CluMMP: Development and Evaluation of Comparative Visualizations of Galaxy Cluster Mergers

Lucas Brito*   Ian Dell'Antonio†

Brown University

ABSTRACT

We present and evaluate *CluMMP* (Cluster Merger Matching Program), a comparative visualization tool for matching observed galaxy cluster mergers to corresponding simulations. Due to the complex dynamics and extremely slow evolutions of galaxy cluster mergers, observationally determining the elapsed time of these events is a difficult and uncertain task. This is typically done by matching a merger to a simulation which is similar in morphology and separation to a given observation. Our tool, CluMMP, aims to expedite this matching process by employing a likely-match simulation algorithm and implementing three visualization methods for studying candidate simulations. Our approach performs a one-dimensional nearest neighbors search on precomputed cluster centroid separations, aligns simulation and observation cluster centroids, and displays simulation images in a browser-based UI using three visualizations: "Side-by-side," "Flicker," and "Difference." Our quantitative evaluation of this tool yielded suggests that the "Difference" visualization is the most effective whereas "Flicker" is least effective, although qualitative evidence suggests that the effectiveness of each visualization method varies with the similarity between candidate simulations.

Keywords: Comparative visualization, application, methodology, evaluation.

1 INTRODUCTION

Galaxy cluster mergers are collisions between two gravitationally bound groups of galaxies. Such collisions represent the last stage in the formation of the structure of our universe and thus are important objects in cosmology. One property of interest of a merger is its timescale, i.e., how long the event has been ongoing; however, cluster mergers are slowly evolving events (in the order of 10^9 years) and thus obtaining this information to good approximation from observation alone is difficult.[5] This is typically done by comparing observed mergers with simulations from which we can extract such information[9]; however, current approaches to characterizing timelines of cluster mergers from simulated collisions are largely non-systematic or heuristic.

This extended abstract presents *CluMMP* (Cluster Merger Matching Program), a tool developed to address the demand for an accessible cluster merger observation-simulation matching workflow. The tool seeks to eliminate several degrees of freedom encountered in the matching process by algorithmically producing likely-match candidates, aligning cluster centroids, and providing users three visualization methods; the design and data processing pipeline of the tool is elaborated on in §3. Our tool also expedites the data acquisition process by providing a web-based UI which allows users to load, view, and interact with remote FITS data.

*e-mail: lucas_brito@brown.edu

†e-mail: ian_dell'antonio@brown.edu

Figure 1: Clusters 1E2215 and 1E2216 experiencing a merger. Composite X-Ray and optical bands.[1]

In addition to presenting CluMMP, this extended abstract presents and discusses the results of an evaluation of the three visualization methods implemented by CluMMP; see §4. In short, we find quantitative evidence suggesting that a "Flicker" visualization yields the shortest task completion time, and qualitative evidence suggesting that the utility of each visualization varies with similarity between candidate simulations. We discuss these results in §5, and open questions and future directions to take the tool are considered in §6.

2 BACKGROUND AND RELATED WORK

There are no existing parameter space exploration tools for cluster mergers, although there are similar tools for analysis of galaxy mergers.[2] In particular, there are no browser-based tools visualization tools designed to facilitate matching observed astronomical data to simulated data products. Furthermore, little work has been done in applying and evaluating the effectiveness of comparative visualization techniques in the context of simulated versus observed images of hydrodynamic phenomena. State of- the art algorithmic image comparison is not sufficiently generalizable to 2D projections of 3D subjects, and one expects the performance of such approaches to decrease given the chaotic nature of cluster dynamics[8][4].

There are existing discussions of comparative visualization techniques, philosophies, and approaches[5][4] and high-level implementations of comparison-focused pipelines[3][7], but little work has been done on evaluating the relative effectiveness of each of these methods, in particular as applied to a simulation-observation matching procedure.

Our tool relies on ZuHone et al.'s Galaxy Cluster Merger Catalog (GCMC) for simulation data; in particular, we access data from the yt Hub (`girder.hub.yt`) via the Girder Python API, and acknowledge that the functionality of CluMMP is reliant on the upkeep and extension of these catalogs and tools.[10]

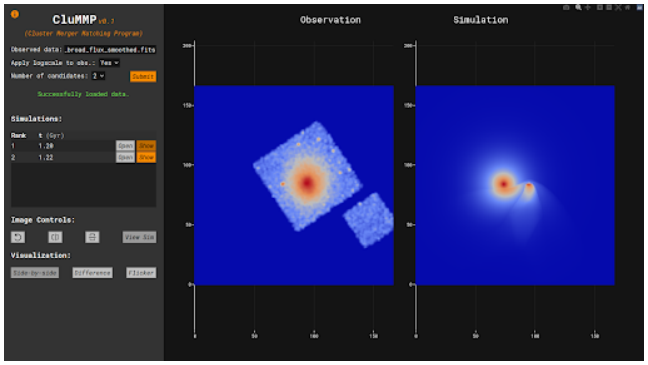

Figure 2: CluMMP with "Side-by-side" visualization.

3 METHODOLOGY

3.1 Data

GCMC simulations were pre-processed in order to obtain cluster X-Ray emissivity centroid separation for usage in the nearest-neighbors candidate suggestion algorithm, and results were stored in a SQL database. Upon receiving a request for n candidates, CluMMP performs a one-dimensional nearest-neighbors search on centroid separation obtained from observation data. CluMMP then uses the Girder Python API (`girder.readthedocs.io`) to download GCMC data products, applies the alignment procedure to simulation and observation data (see §3.2), generates the "Difference" visualization for each simulation, optionally applies a pixel-value stretch to the observation data, then submits a response to the client with appropriate data. Observation data are obtained through a user-provided path to any FITS file in the Oscar computing cluster, which has been mounted to the CluMMP backend.

3.2 Image Alignment for Comparison

Aiming to eliminate superfluous degrees of freedom which may inhibit the observation-simulation matching worfklow, CluMMP eliminates translational and rotational variation among simulations by aligning all pairs of cluster centroids across observation and simulation data. We define the centroid of a cluster as the peak X-Ray emissivity point in the cluster. This point is estimated by computing the level sets of the surface represented by the image's pixel data, obtaining the highest-value level set containing two disconnected paths, and finding the centroid of each path.[1]

We then determine the largest image in area I_{ref} and use this as reference for alignment and resizing in order to prevent loss of information from downsizing. A "source triangle" $\triangle_{\text{src}}$ is constructed by using the cluster centroids of I_{ref} as base vertices $\mathbf{x}_1$ and $\mathbf{x}_2$ and generating a third vertex $\mathbf{x}_3 = \mathbf{F}(\mathbf{x}_1, \mathbf{x}_2)$ from the base vertices such that the triangle is isosceles. A "destination triangle" $\triangle_{\text{dest}}$ is then computed by positioning two base vertices $\mathbf{x}'_1$ and $\mathbf{x}'_2$ along the horizontal midline of I_{ref} such that their midpoint is the center of I_{ref}, then again finding an apex vertex $\mathbf{x}'_3 = \mathbf{F}(\mathbf{x}'_1, \mathbf{x}'_2)$. Then for every simulation/observation image I_i we perform the affine transformation defined by $\triangle_{\text{src}_i} \to \triangle_{\text{dest}}$, and crop I_i on the scale $6 \cdot |\mathbf{x}'_2 - \mathbf{x}'_1|$ at $1:1$ aspect ratio to emphasize local morphology. Users are provided with transform controls to explore the additional degree of freedom of image orientation.

3.3 Visualization Methodologies

We three visualization methods for displaying and comparing a selected simulation against an observation. The "Side-by-side" visualization displays the observed merger next to a selected simulated merger; see Figure 2. The "Difference" visualization displays a heatmap of $|I_{\text{sim}}(\mathbf{x}) - I_{\text{obs}}(\mathbf{x})|$ where $I_{\text{sim(obs)}}(\mathbf{x})$ is the value of the pixel of the simulation (observation) at position $\mathbf{x}$. Lastly, the flicker observation overlays the aligned images and allows the user to toggle between the observation and simulation images.

[1] Defined by $(1/N)\sum_i^N \mathbf{x}_i$ where N is the number of points in that path and $\mathbf{x}_i = (x_i, y_i)$ are the coordinates of that point.

4 RESULTS

4.1 Web Application

A demo distribution of CluMMP can be accessed at `lucasbrito.site/clummp`; note that the demo version of the software is a static webpage and has limited functionality.

4.2 Evaluation

The visualizations implemented by CluMMP were evaluated using the standard speed/accuracy metric. In order to perform the evaluation, five simulations with similar cluster separations were selected from the GCMC, and from each a mock observation was generated using GalSim's photon shooting method.[6] A professional astronomer was given the task of, given a mock observation and the five selected simulations, find the simulation from which the observation was generated. Time was measured in seconds taken to confidently deduce a match-simulation, and accuracy was measured as a Boolean representing whether the chosen simulation was correct. The user was instructed to perform this task using four different workflows: using each of the three visualizations in isolation, and using all three. For the latter workflow, time spent using each of the visualizations was measured; results are displayed in Table 1. These workflows were not compared against the existing workflow due to conflation with efficiency obtained from the simulation suggestion algorithm.

In addition, after the completion of the quantitative evaluation, the user was asked to provide an oral evaluation of the tool. The user stated that the tool is "extremely useful" and "really powerful." The user also postulated that as similarity between simulations increases, "Difference" and "Flicker" visualizations would become more effective, but that for dissimilar simulations "Side-by-side" is the most effective. Lastly, the user suggested that a signed difference visualization would display information useful for the matching procedure.

5 DISCUSSION

The evaluation results displayed in Table 1 suggest that all visualizations provide sufficient information to accurately deduce a correct matching simulation, and that the least effective visualization method is "Flicker" whereas the most effective method is "Difference." For the "All" workflow, wherein the user was allowed to freely use all visualizations, the user spent the majority of the time using the "Flicker" visualization; we postulate that these results are due to a steeper learning curve for the "Flicker" visualization given its interactive component.

6 CONCLUSION

This extended abstract presented CluMMP, a browser-based utility for matching observed and simulated galaxy cluster mergers. We in addition presented our evalution of this tool, which demonstrated qualitatively that the tool is of significant utility to astronomers, and that in particular a "Difference" visualization is most effective in aiding the matching procedure.

Further work on CluMMP entails updating the tool as the GCMC expands, implementing a signed-difference visualization, filtering point sources from the centroid-identification algorithm, and distributing the full version of the application with implemented support for observation data from a user-defined SMB-compliant directory and user-uploaded observation data. Further evaluation work

Table 1: Evaluation results.

Vis	Time (s)	Correct sim.	Time spent using
Side-by-side	49	True	7.8%
Difference	14	True	3.1%
Flicker	66	True	89.1%
All	56	True	100%

entails determining the accuracy of the simulation-suggestion algorithm, performing further user testing with simulation data with varying degrees of similarity, and evaluating a signed difference visualization against the existing "Difference" visualization.

ACKNOWLEDGEMENTS

The authors wish to thank David Laidlaw for contributing crucial ideas to the visualization methodology employed by CluMMP, and Robert Scheidegger for general software engineering advice. The authors also acknowledge the utility provided by the packages Django, Matplotlib, Numpy, OpenCV, Plotly, Girder, Astropy and the front-end framework React.

REFERENCES

[1] *Cluster Merger: Galaxy Clusters Caught in a First Kiss*. NASA, Jun 2019.

[2] J. E. Barnes and J. E. Hibbard. IDENTIKIT 1: A MODELING TOOL FOR INTERACTING DISK GALAXIES. *The Astronomical Journal*, 137(2):3071–3090, jan 2009.

[3] L. Bavoil, S. Callahan, P. Crossno, J. Freire, C. Scheidegger, C. Silva, and H. Vo. Vistrails: enabling interactive multiple-view visualizations. In *VIS 05. IEEE Visualization, 2005.*, pages 135–142, 2005.

[4] Z.-J. Ding, Y. Zhang, A.-Q. Yang, and Dai-Li. Image matching of gaussian blurred image based on sift algorithm. In *2012 International Conference on Wavelet Active Media Technology and Information Processing (ICWAMTIP)*, pages 121–124, 2012.

[5] N. R. Golovich. *Observations and Modeling of Merging Galaxy Clusters*. PhD thesis, University of California, Davis, United States, Jan. 2017.

[6] B. Rowe, M. Jarvis, R. Mandelbaum, G. M. Bernstein, J. Bosch, M. Simet, J. E. Meyers, T. Kacprzak, R. Nakajima, J. Zuntz, H. Miyatake, J. P. Dietrich, R. Armstrong, P. Melchior, and M. S. S. Gill. Galsim: The modular galaxy image simulation toolkit, 2015.

[7] C. Schmid and H. Hinterberger. Comparative multivariate visualization across conceptually different graphic displays. In *Seventh International Working Conference on Scientific and Statistical Database Management*, pages 42–51, 1994.

[8] G. Yu and J.-M. Morel. A fully affine invariant image comparison method. In *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 1597–1600, 2009.

[9] J. A. ZuHone. A parameter space exploration of galaxy cluster mergers. i. gas mixing and the generation of cluster entropy. *The Astrophysical Journal*, 728(1):54, Jan 2011.

[10] J. A. ZuHone, K. Kowalik, E. Ohman, E. Lau, and D. Nagai. The galaxy cluster merger catalog: An online repository of mock observations from simulated galaxy cluster mergers, 2017.