MNRAS 000, 1–3 (2020) Preprint 25 February 2020 Compiled using MNRAS LaTeX style file v3.0

JAS1101 Final Project Exploratory Data Analysis Report

Nathaniel Starkman[1]★ & Qing Liu[1] & Vivian Ngo[2]
[1]Department of Astronomy and Astrophysics, University of Toronto, 50 St. George Street, Toronto, Ontario, M5S 3H4, Canada
[2]Department of Statistics, University of Toronto, Toronto, Ontario, M5S 3H4, Canada

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT

Key words: Globular Cluster – Black Hole

1 INTRODUCTION

Intermediate mass black holes (IMBHs) should help us to understand the evolutionary connection between stellar mass and super massive black holes. However, there is still no clear observational evidence for their existence. The strongest ig

there is still much to learn about their formation and existence. Because of their dim signatures and their low X-ray to radio flux ratio, they do not fulfill the requirements to accrete black holes.

IMBHs are known to strongly affect the spatial distribution of stars in a GC since massive stars sink into the centre and stars that sink closer to the centre are then scattered by the black hole, heating up the core of the GC.

To date, the favored method of constraining intermediate mass black holes (IMBHs) in globular clusters (GCs) is on a case-by-case basis, necessitating a detailed and precise analysis of each GC. Instead, we will attempt to consolidate the full (known) GC catalogue and perform a population-level inference on the favorability of IMBH-inclusive GC models over IMBH-free models. The question we are asking is: for globular clusters as a population, whether the presence of IMBH in the center is statistically significant. And furthermore, from an integral perspective, whether a black hole mass - velocity dispersion ($M_{BH} - \sigma$) correlation holds for IMBH. The answer would shed light on the role that IMBHs plays in globular clusters, and help us better understand the co-evolution of black holes with astronomical systems on different scales.

2 DATA

On the executive summary level, the data consists of 6D phase-space (3 positions, 3 velocities) and photometric (2 colors) measurements from the *Gaia* catalogue, of the population of ~ 150 globular clusters (GCs) in the Milky Way (MW). We use a previously derived catalogue from Vasiliev (2019) to obtain the GCs and some analysis of their properties. From this catalogue we further process the data such that every globular cluster is rescaled in both position and proper motion space by their characteristic radius / proper motion. In rescaled form the GCs may be stacked and analyzed as a population, rather than on an individual basis.

In the following we provide a more detailed description of the data and our efforts thus far to manipulate the data into a form amenable to stacked analysis. We start with a description of the catalogue from Vasiliev (2019).

The list of GCs and their characteristics is derived from two sources: Harris (2010) provides the sky coordinates and distances, except for a couple of corrections; Hilker et al. (2019) provides the line-of-sight velocity, error estimate, and central velocity dispersion for the GCs. For each GC in the combined table the Gaia catalogue is queried for all stars in the cluster. We here describe the query from Vasiliev (2019). Full details can be found in the source paper.

- Select all stars with the full 6D astrometric data and that are within an angular distance $R_{\max}$ (larger than the characteristic radius so as to have a sampling of the background).
- Remove nearby stars (from the parallax).
- Remove stars with large parallax errors.
- Remove stars with large `astrometric_excess_noise` or with a large renormalized unit weight error (Lindegren et al. 2018).
- Remove stars with significant colour excess: `phot_bp_rp_excess_factor`

After obtaining the raw data from the *Gaia* catalogue

Our data will be sourced primarily from the Gaia catalogue, which is publicly available. As needed, we will cross-match with precision photometry datasets, such as Pan-STARRS and 2Mass, to which we also have access. These datasets offer improvements upon the Gaia photometry and augment the Gaia kinematic information. Prior studies, such as Vasiliev 2018, have identified 150 GCs and provided details on how to reproduce a GC catalogue.

3 EXPLORATORY ANALYSIS

In our exploratory analysis we have concentrated on the Globular Cluster 47 Tuc. Our pipeline, in development should work for all globular clusters as well as for a stacked set of globular clusters

Plots:

★ E-mail: n.starkman[at]mail.utoronto.ca

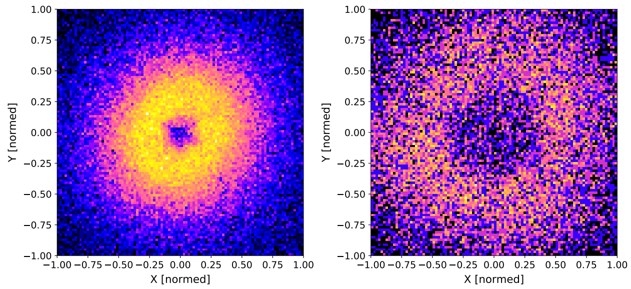

Figure 1. Caption

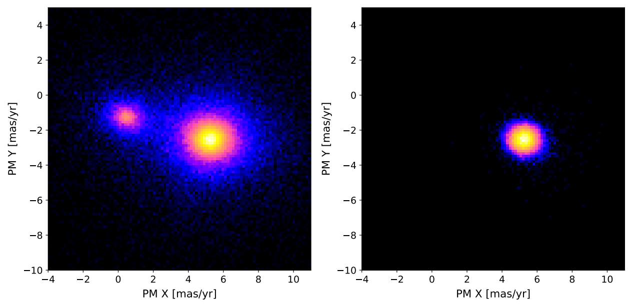

Figure 2. Caption

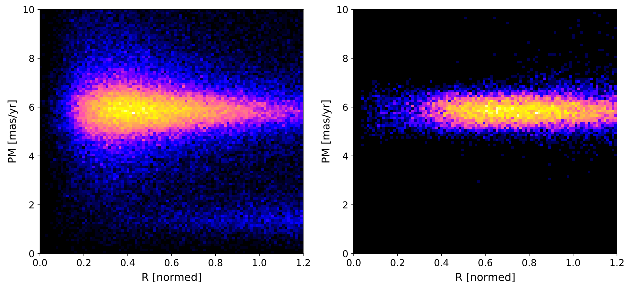

Figure 3. Caption

(i) XY of 47 Tuc
(ii) radial density profile of 47 Tuc, including the scale radius
(iii) The proper motion 2D distribution, without the ‘member-prob‘ to show cleaning problem.
(iv) PM radial profile
(v) Dynesty toy model
(vi) Plummer sphere + BH model
(vii) King Model

3.1

Normally the next section describes the techniques the authors used. It is frequently split into subsections, such as Section **??** below.

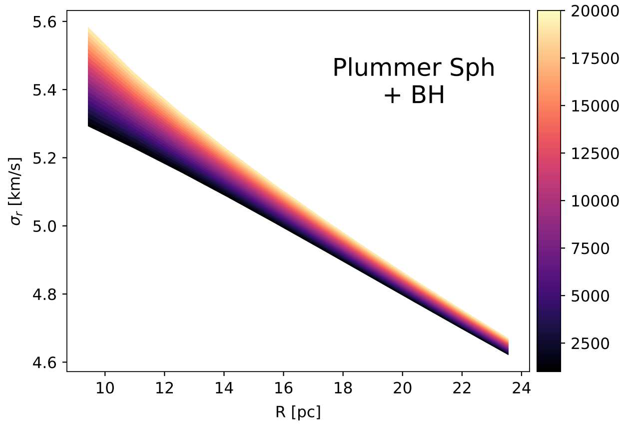

Figure 4. Caption

4 LIMITATIONS

In this section we will describe the issues that we have run into in our data analysis thus far.

5 FUTURE WORK

For the remainder of this project, we intend to investigate the other GCs in the Milky Way Galaxy. So far we have focused on 47 Tucanae to conduct in-depth analysis but we have developed a pipeline that should be able to perform the same analysis for other GCs as well.

Moreover, we want to determine if IMBH’s are significant factors in GC models for GCs in the Milky Way Galaxy. To do this,

C catalogue and perform a population-level inference on the favorability of IMBH-inclusive GC models over IMBH-free models.

The statistical analyses that we plan to apply are model fitting, hypothesis testing, survival analysis, and simulations. One of our goals is to quantitatively describe properties of globular clusters by fitting models to determine differences due to the presence of IMBH. We also plan to conduct hypothesis tests to determine the significance of correlations. Moreover, sometimes only upper limits of M_{BH} can be derived, as a case of censored data. To deal with this, we will leverage survival analysis in our research. Finally, we will also use Markov Chains and Monte Carlo techniques to perform simulations.

ACKNOWLEDGEMENTS

The Acknowledgements section is not numbered. Here you can thank helpful colleagues, acknowledge funding agencies, telescopes and facilities used etc. Try to keep it short.

REFERENCES

Harris W. E., 2010, arXiv e-prints, p. arXiv:1012.3224
Hilker M., Baumgardt H., Sollima A., Bellini A., 2019, arXiv e-prints, p. arXiv:1908.02778
Lindegren L., et al., 2018, A&A, 616, A2
Vasiliev E., 2019, MNRAS, 484, 2832

APPENDIX A: SOME EXTRA MATERIAL

Below are some results of our exploratory data analysis.

This paper has been typeset from a TEX/LATEX file prepared by the author.