0)\overrightarrow{\theta}_{S,t-1} - D^{\kappa}_{t-1}\overrightarrow{\beta})'$$
$$(\overrightarrow{\theta}_{S,t} - (diag(\overrightarrow{a})D^{\kappa}_{t-1} + H_0)\overrightarrow{\theta}_{S,t-1} - D^{\kappa}_{t-1}\overrightarrow{\beta})]\} \times exp\{-\tfrac{1}{2\tilde{\sigma}^2_{\kappa}}(\kappa - \tilde{\kappa})^2\}$$

$$\propto exp\{-\tfrac{1}{2}\textstyle\sum_{t=1}^{T}[((diag(\overrightarrow{a})\overrightarrow{\theta}_{S,t-1} + \overrightarrow{\beta})'D^{\kappa}_{t-1}D^{\kappa}_{t-1}(diag(\overrightarrow{a})\overrightarrow{\theta}_{S,t-1} + \overrightarrow{\beta})$$
$$- 2(diag(\overrightarrow{a})\overrightarrow{\theta}_{S,t-1} + \overrightarrow{\beta})'D^{\kappa}_{t-1}(\overrightarrow{\theta}_{S,t} - H_0\overrightarrow{\theta}_{S,t-1})\tfrac{1}{\sigma^2_{\hat{\theta}_S}}]\}$$
$$\times exp\{-\tfrac{1}{2\tilde{\sigma}^2_{\kappa}}(\kappa - \tilde{\kappa})^2\}$$

We see that κ enters the equation above on the right hand side through D^{κ}_{t-1}, and we employ a Metropolis-Hastings step in the Gibbs Sampling with proposal distribution $\pi_0(\kappa)$ chosen to be a Normal distribution with expectation equal to the κ value at the previous Gibbs step and variance equal to 0.01. But other distributions are possible, for instance a beta distribution.

Parallel to Wikle et al. (1998) we get:

10. $\underline{[\sigma^2_{\hat{\theta}}|\cdot]}$

$$[\sigma^2_{\hat{\theta}} \mid \cdot] \sim IG(XT/2 + \widetilde{q}_{\sigma^2_{\hat{\theta}}}, [\textstyle\sum_{t=1}^{T}(\overrightarrow{\theta}_t - \overrightarrow{\mu} - \overrightarrow{\theta}_{S,t})'(\overrightarrow{\theta}_t - \overrightarrow{\mu} - \overrightarrow{\theta}_{S,t})/2 + 1/\widetilde{r}_{\sigma^2_{\hat{\theta}}}]^{-1})$$

11. $\underline{[\sigma^2_{\hat{\theta}_S}|\cdot]}$

$$[\sigma^2_{\hat{\theta}_S} \mid \cdot] \sim IG(XT/2 + \widetilde{q}_{\sigma^2_{\hat{\theta}_S}}, [\textstyle\sum_{t=1}^{T}(\overrightarrow{\theta}_{S,t} - (diag(\overrightarrow{a})D^{\kappa}_{t-1} + H_0)\overrightarrow{\theta}_{S,t-1} -$$
$$D^{\kappa}_{t-1}\overrightarrow{\beta})'$$
$$(\overrightarrow{\theta}_{S,t} - (diag(\overrightarrow{a})D^{\kappa}_{t-1} + H_0)\overrightarrow{\theta}_{S,t-1} - D^{\kappa}_{t-1}\overrightarrow{\beta})/2 + 1/\widetilde{r}_{\sigma^2_{\hat{\theta}_S}}]^{-1})$$

12. $\underline{[\sigma_\mu^2|\cdot]}$

$$[\sigma_\mu^2 \mid \cdot] \sim IG(X/2 + \widetilde{q}_{\sigma_\mu^2}, [(\overrightarrow{\mu} - \overrightarrow{\mu}_0)'(I - C_c)(\overrightarrow{\mu} - \overrightarrow{\mu}_0)/2 + 1/\widetilde{r}_{\sigma_\mu^2}]^{-1})$$

Distributions for σ_a^2 and σ_β^2 are completely similar.

13. $\underline{[\sigma_M^2|\cdot]}$

$$[\sigma_M^2 \mid \cdot] \sim IG(XT/2 + \widetilde{q}_{\sigma_M^2}, [\textstyle\sum_{t=1}^{T}(\overrightarrow{M}_t - \overrightarrow{\theta}_t)'(\overrightarrow{M}_t - \overrightarrow{\theta}_t)/2 + 1/\widetilde{r}_{\sigma_M^2}]^{-1})$$

REFERENCES

Besag, J., and Kooperberg, C. (1995), "On Conditional and Intrinsic Autoregression," *Biometrika*, 82, 733 – 746.

Cressie, N.A.C (1993), *Statistics for Spatial Data*, (revised ed.), New York: Wiley.

Geller, R.J., Jackson, D.D., Kagan, Y.Y., and Mulargia, F. (1997), "Earthquakes Cannot Be Predicted," *Science*, 275, 1616 – 1617.

Holden, L., Natvig, B., Sannan, S., Abrahamsen, P., and Bungum, H., (2002), "Modelling Spatial and Temporal Dependencies between Earthquakes," in *GEOSTATS 2000, Proceedings of the Sixth International Geostatistics Congress held in Cape Town, South Africa, in April 2000* (Vol. 2), eds. W.J. Kleingeld and D.G. Krige, pp. 439–448.

Ogata, Y. (1998), "Space-Time Point Process Models for Earthquake Occurrences," *Annals of the Institute of Statistical Mathematics*, 50, 379 – 402.

Price, E.J. and Sandwell, D.T. (1998), "Small-Scale Deformations associated with the 1992 Landers, California, Earthquake mapped by Synthetic Aperture Radar Interferometry Phase," *Journal of Geophysical Research-Solid Earth*, 103, 27001 – 27016.

Schoenberg, F.P. (2003), "Multidimensional Residual Analysis of Point Process Models for Earthquake Occurrences," *Journal of the American Statistical Association*, 98, 798 – 795.

Vere-Jones, D. (1995), "Forecasting Earthquakes and Earthquake Risk," *International Journal of Forecasting*, 11, 503 – 538.

West, M., and Harrison, J. (1997), *Bayesian Forecasting and Dynamic Models*, (2nd ed.), New York: Springer Verlag.

Wikle, C.K., Berliner, L.M., and Cressie, N.A.C. (1998), "Hierarchical Bayesian Space-Time Models," *Environmental and Ecological Statistics*, 5, 117 – 154.

Wikle, C.K., Milliff, R.F., Nychka, D., and Berliner, L.M. (2001), "Spatiotemporal Hierarchical Bayesian Modeling: Tropical Ocean Surface Winds," *Journal of the American Statistical Association*, 96, 382 – 397.