The author presents a Monte Carlo method to constrain cluster merger dynamics parameters for "dissociative" merging clusters, in which the cluster gas has been displaced from the galaxies. The intent is to provide a method that can be used to constrain the dynamical state of such mergers without resorting to iterative full N-body hydrodynamical simulations, which are computationally expensive. Furthermore, additional priors may be added to the initial results from the model after the fact to more tightly constrain the parameters. The model is applied to the Bullet and Musket Ball clusters.

The paper is well thought out, and many of the potential concerns with the presented method are addressed. Although the final model parameter constraints are generally not very tight, even when the input observed parameters are reasonably well-known (mostly due to the uncertainty in the angle between the merger axis and the plane of the sky, as noted by the author), the method is certainly instructive, and has its place in assessing and identifying interesting candidates for full numerical simulations. A few questions, comments, and suggestions are given below, which the author might address, aside from which I am happy to recommend this paper for publication in ApJ.

In the abstract (and elsewhere), results from this method are said to be in agreement with N-body simulations to within 10%. While the phrasing given is technically accurate, given the context it is easily misinterpreted as meaning that, given a set of observables, the inferred parameters of the merger agree with those from numerical simulations matched to a particular observation at any time t during the merger (at least that was what I came away with from my initial reading of the abstract). The test that was actually done was to take known parameters from the merger (including d_3D(t_obs), which is not an observable), run this single case through the model (with no Monte Carlo analysis to consider errors), and show that the output v_3D(t_col) and TSC_o are within 10% of the values found in the simulation for the given merger state only (i.e., at t_obs). Although the way this is currently described in the paper is fair, the point could be clarified with some minor changes in language.

On a related note, claiming in the first place that the method is within 10% agreement with N-body simulations (plural) is over-stating things a bit. Once again, all that has been looked at is one snapshot from one simulation of one system. Ideally, one would compare results from the model with simulations for a variety of input parameters (mass ratios, merger velocities, concentration parameters, etc.) at a distribution of times (and a distribution of viewing angles) during the simulations to better characterize the accuracy of the model. I appreciate that running a series of N-body simulations for comparison is beyond the scope of this paper (and, in some sense, defeats the purpose of the model), but on the other hand it has yet to be shown that the model agrees with simulations to within 10% in general. If the author wishes to strengthen their case a bit further, I note that comparing results from the model to the one simulation considered (Springel & Farrar, 2007) over a spread of snapshot times is in principle easily done, as compared with running new simulations for comparison (assuming that merger parameters at other times in the simulation are available).

The abstract could also clarify that the model as it's presented is only applicable to dissociative mergers (since it assumes that the impact parameter b~0), although to be fair it could in principle be extended to include more general cases.

On more than one occasion, the author claims that for SIDM mergers the offset between the galaxies and the dark matter increases with time after core passage (i.e., t_coll) until the subcluster reaches the turn around radius. I suspect that this is not generally the case, and that there is a similar "slingshot" effect with the SIDM as observed with ram pressure stripped gas in merging galaxies and subclusters (i.e., at some time after t_coll and before reaching d_max the gravitational influence of the subcluster DM halo on the subcluster galaxies leads to a decrease in the separation). Has this been carefully checked, or simply assumed?

Third paragraph in the introduction: when discussing various methods for constrained v_3D for merging clusters the author might include the use of merger cold fronts via a comparison of the pressure at the stagnation point and in the free streaming regions (e.g. A3667, Vikhlinin et al. 2006, and several others), although the importance of projection and other systematic effects for this method are unclear.

The question of changing cluster masses during a merger (due to accretion, dynamical friction, tidal stripping, and, in the case of SIDM, momentum transfer) could, in principle, have a significant impact on the accuracy of results from the model. Appropriately, this topic is discussed in the paper, and the author provides reasonable arguments as to why these effects aren't important (at least for the case sigma=0). However, much of the argument is based on the agreement between the model predicted and N-body simulation parameters for the one case considered, while, as noted by the author, Farrar & Rosen find that dynamical friction leads to a 10% reduction in the inferred collision velocity for the subcluster. This apparent disagreement is a bit puzzling, and a possible explanation is not identified. Just as a check, can the author verify that the halo mass parameters given in section 2.3 are at t_obs, and not the initial mass profiles (which, again, are likely to be somewhat different)?

In the first paragraph of section 5.2, the author briefly discusses including a prior to allow for the possibility of a non-zero impact parameter, but indicates that it is not clear what prior to use. As a point of potential interest, Randall et al. (2002) (references by the author elsewhere) gives a probability distribution for the dimensionless spin parameter lambda and describe how to derive the impact parameter from lambda.

The values of M_200_2 differ slightly in the left and right "panels" of Table 3. It is not clear why this should be the case (typo?).

Section 5.2 typo: "in a *manner* similar to..."