Author Response to Referee Comment 1 (J.T. Rougier)

We thank the referee for the detailed and constructive feedback. Point-by-point replies to the comments are provided below.

1) the explanation of MCMC [...] is technically wrong

> We are aware that, mathematically speaking, what underlies MCMC is the property of the generated samples $(\theta_k)_{k=1}^{\infty}$ that the empirical measures $P_K := 1/K \sum_{k=1}^{K} \delta_{\theta_k}$ converge in some sense to the desired measure $P_{\theta|y}$ as $K \to \infty$. However, our goal is to keep the description of MCMC simple and targeted to a general geoscientific audience.
>
> We agree that we could improve the wording and change
>
> "To gain information about $P_{\theta|y}$, we employ an MCMC method, which generates samples that can be used to approximate statistical properties of $P_{\theta|y}$."
>
> to
>
> "To gain information about the posterior distribution $P_{\theta|y}$, we employ an MCMC method, which generates samples that, roughly speaking, behave as if they were drawn from the posterior and can therefore be used to approximate statistical properties thereof."
>
> One aspect of "behaving as if they were drawn from the desired distribution" is that the distribution of θ_k approaches $P_{\theta|y}$ as $k \to \infty$. We know that there is more to it (more on that in issue 2 below). However, if we attempted a more mathematically precise description and talked about convergence of empirical averages, it may still not be clear to a statistically inexperienced reader how that relates, e.g., to the convergence of kernel density estimates.

2) In line 25, it is the stationary distribution of the Markov Chain which converges to the target

> We assume the referee meant "In line 25, it is the stationary distribution of the Markov Chain which EQUALS the target [namely $P_{\theta|y}$]". We agree with this, but we do not see how it contradicts what we wrote. To be clear, what we meant in line 25 by "the sample distribution converges to the desired distribution as $k \to \infty$" is "$\mathcal{L}(\theta_k) \to P_{\theta|y}$ in some metric". We now make this more clear by saying "the distribution of θ_k converges to the [...]". This convergence does hold for aperiodic and ϕ-irreducible MCMC chains (which we admittedly did not cover in the manuscript). The convergence of Cesàro means, i.e. of the empirical measures defined above, also requires verification of ϕ-irreducibility. Since our chain is obviously aperiodic (the transition kernel satisfies $Q(x, \{x\}) > 0 \; \forall x$), we do not think it would be an advantage to talk about the convergence of Cesàro means. On the contrary, we believe delving into Markov chain theory and discussing mean-squared convergence of Cesàro means might deter possible readers. Since we are

not reinventing MCMC theory, we leave it to inclined readers to consider the cited references.

Again, the important part is that readers understand that the samples that we generate behave such that they represent the distribution $P_{\theta|y}$. One advantage of focusing on the convergence of distributions $\mathcal{L}(\theta_k)$ to $P_{\theta|y}$ is that it helps us make a case for parallel tempering. Indeed, the main result of the cited reference is that this convergence can be accelerated by parallel tempering.

3) The authors will want to use a complicated Π and this means, typically, that it will be simple to sample from $\Pi|\psi$ but very hard to evaluate $p(\Pi|\psi)$ [...] I am skeptical of whether the authors are able to evaluate $p(\Pi|\psi)$ [we replaced all occurence of ϕ by ψ]

One of our proposals indeed consists in resampling the entire process from the prior, as suggested by the referee.

Additionally, we sometimes redraw only subintervals of the Wiener process that drives our earthquake generating process (see Equation (4) on page 7). By definition, this Wiener process is independent of the remaining components of the prior, so this can be done by a standard Brownian bridge construction (see lines 11ff on page 8).

Furthermore, we sometimes use local proposals to redraw the drift and volatility values of our earthquake generating process as well as the times at which those values change. Since these values together with their switch points constitute a marked Poisson process, the required Metropolis-Hastings ratio calculations can be found in the cited reference on "Reversible jump MCMC" (see line 10 on page 8 and issue 7 below).

4) "realistic earthquake models are much more complicated than marked Poisson processes"

While its drift and volatility coefficients are a marked Poisson process, our earthquake generating process itself is *not* a marked Poisson process. In any case, we agree that the model is not perfectly realistic. Nonetheless, our model is an improvement over the state-of-the-art Brownian passage time model, which is based on constant drift and volatility.

5) "it is not clear in the manuscript whether this is a normal or geometric Brownian motion"

Mathematically speaking, our earthquake generating process is an Itô diffusion (when conditioned on the drift and volatility coefficients). As such, it is driven by a standard Brownian motion. This is clear from Equation (4). In general, we don't think "Brownian motion" could seriously be used without qualifier to refer to a "geometric Brownian motion".

6) In their MCMC scheme, $prop(\theta, \phi)$ is a random walk, although I caution the authors to make sure that they include the Jacobian term if using a transformation

Most of our proposals do not require a Jacobian term. For example, for simple production parameters, we randomly use either a proposal from the prior or a local random walk type proposal. Whenever the latter proposal proposes values outside of the bounded interval, the proposal is simply rejected. No log-transformations etc. are used. (As a sidenote, the fact that sometimes proposals are outside of the bounded prior intervals is one way to check $Q(x, \{x\}) > 0$, which we claimed above)

7) I do not understand the Reversible Jump part at all. $\Pi|\psi$ is a point process; the fact that the number of components varies is irrelevant.

We agree that the fact that the number of components of the marked point process varies is irrelevant from a theoretical point of view. However, from a practical point of view, the corresponding calculations are not completely trivial, in particular concerning the careful treatment of Jacobians in the computation of MH ratios when draws are not from the prior distribution. The purpose of the cited work on "Reversible jump MCMC" is simply to present these calculations in theory and practice.

8) There is also some confusion in sec 3.2 which I will come back to below[...] The description of the proposal in sec 3.2 is confused.

We are not sure which part is confusing. The referee states he believes "some issues have been resolved above, by always sampling $\Pi|\psi$ from the prior" and we hope this is the case since we indeed mostly use simple proposals from the priors and local random walk type proposals (see section 3.2 of the submitted manuscript). The only exception of non-trivial proposals are those for the marked Poisson process of drift and volatility values, which is why we added the Reversible Jump MCMC reference. In general, we tried to keep the description in the manuscript easy to understand and read and attached the MATLAB code of our implementation as reference for implementation details.

9) In terms of diagnostics, the authors will need to demonstrate that their very intricate MCMC does indeed have the correct target distribution, using [...] Cook [...] Simpson

We thank the referee for bringing these very interesting verification tools to our attention. We implemented the test by Talts et al. (2018), the results of which are shown in Figure 1. However, since it is a very expensive test, we could only generate 200 samples in the month until this revision deadline, and each was only run until a point where the GR diagonistic was around 1.5 (200 samples means 200 synthetic test cases like the one presented in the manuscript). We changed the sentence

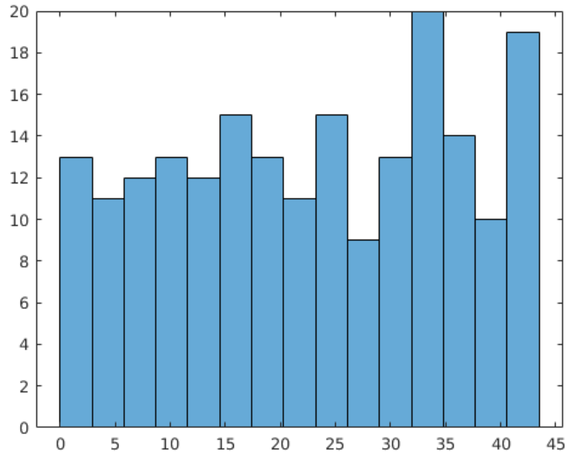

Figure 1: MCMC implementation verification following Talts et al. (2018): We generated 200 synthetic truths from the prior distribution and ran our MCMC algorithm for each. We used 10% burn-in and thinned each chain by factor of 250 to obtain quasi-independence. This left us with 42 posterior scenarios for each of the 200 runs of the MCMC algorithm. The distribution of the rank of the truth among these 42 scenarios is plotted above (where the ranking is based on the accumulated displacement at -7ka, a time at which the average fault has reached approximately half of its final height). A χ^2-test on uniform distribution of these ranks returned the p-value 0.54 (which is considerably too large to reject the null hypothesis of uniform distribution).

"We verify global convergence with the diagnostic of Gelman and Rubin (1992)."

in the introduction to

"We verify correct implementation of our MCMC algorithm by the test of Talts et al. (2018) and global convergence with the diagnostic of Gelman and Rubin (1992)."

We prefer to not include Figure 1 in the manuscript. In the end, this is just one of many possible implementation tests and we believe that the best way to verify software is to make it open and free. For example, even with the implementation of the MCMC algorithm verified, the simulation of 36Cl concentrations and the sparse grid surrogate for attenuation factors, neither of which are much easier to implement than the MCMC algorithm itself, can only be verified by applying our algorithm to real case studies or by looking at its source code (or by a handful of sanity checks, such as case-wise comparison with alternative simulators, monotonocity/positivity checks, application to trivial earthquake scenarios with obvious solution properties, etc., which we did perform but didn't include in the manuscript either). We do prefer to keep the much cheaper GR diagnostic for simple convergence testing. To use 8 instead of 2 independent chains, users may simply add `settings.group_size=8` to the case study text file.