Title: The return of the merging galaxy subclusters of El Gordo?
Manuscript ID: MN-14-3494-MJ
Authors: Ng, Karen (contact); Dawson, William; Wittman, David; Jee, Myungkook; Hughes, John; Menanateau, Felipe; Sifon, Cristobal

This paper is investigating the merger properties of the "El Gordo" merging cluster using a wide range of data set. Although the analysis sounds convincing, it would benefit of a better presentation so that the reader have a better understanding on the importance of various constraints. Once this is corrected, the paper can certainly be accepted.

Introduction:

> more references on cluster merger and in particular regarding the typical speed of a merger should be given.

We agree some typical speeds will help readers make better comparisons. We have added the typical speeds along with some citations on the 9th line of the introduction.

> Figure 1 is key for the understanding of the paper. I would suggest to make it larger (by rotating it by ~45-50 deg in order to have the elongation axis of the cluster horizontal)

This is an excellent suggestion and we agree with the referee about the importance of the figure. We did not rotate the figure because there is valuable information in the RA, Dec coordinates of the various components, so we made the figure two columns wide instead.

> showing possibly the galaxy luminosity contours.

We will have an overplotting issue if the galaxy luminosity is added. We did not present any scientific results based on the galaxy luminosity contours. The galaxy luminosity showing bimodality is in Menanateau 12 and we have put a note in the main text to refer the readers to the reference.

> Can the relic radio data be plotted on top instead of some schematic of it? (it looks like the size of the relic does not match the size given in Lindner et al 2014).

We emailed Robert Lindner for the relic contours but have not received any reply. The relic size depends on the wavelength and the contour of the radio relic image that one examines. We referred to Figure 5 and Figure 8 in Lindner et al. 2014 for estimating the extent of the radio relic and have double checked that the extent are consistent with the outer contours. The schematic is just to give the reader an idea of where the relic is compared to the rest of the cluster. We used the numerical locations given by L13 for all the calculation (See later discussion).

> The concept of time-since-pericenter is interesting, but it need to be clarified (the time at pericenter is only defined in section 3.3).

We completely agree with this suggestion. The definition has been added to second paragraph of P.3, the introduction section of the paper.

> I would suggest that you draw a diagram of the merger (may be at different time step) so that the reader can have a clearer idea of the geometry and evolution of the system (on the plot all the quantities used such as distances and velocities must be indicated). There you can also define what are the different merger scenarios that will thereafter discussed (the outgoing and the returning scenario). At the moment, the reader needs to go through the literature to understand what are the effective geometric assumptions used.

This suggestion really improves the presentation a lot. We have added the different merger scenario the introduction on P.3 at the end of the second paragraph. A new illustration has also been added as Fig. 2 to go along with the definition.

Section 2

> The presentation of the data is not fully quantitative. A table is given for the WL data, but the paper is also using the radio data relic, and a summary table would be great. This table should underline the key number from the radio data used in the analysis (position, Mach Number/velocity?, polarisation? ...)

We have tried to make the data as fully quantitative as possible and put the relevant quantities in a section where the calculation making use of the quantities are explained.

Only the polarization information of the NW relic has been used i.e. 1 best estimate and 1 uncertainty, is present in Section 3.3 where we explain how to calculate the polarization prior. The E relic data is not used as we explained in the paper that the E relic does not have polarization data available.

The two observed positions (RA, DEC) used for the NW and the E relic have been added to the end of section 3.4, along with the projected separation in the center-of-mass frame. Section 3.4 explains how the probabilities of different merger scenarios are calculated.

We did not use any Mach number in our calculation and we have explained why Mach numbers give unreliable estimates of the shock velocities in the center of mass frame in Section 4.1 paragraph 3.

> What about the velocity data? It is described in section 3.1.1 but should it not be moved to section 2?

Yes, we agree section 2 is a better location for describing how the velocity data was obtained. We have moved the descriptions of the relative velocity data from the first paragraph of section 3.1.1 to section 2.

Section 3

> I believe the vector D is representing the data, but this could be clearer, and it would be good to clarify which data is effectively used. Table 1 seems to only give part of the data used in the analysis.

Table 1 gives all the PDFs that we draw directly from as inputs to the simulation, i.e. $M_{200_c NW}, c_{NW}, M_{200_c SE}, c_{SE}, z_{NW}, z_{SE}, d$
.
We did not draw samples directly from from the radio relic data, so the radio relic data does not belong in Table 2 which specifies only properties that are used as the input PDF of the Monte Carlo simulation. The radio relic data are only used to compute Monte Carlo weights.

We have edited the descriptions in Table 1 to clarify that the table is only for input sampling PDFs and not for data used for computing weights or merging scenario.

The radio relic data that we do make use of include: 1) the polarization fraction and the uncertainty ($33\% \pm 1\%$) which is specified in section 3.3.1 where the physics of the polarization are explained 2) the observed location of the relics (RA, DEC and the separation from the center of mass) that we used for the calculation for the different merger scenarios have been added to the last sentence of section 3.4. This particular section explains the motivation and how the projected relic location is calculated.

> It would be helpful to describe a little bit more the MC simulation code. Does the simulation use a large number of particules? Or is it just using 2 "particles" with a NFW mass profile. It would be good to remind the reader of some of the key element in D13.

We agree that we should provide the details of the simulation but the original paper (Dawson 13) did not describe the setup more than having two NFW profiles for evaluating the gravitational force. After going through the code base, we have added The description of how the gravitational attraction was numerically evaluated at fixed grid points of each of the two analytic NFW halo profiles (10000 grid points each) in section 3, at the end of the first paragraph.

> Are the galaxies introduces as test particles in the simulations? It seems not, but would this be a way to better model and possibly constraints the merger? I computed a Delta_v_rad of 463 km/s (based on the 2 redshift: 0.8684 and 0.8713) instead of the number of 476km/s given in the text, can you explain why?

We thank the referee for double-checking our calculation. (1) The calculations that led to the results were done with the entire probability density functions (PDFs), i.e. carrying out this calculation for EACH of our 2 million realizations, then we take the biweight location as the estimate. In the referee 's calculation, only the best estimate (this approach neglects any correlation in the inputs) was used. Therefore, the discrepancies can arise from the different inputs. (2) We only report to 2 significant figures since some the inputs of the data did not have more than 2 significant figures. (3) The uncertainty that we give is much bigger than the discrepancy between 463 and 476 km/s.

> Similarly the d_{proj} I found is 0.744 instead of .74 (based on the RA, DEC given)

We also used the full PDFs for computing d_{proj} which is a more precise calculation. Other explanations from the previous answer also follows.

> It would be much better to define the output parameters using a diagram of the different merger scenarios.

We have added Figure 2 that outlines each merger scenario with the corresponding time-since-pericenter and the corresponding projected separation of the two subclusters.

> Would there be a correlation between beta and TSP?

No. The parameter β is not a property, nor an output of each Monte Carlo simulation. It is a choice of parametrization of our uncertainty, i.e. you cannot find one β value for each simulation so you cannot compute a correlation between β and TSP. The variable TSP is an output and has one value for each Monte Carlo simulation.

Section 4

> It would be good to move some of the key likelihood plot from the annex to the main text. Indeed, the start of section 4, starts with results that are basically described in the appendix, which is not ideal.

We have moved table 2 to be right before the result section and provided an additional sentence to guide the reader to find the table in the result section. The table contains the key summaries of the results. Figure 6 also contains the marginalized PDF of one of the key variables, TSP vs the relative 3D velocities at pericenter. The plots in the appendix is only for visualizing possible correlation but otherwise contain the same information as table 2.

Section 5

> For the comparison with other analysis, I would suggest that the author summarizes the comparison in two tables. One table comparing results with other El Gordo modelling.

We agree with the referee and we originally attempted summarizing the results in table form. However, the hydrodynamical simulations do not directly give estimates for most of the output parameters that we have. If we try to compute a table, a lot of the entries from the hydrodynamical simulations will be rough estimates that are estimated from their plots and would require explanation of how they might have under sampled the number of snapshots so their time estimates have larger uncertainties.

> And one table comparing the properties of the different merger described with the same modelling principle, in particular comparing El Gordo to the bullet cluster and the musket cluster.

In section 5.2, we have summarized the main similaries and differences due to TSP that can contribute to the morphological features. We feel like a numerical comparison from a table is less compelling than the description of the physics at the different merger stages as in section 5.2. Furthermore, the Musketball Cluster around an order of magnitude less massive and is much more further progressed than both El Gordo and the Bullet Cluster. The Musketball also does not possess neither the cool core nor radio relics. Therefore, we found the Musketball different enough that we did not include a direct comparison of the Musketball Cluster in the paper. Interested readers can find the full numerical results of Musketball and Bullet Cluster in Dawson 13 for a comparison themselves.