**Reviewer's report:**
**Bayesian Earthquake Dating, GMD-2018-94**

I enjoyed this paper, but I will focus on section 3, as this is where I have some doubts about what the authors have done, and whether it is correct. These doubts stem from their explanation of MCMC at the bottom of p6, which is technically wrong. In line 25, it is the stationary distribution of the Markov chain which converges to the target, and there is a similar misunderstanding in line 29. This convergence implies that Cesàro means converge in mean-square to expectations; i.e. a sample from the chain can be used to estimate expectations. There is also some confusion in sec 3.2 which I will come back to below.

The authors have a statistical model with some parameters $\theta$, a stochastic process $\Pi$, and some parameters controlling the stochastic process, $\varphi$. They do not name $\Pi$ and $\varphi$ explicitly, but $\Pi$ is the stochastic process generating $\{(t_1, d_1), \ldots, (t_N, d_N)\}$. The authors will want to use a complicated $\Pi$, and this means, typically, that it will be simple to sample from $\Pi \mid \varphi$ but very hard to evaluate $\mathrm{p}(\Pi \mid \varphi)$. $\Pi \mid \psi$ is complicated for two reasons. First, realistic earthquake models are much more complicated than marked Poisson processes; second, they will want to impose the constraints that $d_{\min} \leq d_i \leq d_{\max}$, and $\sum_{i=1}^{N} d_i = H_{sc}$. This would never happen by chance in an unconstrained simulation, and so it must be built-in to the prior, as explained at the bottom of p7.

I am skeptical of whether the authors are able to evaluate $\mathrm{p}(\Pi \mid \psi)$ for their complicated model, which is a highly non-linear function of a Brownian motion (it is not clear in the MS whether or this is normal or geometric Brownian motion). I am also confused about the proposal described at the start of sec 3.2. I would like to point out that in a Metropolis-Hastings MCMC scheme they do not need to evaluate $\mathrm{p}(\Pi \mid \psi)$, if they always propose $\Pi \mid \psi$ from the prior. The MH acceptance ratio is

$$\alpha = \frac{L(\theta', \Pi', \psi')\,\mathrm{p}(\theta')\,\mathrm{p}(\Pi' \mid \psi')\,\mathrm{p}(\psi')}{L(\theta, \Pi, \psi)\,\mathrm{p}(\theta)\,\mathrm{p}(\Pi \mid \psi)\,\mathrm{p}(\psi)} \times \frac{\mathrm{prop}(\theta, \psi)\,\mathrm{prop}(\Pi \mid \psi)}{\mathrm{prop}(\theta', \psi')\,\mathrm{prop}(\Pi' \mid \psi')}$$

where $L$ is the likelihood function, primes indicate the proposed values, and 'prop' is the proposal distribution, which may depend on $(\theta, \Pi, \psi)$. So if $\mathrm{prop}(\Pi \mid \psi) = \mathrm{p}(\Pi \mid \psi)$, then the MH acceptance ratio simplifies to

$$\alpha = \frac{L(\theta', \Pi', \psi')\,\mathrm{p}(\theta')\,\mathrm{p}(\psi')}{L(\theta, \Pi, \psi)\,\mathrm{p}(\theta)\,\mathrm{p}(\psi)} \times \frac{\mathrm{prop}(\theta, \psi)}{\mathrm{prop}(\theta', \psi')}.$$

Now it is quite true that this chain will be slow to mix, if the likelihood is highly concentrated. But that is exactly why tempering is a good idea. Tempering is a good trick for whenever we are forced to propose from the prior, owing to the difficulty of computing the probability density. I think there is an interesting message in this paper, which is that this method is applicable even though $\Pi \mid \psi$ is very complicated.

In their MH-MCMC scheme, $\text{prop}(\theta, \varphi)$ is a random walk, although I caution the authors to make sure that they include the Jacobian term if using a transformation: e.g., random walking in logs for non-negative parameters, or logits for bounded parameters, or else be clear about reflection at the boundaries. They might also consult, e.g., Andrieu and Thoms (2008), to implement an adaptive phase at the start of their chain, and to make cautious proposals using the principle components of the estimated posterior variance matrix. All of this needs to be stated in the MS.

The description of the proposal in sec 3.2 is confused. I believe that some issues have been resolved above, by always sampling $\Pi \mid \psi$ from the prior, and tempering. If the authors do not do this, then they will have to give explicit forms for $\text{p}(\Pi \mid \psi)$ and for $\text{prop}(\Pi \mid \psi)$, so that readers can implement the algorithm themselves. I do not understand the Reversible Jump part at all. $\Pi \mid \psi$ is a point process; the fact that the number of components varies is irrelevant.

In terms of diagnostics, the authors will need to demonstrate that their very intricate MCMC does indeed have the correct target distribution, using, e.g., the method of Cook et al. (2006); see also Dan Simpson's update at `http://andrewgelman.com/2018/04/18/better-check-yo-self-wreck-yo-self/`. This will also check that the authors are using a long-enough burn-in, and so separate Gelman-Rubin diagnostics are not required in the MS, although they may be helpful when setting-up the chain. As an aside, 2 independent chains is not nearly enough for Gelman-Rubin: 8 is much better.

Jonathan Rougier
University of Bristol
July 2018

## References

Andrieu, C. and Thoms, J. (2008). A tutorial on adaptive MCMC. *Statistics and Computing*, 18:343–373.

Cook, S., Gelman, A., and Rubin, D. (2006). Validation of software for Bayesian models using posterior quantiles. *Journal of Computational and Graphical Statistics*, 15(3):675–692.