This work focuses on the ability of data assimilation in providing estimates of fault states (slip rate and shear stress) in the presence of friction parameter biases by using a particle filter. Simultaneous state-parameter estimation seems to be an interesting endeavour in this challenging problem. The manuscript is organized and well written. My recommendation is minor revision before acceptance.

Please, see below my comments:

Comment 1
The study uses particle filter as the data assimilation method to solve the problem. But it would be good to inform the reader which other data assimilation methods were previously used in studies related to earthquake modelling and if those were successful. If not, which were the main issues and why a particle filter would suit better in this problem compared to those. This would mainly situate the reader on the importance of your choice on the method for this study.

Comment 2
Equations 4 and 5: What does j mean?

Comment 3
The authors use a Lorentz function instead of a Gaussian to prevent filter degeneracy. In addition, a SIR step is used to further avoid this issue. Were these enough to avoid filter degeneracy or the system still presents the problem?
The authors mention that real earthquakes are far from being periodic but, as they have considered a 0D model, their system generates periodic cycles. I wonder how far from a real state this 0D model is and why this study has not used a 1D model, in which at least a minimum spatial dimension would be considered. It will be interesting to address in the manuscript why the 0D model was chosen in this case.

Comment 5
I suggest different colors for the trajectories in the phase diagram in Figure 2, as it is hard to distinguish between them.

Comment 6
4 time steps in this model correspond to which portion of a seismic event? Please, describe it in the manuscript to situate the reader on the frequency of the assimilation steps, as earthquake cycles may not be a subject well understood by many.

Comment 7
Have you tested the impact of the use of less particles in the filter? If so, it would be good to share these results as well.

Comment 8
How do the orders of magnitude of the observation errors compare to the states? Are those the typical magnitude of the real measurement errors?

Comment 9
It would be good to explain in the label of Figure 5 what each of the lines in the pdf represent.

Comment 10
The authors mention that by using different assimilation settings, it is possible to inflate the ensemble. But it seems that the ensemble spread is not exactly the problem, as the state-parameter estimation presents much better results than the other tests which have an improved ensemble spread. Can the authors explain more clearly the effects of the spread on this specific seismic system?

Comment 11
Figure 8b): What happens after nearly every 250 time steps, in which the periodic behaviour is lost by a double peak? What is the influence of the double resampling in these patterns?

Comment 12
Still on Figure 8, it seems that Fig 8a) presents better results than Fig 8c), which makes me wonder if the double resampling is really helping the system...

Comment 13
The results for the state-parameter estimation are indeed promising and I congratulate the authors for this, but I would expect a comparison of these results with any other study (if they exist) using 1D models with or without data assimilation. The manuscript lacks information on other results found by studies which used other data assimilation methods and/or models with different dimensions.