Bayesian Earthquake-Risk Estimation

Harold A. Hernández Roig

December 3, 2018

1 INTRODUCTION

In this study we face the problem of estimating the earthquake risk at a cube near Fiji. We have data from 1000 seismic events that occurred since 1964, although we have no knowledge or dependency on time for them. In order to describe better our data we introduce some terminology from earthquake research field [11]:

- *Epicenter:* The latitude and longitude of the point of first motion of the earthquake. We have this information in variables $lat \in [-38.59°; -10.72°]$ and $long \in [165.7°; 188.1°]$ respectively.
- *Depth:* The depth below the surface of the point of first motion. For our data, every observation was recorded in the variable *depth*, with range [40; 680] km.
- *Magnitude:* A measure of the size of the event, quoted in this case in the Richter scale. All the earthquakes reported in our dataset have the corresponding magnitude in variable $mag \geq 4.0$, being 6.40 the highest value reported (all in log scale).
- *Stations:* Usually the number and locations of stations reporting the event. We only have the number in variable *stations*, with a range of values in [10; 132].

The term "earthquake risk" is in fact loosely used. In order to distinguish from other studies, we follow the terminology from [11] and assume as definition:

> "The *geophysical risk* (or *earthquake hazard*) is the expected rate of occurrence (number of events per unit time) within a prescribed region and over certain magnitude level."

As we have no information regarding the *origin time*, we focus on estimating the geophysical earthquake risk of the observed area, without dependence on time. This could be thought as the *long-term average*, *static*, or *background risk*.

1.1 DATA TRANSFORMATION

In order to estimate "number of events" we construct a set of grid cells of $0.5° \times 0.5°$ (approximately $50 \times 50 km^2$, although correspondence from degrees to kilometers is not the same in every region of the earth surface), analogously to [8]. Then we count the number of occurrences (our response variable Y in the future model) per grid cell and take the average depth, magnitude and number of stations reporting for each of them (these could be regarded as a 3-covariates matrix X). This data transformation leads to a set of 2726 cells or locations

over the observed region, for which we have only 366 non-zero counts on $\{Y; X\}$, those corresponding to locations where at least a single earthquake occurred. We also take logarithms of the variables *depth* and *stations*, so they are in the same scale as *mag*. Codes for data transformation are available at *transform_data.R*.

We propose two approaches based on Bayesian Poisson Regression. In Section 2 we regress the non-zero responses in terms of the covariates in matrix X, including a (previous) Bayesian Variable Selection procedure. In Section 3 we take into account the spatial structure of the observed area, thus we include an Intrinsic Conditional Auto-Regressive (ICAR) component, in order to explain the correlation between neighboring regions. This is known as a Besag-York-Mollié (BYM) model. All computations are carried out with an Intel(R) Core(TM) i7-6500U CPU @ 2.50GHz 2.59GHz and RAM 12.0GB.

2 Bayesian Poisson Regression and Variable Selection

Consider the 366 non-zero counts over the constructed grid. The aim is to regress this values in terms of the observed average magnitude, depth and number of stations reporting. We choose a Poisson model to describe this:

$$Y|X \sim \text{Poisson}\big(\exp(X\beta)\big) \tag{1}$$

where β is the vector of parameters $(\beta_0, \beta_1, \beta_2, \beta_3)^t$, to be estimated. Each indexing $i = 0, 1, 2, 3$ corresponds to intercept, average magnitude, depth and number of stations, respectively. We start by performing a Bayesian Variable Selection and then we use the "selected" variables to fit a Poisson Regression with Stan.

2.1 Variable selection

We use R-package *BAS* for Bayesian Variable Selection on model defined by eq.1. This package allows us to find high probability models, which can therefore be used to select the "relevant" variables. The core of the package is the implementation of Bayesian Adaptive Sampling [4]. As major drawback to our application is that this procedure works well when variables has small correlation, and this is not guaranteed (at least intuitively) for our observations (for example: stations could depend on magnitude).

We proceed by fixing *"MCMC + BAS"* as method, and set an Empirical Bayes Prior Distribution (*"EB-prior"*) for the coefficients. The method uses an initial Metropolis Hasting algorithm to calculate marginal inclusion probabilities and then samples without replacement with Bayesian Adaptive Sampling. The EB-prior allows us to select an initial distribution depending on data. This could be thought of as unbiased but weak prior information about the betas. All codes are available in script *bas_poi_regr.R*.

Figure 1 gives a graphical representation of the "top models". In there, the models are sorted by their posterior probabilities from best (left) to worst (right). We observe that the "best" is

the one having all 3 variables and the intercept. Moreover, posterior inclusion probabilities (pip) are 1 for all the variables (check codes), which indicates they are relevant to explain the response.

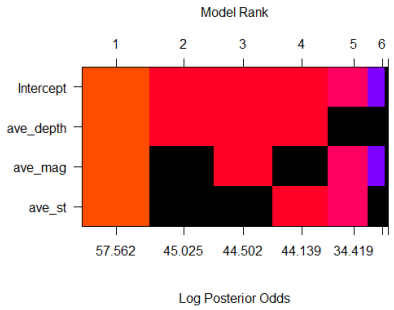
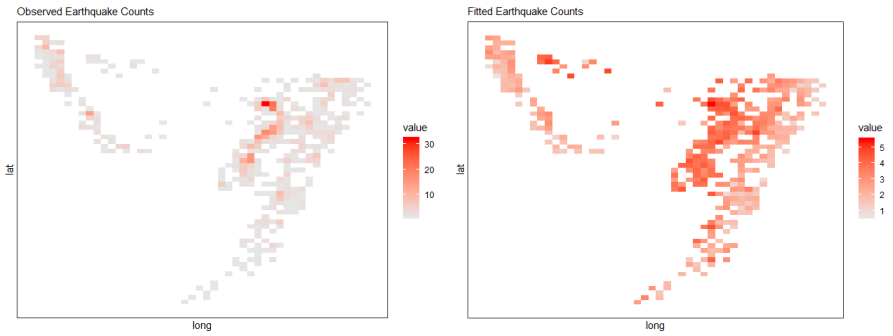


Figure 1: Top models in terms of posterior probability. The "best" is the one having all 3 variables and the intercept.

We include a summary of the coefficient distributions, under Bayesian Model Averaging (BMA). We observe they all have thin Highest Posterior Density intervals (not as thin for the intercept), and none of them includes the 0 :

	2.5%	97.5%	mean.beta
Intercept	0.5481389	3.1509962	1.8856484
ave_depth	0.1966188	0.3528359	0.2768945
ave_mag	-1.3623147	-0.6378312	-0.9900182
ave_st	0.4108760	0.9011038	0.6627843

Moreover, we can check fitted responses under the BMA model. Figure 2 shows two spatial plots: (a) observed values in the given region, and (b) mean of the fitted values in each cell-grid. It is straightforward to infer that this model cannot estimate "peaks" describing zones with high occurrences of events!



(a) Observed earthquake counts (b) Fitted earthquake counts (mean value)

Figure 2: We observe that the (mean) fitted values cannot estimate correctly high values of the observed response variable.

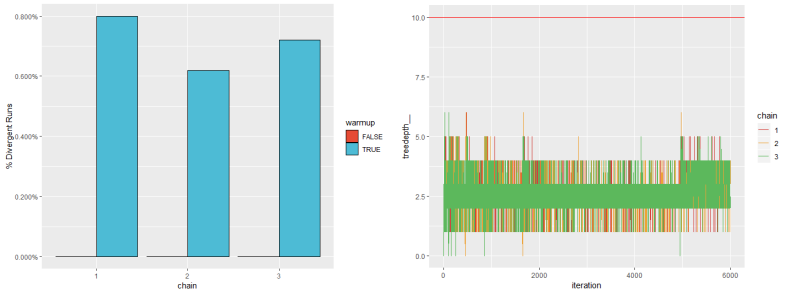
2.2 POISSON REGRESSION WITH STAN

Motivated by Subsection 2.1, we accept that all 3 covariates in X are relevant to fit a Poisson model. Nevertheless, we saw that the BMA model does not arise good fitted values, since (at least in mean) cannot recover peaks representing regions with high occurrences of events. We therefore use RStan [10] to adjust the Poisson model in terms of the covariates in X. We relay on the advantage of diagnosing models in Stan to decide whether our proposal is good or not. Codes for fitting the model are available in *fit_poi_regr_weak_prior.R* and *poi_regr_weak_prior.stan*.

Following [5], we fix the "default" weakly informative prior distribution for our coefficients: $\beta \sim \text{Cauchy}(0; 2.5)$. We scaled our data (just inputs variables) to be centered at 0 and with standard deviation 0.5. We should add, that recent studies also indicates the usage of *t-Student* distributions $\beta \sim \text{student\_t}(\nu, 0, s)$ but with $3 < \nu < 7$ and s chosen to provide weak information on the expected scale. This ideas were extracted from RStan-Wiki: "Prior Choice Recommendations" [1]. We also set 3 chains and 6000 iterations with a warm-up of 5000.

Although not included, the estimated coefficients showed similar results: there are not changes in sign for the estimated mean value and the credible intervals are thin. The only difference arise from the different scale in the input data.

Let's diagnose the Stan-fit (code available at check_fit_poi_regr_weak_prior.R). We start by checking convergence. Figure 3 (a) shows there are no divergences during the sampling (post warm-up) period. Next we check the *treedepth* (default maximum value is 10). This parameter acts as a "cap" restricting the draws from the Hamiltonian Markov Chain procedure. If our model "hits" this value too many times, it means that we are selecting draws based on hitting this maximum value, rather than properties of the posterior probability. Figure 3 (b) shows we do not have such a problem since all iterations have treedepth values below 10.

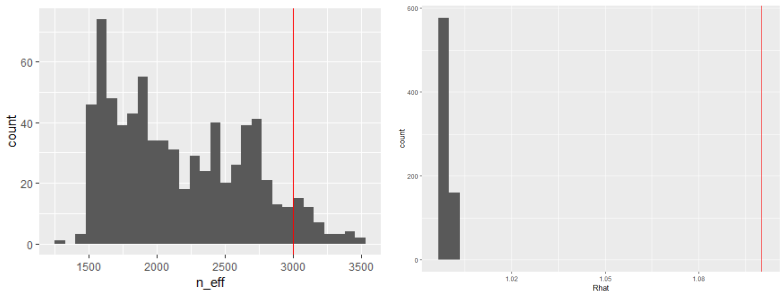


(a) Percent of divergent transitions during warm-up and post warm-up. We had no divergences during the sampling period in any chain.

(b) Treedepths for every iteration were below the maximum allowable.

Figure 3: Diagnose on divergence and treedepth.

We turn now to diagnose individual parameter estimates. There are two basic statistics that helps us on this: the effective sample size (*n_eff*) and the Gelman and Rubin potential scale reduction statistic (*Rhat*) [3] [6]. The *n_eff* measures the amount of effectively independent sampling iterations across all 3 chains ($\leq 3 \times 1000$). The *Rhat* statistics can be thought of as a value describing whether a chain has reached a stable posterior distribution. Following [3] [6], we should be aware of values of $Rhat > 1.1$. Figure 4 (a) shows that most of our parameters have $n\_eff > 1000$. Moreover, in Figure 4 (b) all values of $Rhat < 1.1$, so we confirm these effective sample sizes are good enough.



(a) *n_eff* (b) *Rhat* statistic

Figure 4: Diagnose on parameters. We have no problems with *n_eff* nor the *Rhat* statistic.

So far now our model looks quite good, we then perform some graphical Posterior Predictive Checks (PPcheck). Stan helps us by dropping realizations from the posterior distribution of the outcomes, thus we focus on the histograms of the observed Y and some replicates for it. Figure 5 shows that replicated outcomes cannot reproduce the overall shape of the observed data. Moreover, the mean of fitted values for the cell-grids cannot recover the peaks representing regions with high occurrences of events (figure is not included in the report), showing the same behavior as in Figure 2. Although diagnose is encouraging, the fitted values are not that good.

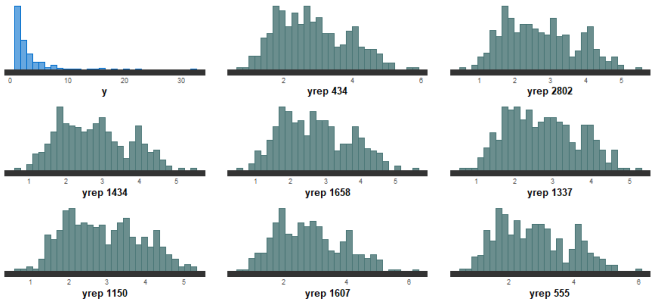


Figure 5: PPcheck (taken from ShinyStan). The replicated values cannot reproduce the overall shape of the observed data.

3 Spatial Model: BYM model

In this section we turn to adjust a BYM model [2], which is a lognormal Poisson model originally proposed to count data in biostatistics and epidemiology. Such a model includes an ICAR component for spatial smoothing and an ordinary random-effects component for non-spatial heterogeneity:

$$Y_i|\psi_i \sim \text{Poisson}(\exp(\psi_i)); \qquad i = 1,\dots,N \tag{2}$$

where $\psi = X\beta + \theta + \phi$. We already declared X as the matrix of covariates (now including a column of 1's to account for the intercept). θ is an ordinary random-effects component for non-spatial heterogeneity, and ϕ is the ICAR spatial component.

Following [7] (see references therein), for a set of N different areal units of a region, spatial interactions between a pair of units N_i and N_j ($i \sim j$ defines neighbors) can be modeled conditionally as a spatial random variable $\phi = (\phi_1,\dots,\phi_N)$. The corresponding conditional distribution specification for the components can be written in terms of the amount of neighbors $d_{i,i}$ (define diagonal entries of a matrix D) for region N_i and a precision parameter τ:

$$p(\phi_i|\phi_j, j \neq i, \tau_i^{-1}) = \text{Normal}\left(\frac{\sum_{i\sim j}\phi_i}{d_{i,i}}, \frac{1}{d_{i,i}\tau_i}\right)$$

Moreover, [7] shows that this last expression can be rewritten in a *pairwise difference formulation*:

$$p(\phi|\tau) \propto \tau^{\frac{N-NC}{2}} \exp\left\{-\frac{\tau}{2}\sum_{i\sim j}\left(\phi_i - \phi_j\right)^2\right\}$$

adding the constraint that $\sum_i \phi_i = 0$, for identifiability of the joint distribution (in practice, we set up a "soft sum-to-zero constraint" using a prior on ϕ which tightly constrains its mean to be within some epsilon of zero). The term NC is the number of components in the graph over all areal subregions defined by the spatial proximity among components. We set $NC = 1$ in our case, because we consider a fully connected areal graph (every subregion can be reached from every other subregion via a sequence of neighbors). Inside the BYM, the ICAR specification is used as part of a hierarchical model.

We specifically follow the idea of the BYM2 model exposed in [7] (and taken from [9]), which improves the parameterization of $(\theta + \phi)$ by placing a single precision parameter on the combined component, and mixing the parameters for the amount of spatial/non-spatial variation:

$$\theta + \phi = \sigma\left(\sqrt{1-\rho}\,\theta^* + \sqrt{\rho}\,\phi^*\right)$$

where $\sigma \geq 0$ is the overall standard deviation; $\rho \in [0,1]$ models how much of the variance comes from the spatially structured effect and how much comes from the spatially unstructured effect; $\theta^* \sim \text{Normal}(0, I)$ is the unstructured random effect with fixed standard deviation (= 1); and ϕ^* is the ICAR model scaled so $\text{Var}(\phi_i) = 1$. Although we won't cover the

details, for each i the model is restricted to obtain $\text{Var}(\phi_i) \approx \text{Var}(\theta_i) \approx 1$, in order for σ to "legitimately" be the standard deviation of the random effect. This is carried out in our code by scaling the model so the geometric mean of these variances is 1, using package R-INLA on the adjacency matrix (W, a $N \times N$ sparse matrix with non-zero components w_{ij}, w_{ji} whenever $i \sim j$).

3.1 Fittting a BYM2 model on Quakes data

The optimal choice for encoding adjacencies (neighboring regions) in Stan is through an undirected graph with set of N nodes and a set of edges, one per pair $i \sim j$ and $j \sim i$. We therefore create two parallel arrays *node_1* and *node_2*, each with length N_edge, describing neighboring cell-grids. This is carried out in script *transform_data.R*, resulting on $N = 2726$ (amount of grid-cells) and $N\_edge = 21182$, which corresponds to the 8 grid-cell neighbors (fewer on the borders of observed area) around a given location.

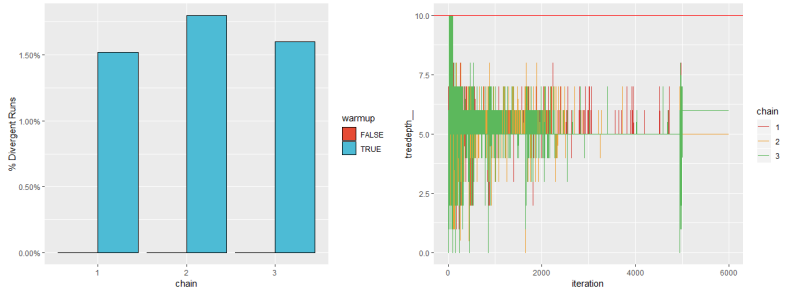
We decide to adjust the BYM2 model to our transformed-quakes data,but including only the intercept term β_0 in eq.2. We actually tried several scenarios including the 3 covariates as in Section 2 but divergences problems made us fix a bigger *treedepth* value and number of iterations. Even though, we could not reach an acceptable solution in a reasonable time (50% of the iterations were taking around 10 hours).

Codes for adjusting the BYM2 model on quakes data are available in fit_quakes_bym2_no_beta.R and bym2_no_offset_no_beta.stan. Following recommendations from [7], we set a standard half-normal prior on σ, a Beta(0.5;0.5) as prior for ρ and a Normal(0;2.5) on the β_0. As in previous Section we run 6000 iterations (5000 for warm-up) on 3 chains, and leave all other parameters in their default values. We now conduct a diagnose on the obtained Stan-fit (codes available at check_fit_quakes_bym2_no_beta.R).

We start by checking if there is any divergence problem in our transitions (after burn-up). Figure 6 (a) shows there are no divergences during the sampling (post warm-up) period. On the other hand Figure 6 (b) shows we have some iterations "touching" the maximum allowable treedepth value. Nevertheless, it does not seem to be a common behavior, so we conclude there are no divergence/treedepth problems

Let's diagnose individual parameter estimates through the *n_eff* and the *Rhat* statistic. Figure 7 (a) shows that, even though most of our parameters have a high effective sample size, there are some with lower values. Therefore we contrast this information with results from Figure 7 (b). As all values of $Rhat < 1.1$, we confirm these effective sample sizes are good enough.

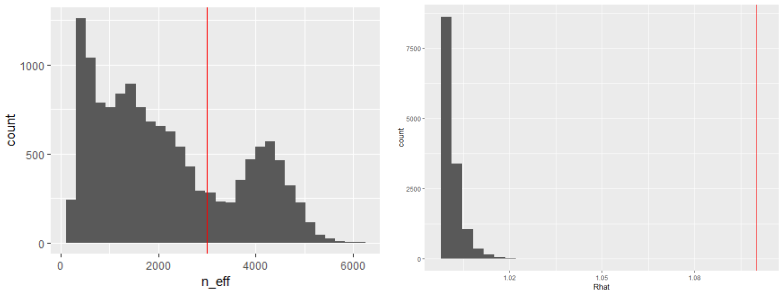
Finally, we do some PPcheck using ShinyStan. Figure 8 shows the histogram of the observed data, as well as those for some replications. Now, we have better results compared to those from Subsection 2.2; the replicated values do reproduce the overall shape of the observed data. Moreover, there are some small bars representing what we already called "peaks", over



(a) Percent of divergent transitions during warm-up and post warm-up. We see that across all 3 chains, we had no divergences during the sampling period

(b) Treedepths for most of the iterations accross chains were below the maximum allowable.

Figure 6: Diagnose on divergence and treedepth. Spatial Model.



(a) *n_eff*

(b) *Rhat* statistic

Figure 7: Diagnose on parameters from the Spatial Model. Although some *n_eff* seem to be low, the *Rhat* statistic guarantees that those sample sizes are good enough.

regions with many event occurrences. We also illustrate in Figure 9 the mean of the fitted values for each cell-grid and compare it to the observations. It is straightforward to infer that the BYM2 model can recover the earthquakes occurrences over the observed area. Also, the neighboring structure implies smoother and more spread estimates around areas where events were recorded. We can therefore encourage the use of this model for earthquake risk estimation, pointing out that it could be improved with real information about the neighboring structure, for example, by considering as neighbors only those regions around plate junctions.

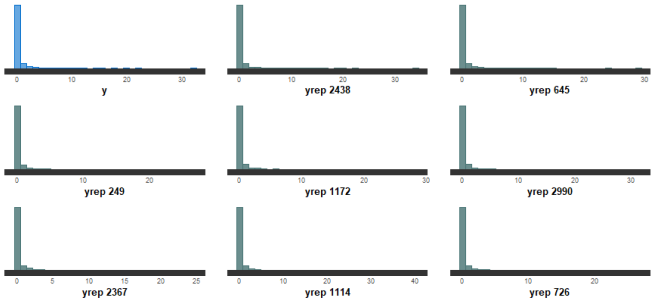
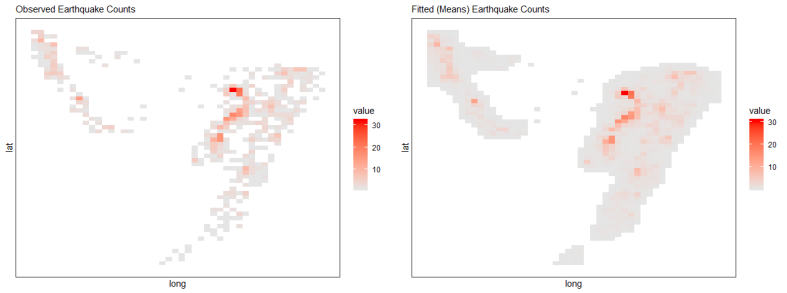


Figure 8: PPcheck for the Spatial Model (taken from ShinyStan). The replicated values seem to reproduce the overall shape of the observed data. Although the obvious is that the replicates reproduce quite well the zero-inflation, there are some small bars representing what we already called "peaks".



(a) Observed earthquake counts (b) Fitted earthquake counts (mean value)

Figure 9: We observe now that the (mean) fitted values can estimate with good accuracy the peaks over regions with higher occurrences of events. Also, the neighboring structure implies smoother and more spread estimates around areas where events were recorded. Spatial Model.

4 Conclusions

- We showed two main models to estimate *geophysical* earthquake risk on an area around Fiji, based on an earthquake database with time-independent occurrences.
- We performed a Bayesian Variable Selection (with R-package BAS) procedure to sustain the usage of a Poisson regression to estimate earthquake occurrence over a grid-cell of the observed area, in terms of the average depth, magnitude and number of stations reporting.
- We also obtained a BMA model that could not estimate correctly the peaks over the observed area.
- We fitted a Bayesian Poisson Regression with weakly informative priors using RStan and including the 3 variables. Although the model diagnose was encouraging, the fitted

outcomes showed the same behavior as for the BMA model.

- We introduced an ICAR component into our Poisson model (now a BYM model) to take into account spatial smoothing and random-non-spatial heterogeneity. By fitting without covariate matrix X, the diagnose was correct and the fitted outcomes could reproduce the overall shape of the observed data.

REFERENCES

[1] Prior choice recommendations, https://github.com/stan-dev/stan/wiki/Prior-Choice-Recommendations. Consulted on November 29th, 2018.

[2] Julian Besag, Jeremy York, and Annie Mollié. Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*, 43(1):1–20, mar 1991.

[3] Stephen P Brooks and Andrew Gelman. General Methods for Monitoring Convergence of Iterative Simulations. *Journal of Computational and Graphical Statistics*, 7(4):434–455, 1998.

[4] Merlise A. Clyde, Joyee Ghosh, and Michael L. Littman. Bayesian adaptive sampling for variable selection and model averaging. *Journal of Computational and Graphical Statistics*, 20(1):80–101, 2011.

[5] Andrew Gelman, Aleks Jakulin, Maria Grazia Pittau, and Yu Sung Su. A weakly informative default prior distribution for logistic and other regression models. *Annals of Applied Statistics*, 2(4):1360–1383, 2008.

[6] Andrew Gelman and Donald B. Rubin. Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science*, 7(4):457–472, 1992.

[7] Mitzi Morris. Spatial models in stan: Intrinsic auto-regressive models for areal data. *Stan Case Studies*, http://mc-stan.org/users/documentation/case-studies/icar_stan.html. Consulted on November 29th, 2018.

[8] Bent Natvig and Ingunn Fride Tvete. Bayesian hierarchical space-time modeling of earthquake data. *Methodology and Computing in Applied Probability*, 9(1):89–114, 2007.

[9] Andrea Riebler, Sigrunn H. Sørbye, Daniel Simpson, Håvard Rue, Andrew B. Lawson, Duncan Lee, and Ying MacNab. An intuitive Bayesian spatial model for disease mapping that accounts for scaling. *Statistical Methods in Medical Research*, 25(4):1145–1165, 2016.

[10] Stan Development Team. RStan: the R interface to Stan, 2018. http://mc-stan.org.

[11] David Vere-Jones. Forecasting earthquakes and earthquake risk. *International Journal of Forecasting*, 11(4):503–538, 1995.