Piana Agostinetti and Giulia Sgattoni present a study that focuses on partitioning / clustering double difference seismic travel time measurements using a transdimensional Bayesian-based method that they have devised. The authors apply their new method to a sample dataset which comprises an earthquake cluster near Katla volcano in Iceland. Given the null hypothesis of all 1141 events within the earthquake cluster being co-located beneath invariant subsurface elastic field, the authors test how their new method can quantitatively partition the data into different groups and explain the waveform similarities within groups as well as the seismicity rate in the region.

I find this manuscript interesting and think that it can potentially be an important contribution to more objectively selecting key parameters (e.g., the number and length of the time-windows) in double difference data. This can eventually enhance the robustness of subsequent seismological analysis such as seismic tomography or earthquake relocation. I have a few points worth further clarification before the manuscript is suitable for publication.

*Structure of the manuscript:*
The Data, Method, and Result sections are juxtaposed in the current form of this manuscript. This may be a good way to summarize / discuss with someone who is already familiar with this work but personally, it was difficult to follow. I think this manuscript would be much clearer if authors delineated the main contents (before the Discussion section) more strictly so that each section can be fully understood on its own. Furthermore, it would be great to include a flow chart that summarizes and describes the overall methodology and the corresponding sub-steps (as well as descriptions of the priors).

*Application to a complex dataset before validating the method:*
For a methodology-oriented research paper, I think it is logical to first test the new method on a very simplistic case (e.g., synthetic dataset such as Fig. 1) prior to its application to more complicated datasets. As the authors state in the introduction, the selection of the key parameters in double difference data is nonunique and largely qualitative. Therefore, even if the authors find some plausible posterior distributions of the hyperparameters with their new Bayesian approach, a minimum level of manual supervision is expected to draw a final set of the optimal parameters. That said, this final step can be very difficult to assess when the dataset that is being analyzed itself is complex. The dataset that the authors used in the manuscript is a rather complex one, at least for validation purposes. As the authors pointed out in the main text, their dataset includes some complexities: 1) changes in station network configuration that resulted in loss of data and degraded the quality of earthquake locations as well as origin times; 2) other temporal changes in waveforms that were identified through previous work. Why not start from a simple synthetic dataset for more straightforward assessment to understand whether the number and time-length of their analysis windows accurately capture elastic changes of the subsurface? It sounds like the authors already have tested their method with some synthetics (e.g., L251-253) so why not share them more extensively in the manuscript?

*Sensitivity of the analysis on data collected near Katla volcano:*

-How significant is the number of changepoints inferred from the analysis? For example, if the max. number of changepoints is fixed to a certain value, would the prediction still be able to explain the data equally-well (the largest probable changepoint, #3 seems pretty good in terms of clustering the data into ones that have stable CCs vs. ones that have variable CCs while the rest of changepoints don't seem to correlate well with the CCs)? In addition, the evolution of the number of changepoints in Fig. 5 generally shows that there are mainly 3-5 clusters / families of models which are stable across the model ensemble. What are the differences between those models? Do all of them show similar behavior shown in Fig. 7-8?
-Another related note here is how significant are those less-probable changepoints (other than changepoint #3); 1) would they correlate with other variables that are not currently discussed, e.g., variances of the CCs within the data delineated with gray lines in Fig. 7 or changes in seismicity rate in the region (e.g., in addition to what you state L322-323)? 2) if you were to compare the resulting clusters to any other widely used clustering algorithms (perhaps with an optimization-based approach that would be much faster to converge) would you expect similar segmentation of the data?
-It appears to me that the time-occurrence of the resulting changepoints is only discussed based on the timings of the event cluster (e.g., L190-193). Could you comment on any observable spatial patterns (e.g., Fig. 2b)?
-To test the significance of your optimal parameters, can you evaluate a "null hypothesis" model in which you randomly assign those two parameters within the distribution of the actual data?

Despite my comments / questions and reservations, I am confident that the new method presented in this manuscript will be a useful tool by providing a complementary solution for choosing the hyperparameters associated with double difference data. This would allow the community to move beyond more common, largely manual, hard-partitioning approaches. I hope the authors find my comments constructive and helpful as they make revisions.