Dynamics of External Galaxies: Stellar Kinematics, Black Holes, and Machine Learning Modelling

David A. Simon

Wolfson College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Michaelmas 2025

Acknowledgements

As I reflect on the last four years, I cannot help but feel an immense sense of gratitude for all of the kindness and support given to me by so many people. Foremost among these is my adviser, Michele Cappellari. Working with him has been an immense privilege and pleasure. Michele, in addition to being supremely knowledgeable and insightful, is also extremely kind, willing to invest enormous amounts of time and effort into his students. I have learned an enormous amount over these last years—not only about astrophysics, but also about how to think critically, communicate clearly, and approach problems with both curiosity and rigour. His patience and guidance have shaped my development as a researcher more than I can express.

I also owe an enormous debt of gratitude to the Oxford Astrophysics Department and my fellow DPhil students. The department has been one of the warmest, most welcoming places that I have ever had the pleasure of being a part of. The collaborative spirit and sheer enthusiasm for discovery have made Oxford an inspiring environment in which to learn and grow. A big thank you to my DPhil cohort, Phil, Fergus, Rohan, James, Matthew, and Haochuan, who added many moments of joy along the DPhil journey. I'd also like to thank Eric Liang, Julien Devriendt, Bence Kocsis, Jonathan Patterson, Ashling Gordon, and Johanna Hartke. I can't imagine a better place to have spent these last four years.

I have also been very fortunate during my time to visit and study with professionals from outside of Oxford. I would like to give a special thank you to Thorsten Naab and his group, especially Antti Rantala, for hosting me at the Max Planck Institute for Astrophysics in Garching. My time there was both scientifically enriching and personally rewarding, and I am grateful for the opportunity to engage with such a vibrant and talented research community.

I would also like to give a very big thank you to Shude Mao and the astronomy students and support staff at Tsinghua University for their support during my visit there. I would especially like to thank Kai Zhu, Youquan Fu, Haitong Wang, Jiani Chu, Dandan Xu, Qiyue Qian, Jiyuan Zhang, Xiaohan Wang, Haozhe Jiang, Xikai Shan, Yunpeng Jin, and Xiang Liu, among many others. I was made to feel very welcome and at home during my time visiting Beijing for which I am very grateful.

Additionally, I would like to thank my examiners, Eugene Vasiliev and Niranjan Thatte. Their diligent reading of the thesis and deep and interesting questions have significantly improved the quality of this thesis.

Last but not least, I would like to thank my parents for their endless love and for the many sacrifices that they made to give me many opportunities in my life. Their support and belief in me have opened doors I never could have reached on my own.

Abstract

Spectroscopy has played an important role in many of the major discoveries in astronomy over the last century. Integral field spectroscopy has continued this trend. Originating in the late 1980s, it is now a fully mature technology and a staple component of modern observing facilities.

I showcase the power of integral field spectroscopy by using MUSE and OASIS data to measure the supermassive black hole mass in the early-type galaxy M87. M87 is the poster child for black hole studies, but presents a number of challenges in the reduction and analysis of the data. To overcome these issues, we pioneer several new techniques for handling the dynamical modelling, as well as determining key parameters such as the PSF and central stellar distribution.

From here, I turn my attention to some of the details of integral field kinematics and dynamical modelling. Making the most out of integral field data typically requires two steps. The first step is to measure the spatially resolved line-of-sight velocity distribution (LOSVD) using a spectral fitting code. This involves a number of choices, from the choice of stellar library to the parameterisation of the LOSVD. The second step is to perform dynamical modelling using these extracted kinematics.

I investigate the impact these choices make at two levels using a high-resolution N-body simulation. First, the choice of the parameterisation of the LOSVD is typically done using a truncated Gauss-Hermite series. Different choices of truncation, that is, different numbers of moments, can induce systematics in the recovered kinematics. At another level, template mismatch and the specific choice of stellar library can induce systematic offsets in the kinematics which impact dynamical models. I investigate each of these independently, and determine some of the best practices for fitting stellar kinematics.

Using these stellar kinematics, I perform JAM, axisymmetric Schwarzschild, and triaxial Schwarzschild modelling and describe the systematic differences between these methods and how well they are able to recover the true black hole mass.

The role of systematics in dynamical models is also important. Most dynamical modelling methods make a number of assumptions which don't perfectly hold for real galaxies. In the penultimate chapter, I discuss an interesting new direction, where some of these assumptions can be broken using machine learning models which replace traditional dynamical models. I conclude with a very optimistic picture of the future of this field.

Statement of Originality

I carried out the work in this thesis at the Department of Astrophysics, University of Oxford between October 2021 and January 2026 under the supervision of Professor Michele Cappellari. My research was funded by a UKRI/STFC studentship. I declare that no part of this thesis has been submitted in support of another degree, diploma or other qualification at the University of Oxford or any other university. Except where otherwise stated, the work in this thesis is all mine.

This research is based on observations collected at the European Southern Observatory under ESO program 0103.B-0581. This research is also based on observations obtained at the Canada-France-Hawaii Telescope (CFHT) which is operated by the National Research Council of Canada, the Institut National des Sciences de l'Univers of the Centre National de la Recherche Scientifique of France, and the University of Hawaii. This research is based on observations made with the NASA/ESA Hubble Space Telescope obtained from the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS 5–26555. These observations are associated with program GO-9401.

In this thesis I have made extensive use of SDSS photometry and MaNGA IFS observations. Funding for the Sloan Digital Sky Survey IV has been provided by the Alfred P. Sloan Foundation, the U.S. Department of Energy Office of Science, and the Participating Institutions. SDSS acknowledges support and resources from the Center for High-Performance Computing at the University of Utah. The SDSS web site is `www.sdss4.org`.

SDSS is managed by the Astrophysical Research Consortium for the Participating Institutions of the SDSS Collaboration including the Brazilian Participation Group, the Carnegie Institution for Science, Carnegie Mellon University, Center for Astrophysics — Harvard & Smithsonian (CfA), the Chilean Participation Group, the French Participation Group, Instituto

de Astrofísica de Canarias, The Johns Hopkins University, Kavli Institute for the Physics and Mathematics of the Universe (IPMU) / University of Tokyo, the Korean Participation Group, Lawrence Berkeley National Laboratory, Leibniz Institut für Astrophysik Potsdam (AIP), Max-Planck-Institut für Astronomie (MPIA Heidelberg), Max-Planck-Institut für Astrophysik (MPA Garching), Max-Planck-Institut für Extraterrestrische Physik (MPE), National Astronomical Observatories of China, New Mexico State University, New York University, University of Notre Dame, Observatório Nacional / MCTI, The Ohio State University, Pennsylvania State University, Shanghai Astronomical Observatory, United Kingdom Participation Group, Universidad Nacional Autónoma de México, University of Arizona, University of Colorado Boulder, University of Oxford, University of Portsmouth, University of Utah, University of Virginia, University of Washington, University of Wisconsin, Vanderbilt University, and Yale University.

This thesis has made use of the NASA/IPAC Extragalactic Database (NED), which is funded by the National Aeronautics and Space Administration and operated by the California Institute of Technology. This research has made use of NASA's Astrophysics Data System Bibliographic Services. This research made use of Montage. It is funded by the National Science Foundation under Grant Number ACI-1440620, and was previously funded by the National Aeronautics and Space Administration's Earth Science Technology Office, Computation Technologies Project, under Cooperative Agreement Number NCC5-626 between NASA and the California Institute of Technology.

I have had the great benefit during this thesis to have access to a number of computing facilities. I would like to thank Jonathan Patterson for his assitance with the Oxford astrophysics computing cluster Glamdring. I would also like to thank the Max Planck Computing and Data Facility (MPCDF) for access to computational resources. I would like to acknowledge Tsinghua University for the Tsinghua Astrophysics High Performance Computing platform, as well as Westlake University for the High-Performance Computing platform that provided computational and data storage resources that have contributed to the results in this thesis.

I wish to acknowledge CSC – IT Center for Science, Finland, for computational resources.

This research made use of Python (Python Software Foundation, 2024), Jupyter (Pérez & Granger, 2007; Kluyver et al., 2016), NumPy (Harris et al., 2020), SciPy (Virtanen et al., 2020), Matplotlib (Hunter, 2007), Astropy, a community-developed core Python package for astronomy (Astropy Collaboration et al., 2013, 2018), emcee (Foreman-Mackey et al., 2013), Schwimmbad (Price-Whelan & Foreman-Mackey, 2017), MgeFit (Cappellari, 2002), Vorbin Cappellari & Copin (2003), JamPy (Cappellari, 2008, 2020), pPXF (Cappellari & Emsellem, 2004; Cappellari, 2017, 2023), CapFit (Cappellari, 2023), PaFit (Krajnović et al., 2006), the Schwarzschild code of van der Marel et al. (1998); Cappellari et al. (2002, 2006), DYNAMITE (van den Bosch et al., 2008; Jethwa et al., 2020; Thater et al., 2022b), scikit-learn (Pedregosa et al., 2011), scikit-optimize (Head et al., 2021), PyTorch (Paszke et al., 2019), and Mathematica (Wolfram Research Inc., 2025).

This work has made use of generative AI for the purposes of providing coding templates and for improving the clarity of the text. No written part of this thesis is taken directly from AI. Where relevant, I mention where I have used generative AI in the text.

Parts of this thesis have been submitted for publication in (or are to be submitted to) the following refereed journals:

Chapter 2: is based on Simon D. A. et al., *Supermassive black hole mass in the massive elliptical galaxy M87 from integral-field stellar dynamics using OASIS and MUSE with adaptive optics: assessing systematic uncertainties*, 2024, MNRAS, 527, 2341. The MUSE integral field data underlying this chapter is publicly available through the ESO science archive facility (`http://archive.eso.org/cms.html`). The SAURON integral field data is available on the ATLAS^{3D} website (`https://www-astro.physics.ox.ac.uk/atlas3d/`). The HST ACS/WFC F850LP image is publicly available through the Mikulski Archive for Space Telescopes (`https://archive.stsci.edu/`). The r-band SDSS image is available from `https://dr12.sdss.org/`. Johanna Hartke performed the the data reduction for the MUSE cube. Michele Cappellari modified the JAM code to allow for

a varying anisotropy. Michele Cappellari provided Figure 2.12 and Figure 2.13.

Chapter 3: is based on Simon D. A. et al., *Accuracy of the Gauss-Hermite Expansion for Stellar Dynamical Modeling and the Recovery of Supermassive Black Hole Masses in High Resolution N-Body Simulations*, to be submitted.

Chapter 4: is based on Simon D. A. et al., *Recovering the Line of Sight Velocity Distribution with Simulated Galaxies: pPXF Best Practices*, to be submitted. The MaNGA data underlying this chapter is publicly available through the online interface Marvin (`https://magrathea.sdss.org/marvin/`, (Cherinka et al., 2019)).

Chapter 5: is based on Simon D. A. et al., *Accuracy of the Gauss-Hermite Expansion for Stellar Dynamical Modeling and the Recovery of Supermassive Black Hole Masses in High Resolution N-Body Simulations*, to be submitted. Haitong Wang and Youquan Fu made some modifications to the axisymmetric Schwarzschild code pertaining to the parallel optimisation and interface.

Chapter 6: is based on a conference proceedings Simon D. A. et al., *Dynamical Modelling of Galactic Kinematics using Neural Networks*, 2026, to appear in the book (Bufano et al., 2026)

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Summary

Spectroscopy has played a major role in many of the discoveries in astronomy over the last century. Integral field spectroscopy is the latest in this history. I give a brief history of extragalactic astronomy with an emphasis on early-type galaxies and how spectroscopy, especially knowledge of galactic kinematics, has improved our understanding of the Universe. This culminates in the development of integral field spectroscopy and some of the key discoveries about elliptical galaxies that resulted from it.

1.2 Discovering Galaxies

1.2.1 The Great Debate

For as long as humans have looked up at the night sky, it has inspired the enduring question: what is our place in the Universe? Throughout history, this question has been answered and reanswered as our understanding of the cosmos has expanded. Copernicus moved the Earth from the centre of the Universe, Galileo revealed that the heavens were filled with worlds, and Newton described a Universe bound together by gravity. A new step came in the 1920s, as astronomers debated the nature of the spiral nebulae, objects in the night sky that appeared as extended objects with spiral arms. Were these spiral nebulae internal to the Milky Way? Or were they merely one of countless "island universes," each a galaxy of its own, scattered across unimaginable distances? When astronomers Harlow Shapley and Heber D. Curtis met to debate this question, later known as the Great Debate[1], it marked another

[1]For a thorough history of this and early 20th century astronomy, see Bartusiak (2010).

turning point in humanity's ongoing effort to locate itself within the vastness of space.

Shapley argued that the Milky Way was an enormous galaxy encompassing the entire known Universe, possibly 300,000 light-years across, with the Sun located far from the centre. He based his reasoning on his studies of globular clusters. Using Cepheid variable stars as distance indicators (Leavitt, 1907; Leavitt & Pickering, 1912), he estimated that the most distant globular clusters were over 200,000 light-years away from Earth (Shapley, 1918). Furthermore, the globular clusters he observed were not uniformly distributed on the sky but had a greater number density in the direction of the Sagittarius constellation. Assuming that globular clusters are uniformly distributed in the galaxy, this suggests that the Earth is far from the galactic centre, with Sagittarius near its heart. Consequently, he believed that the spiral nebulae, such as the Andromeda Nebula, were relatively small gas clouds within our own galaxy, not separate "island universes." He had a few reasons to think this. The first was that a few very bright stars, brighter than the entire nebulae, had been observed in Andromeda. A remarkable mechanism would be required to explain how such an object could become so bright at such a great distance. Additionally, observations by Adriaan van Maanen had shown that some of these spiral nebulae appeared to have undergone rotation across different photometric observations (van Maanen, 1916). To observe the rotation of spiral nebulae on human time scales would very strongly constrain the size of these objects such that they would have to reside within the Milky Way.

Curtis, on the other hand, contended that the Milky Way was significantly smaller, at about 30,000 light-years across, and that the spiral nebulae were in fact independent galaxies, or "island universes," similar to the Milky Way. Curtis disputed the globular cluster distances presented by Shapley, arguing that there was a very high disagreement between different approaches measuring the distances to globular clusters. As an example, Curtis compared several distances determined to M13. The results ranged from 170 light-years at the low end to over 100,000 light-years (Shapley's measurement) at the high end[2]

Curtis also noted that stars, globular clusters, and planetary nebulae were all concentrated along the galactic plane. The exception to this was spiral nebulae which were more closely concentrated outside of the galactic plane. This was a big problem as the leading idea apart from spiral nebulae being island universes was that they were primordial solar systems. But if this were the case, shouldn't they be most densely situated in the Milky Way disk? This would only make sense if they were

[2]A modern measurement gives around 24,000 light-years.(Gontcharov et al., 2020).

beyond the dark lanes of the Milky Way. And as an additional point, these spiral nebulae had also been observed to have their own dark lanes.

After the Great Debate, Shapley soon left Mount Wilson Observatory to become the head of the Harvard College Observatory. Around this time, a new PhD graduate named Edwin Hubble was hired to work at Mount Wilson Observatory where he would take up the mystery of the spiral nebulae.

In 1923, Hubble started observing the Andromeda Nebula. In one exposure, he observed three new stars which he assumed were novae. However, when he compared his results with older observations, he found that the star reappeared. This strongly suggested that what he was observing was not a nova but rather a Cepheid variable star within Andromeda.

To prove this, he made a series of observations in 1924 to measure the period of the fluctuations of this star. He found a period of 31 days. Using the apparent magnitude of the star, he determined the distance to this star to be about 1 million light-years, well beyond the size of the Milky Way proposed by either Curtis or Shapley.

The results of this first went public in November of 1924, when Edwin Hubble was featured in a New York Times article (The New York Times, 1924) claiming that for the first time he had "confirmation of the view that the spiral nebulae, which appear in the heavens as whirling clouds, are in reality distant stellar systems, or 'Island universes'." Hubble continued observing the Andromeda Nebula, now the Andromeda Galaxy, ultimately finding 50 variable stars. He published these results in Hubble (1925, 1926a, 1929b).

With regard to the Great Debate, both Shapley and Curtis had been partly correct and partly wrong. Shapley was correct that the Sun was off-centre in the galaxy, and his size estimate of the Milky Way wasn't too far from the truth (a modern estimate of the diameter is closer to 100,000 light-years). But what about his objections to the spiral nebulae as galaxies? The observation of a star outshining the Andromeda Galaxy (what we now refer to as SN 1885A) would only be explained later with the discovery of supernovae. What about the observations of spiral nebulae rotating? This would be a killer of the island universe theory if true. Unfortunately for van Maanen, subsequent attempts to reproduce his findings failed. He had made his determination using photographic plates taken from different observatories meaning that any misalignment could induce the appearance of rotation. Ultimately, nearly two decades after his initial publication, van Maanen retracted his claims, saying that the observed rotation was within the observational systematic errors (van Maanen, 1935).

This didn't quite mark the end of the Great Debate. Shapley was initially sceptical of these results from Hubble until Hubble personally sent him a copy of the light curve of his first variable star. It was at this point that Shapley, who was very familiar with Cepheids from his work on globular clusters, was convinced, and reportedly told his colleagues that 'this is the letter that destroyed my Universe.' But the destruction of Shapley's Universe meant the creation of a much larger one. One with limits not set by the bounds of the Milky Way but by our ability to look out into the Universe.

1.2.2 Classification of Galaxies

Once it was confirmed that these 'island universes' were not native to the Milky Way but were of extragalactic origin, it was natural to begin studying them by classifying them according to their observed characteristics. In fact, attempts at classifying novae date back as far as William Herschel (Herschel, 1786).

The first modern classification scheme closely resembling the well-known Hubble tuning fork diagram was put forth by John Reynolds in Reynolds (1920)[3]. In this work, he describes one continuous classification sequence with 'spirals consisting entirely of amorphous nebulosity' (elliptical galaxies) at one end, and pure spirals at the other end. This system, however, is incomplete as it failed to account for the inclusion of barred spirals (Curtis et al., 1918).

Hubble built on this, putting forward a preliminary classification system in Hubble (1922) and his famous classification system in Hubble (1926b). A key detail, however, is that the famous 'tuning fork' diagram does not appear in Hubble (1926b). The use of a tuning fork type was first put forward by James Jeans in Jeans (1928) where he notes that a 'proper representation demands a Y-shaped diagram.' The first appearance of the modern 'Hubble tuning fork' diagram appeared in the popular book "The realm of the nebulæ" (Hubble, 1936). A plot of this diagram, along with Jeans' proposed diagram, is shown in Figure 1.2. In this scheme Hubble calls the ellipticals 'early-type' and spiral galaxies 'late-type.'[4].

[3]Hubble famously failed to cite Reynolds when putting forth his own (for a recent analysis of Hubble's legacy, see Way (2013)). He does, bizarrely, accuse fellow astronomer Knut Lundmark of plagiarism in Hubble (1926b) over a similar classification put forward by Lundmark in Lundmark (1926). Lundmark later published a full response, ending with the quote 'as to Hubble's way of acknowledging his predecessors I have no reason to enter upon this question here.'

[4]It is often pointed out that the terms early and late-type are actually reversed with regards to the temporal progression of galaxies. Hubble himself notes in Hubble (1926b) that the terms early and late aren't meant to denote a temporal progression, but rather a progression from simple (elliptical) to more complex (spiral).

One subtlety of the tuning fork diagram is the implicit recognition of elliptical nebulae as external galaxies. Most of the work done on 'island universes' was done with respect to spiral galaxies. It was only in Andromeda and M33 that Hubble observed Cepheid variable stars allowing him to make distance measurements, not in elliptical nebulae. Hubble presents several arguments for why these objects are external galaxies in Hubble (1926b). The first is that these objects have a clear correlation between their apparent size and luminosity, suggesting that this is the same class of object but observed at different distances (Figure 1.1). The second, as is emphasized in the tuning fork diagram, is that these elliptical nebulae often resembled the bulges of spiral galaxies, suggesting an evolutionary path between them.

1.3 Extragalactic Spectroscopy

1.3.1 Redshifts

One gets the impression from the lack of emphasis on spectroscopy in the Great Debate that spectroscopy, a tool which ought to be able to readily discern the nature of the spiral nebulae, was either technologically immature or not widely available. In fact, neither of these is the case. The first extragalactic spectrum was taken by J. Scheiner in Scheiner (1899); Bronkalla & Oleak (1987)[5]. Scheiner observed that the spectrum was continuous, bearing close resemblance to a solar spectrum. This proved, before it was confirmed that spiral nebulae were island universes, that they were stellar systems rather than gaseous systems.

The next step was to measure the motions of spiral galaxies. The first redshift was calculated by Vesto Slipher for the Andromeda Galaxy in Slipher (1913). He expanded this to a larger group of 25 galaxies in Slipher (1917), observing that all but four were moving away from the Earth, with some having a recession velocity of up to 1100 km s^{-1}, or twice the modern value of the Milky Way escape velocity. It is worth remembering that at the time it wasn't clear what to do with these discoveries as it hadn't been confirmed that these objects were in fact external galaxies and consequently these results were treated with some scepticism. Slipher tries to make sense of this bias towards a recession velocity by claiming that the objects he observes are not uniformly distributed on the sky, and that this may result in the observed

[5]The spectrum itself remained unpublished until 1987 when Vera Rubin had requested to locate it at the Potsdam Observatory. A remarkable discovery given that Germany had since been through two World Wars (Rubin, 1995).

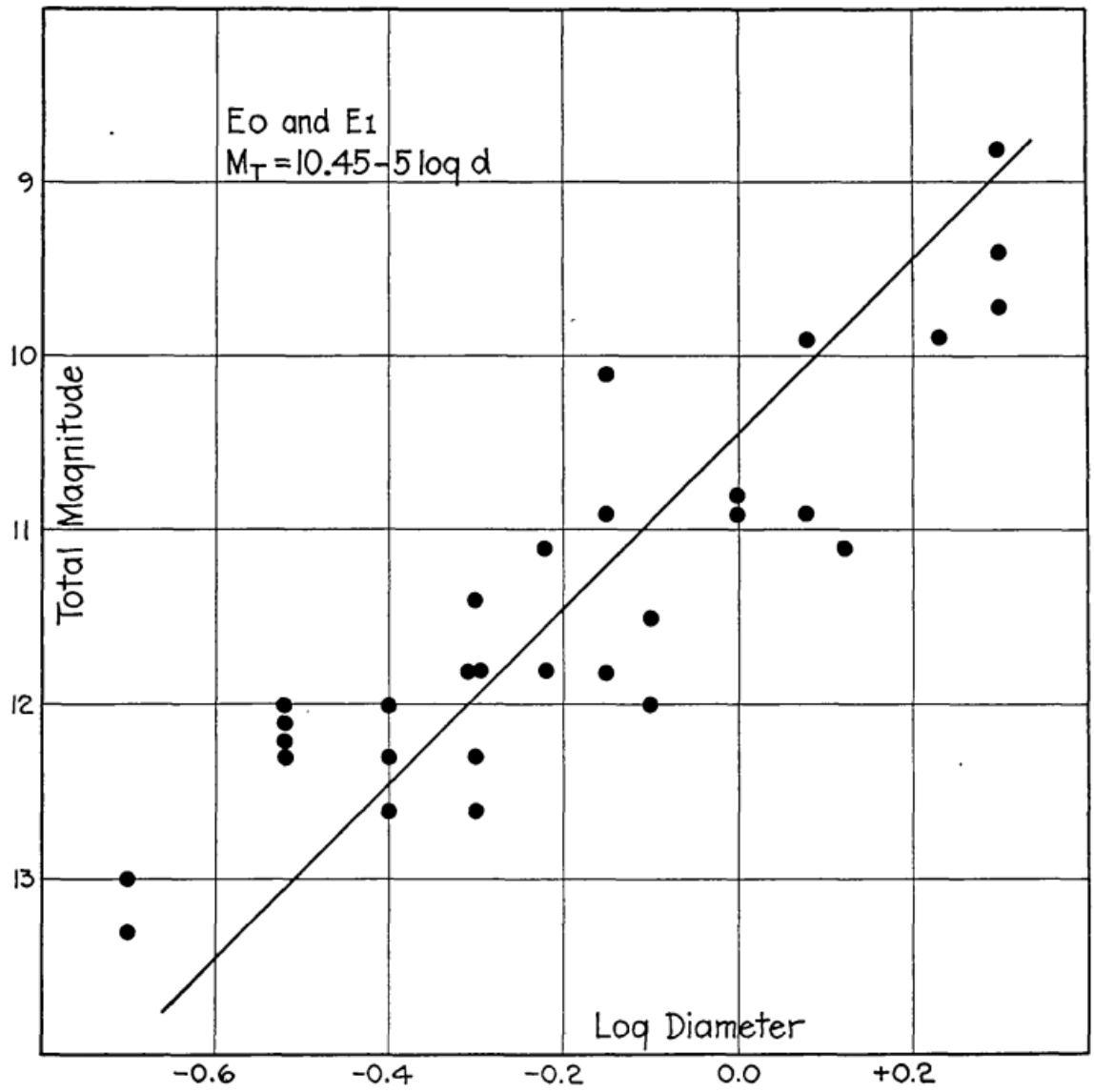

Figure 1.1: Relation between luminosity and diameter among E0 and E1 nebulae (taken from figure 2 of Hubble (1926b)). A linear relationship suggests that E0 and E1 nebulae with different sizes are the same class of object viewed at a different distance.

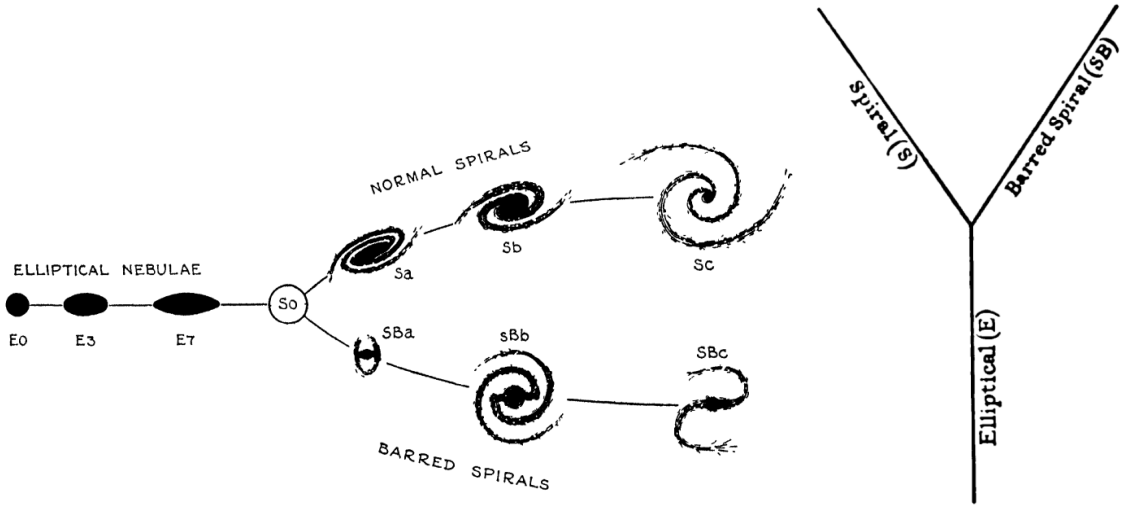

Figure 1.2: The left is Hubble's classification scheme of Hubble (1936), and the right is the proposed Y-shaped diagram put forward by James Jeans in Jeans (1928).

bias. But the magnitude of the velocities, he states, is clear evidence for the island universe theory.

In retrospect this appears a knockout proof that spiral nebulae are external galaxies. However, the historical context is a bit more complicated (see Freeman (2013)). The key confusing factor was van Maanen's work suggesting that spiral galaxies were rotating on human time scales (van Maanen, 1916). Without this confusion, it seems likely that the island universe debate may have been settled several years earlier.

Slipher's measured recession velocities, broadly speaking, were accurate and very close to the contemporary values. Though they were not fully appreciated in the context of the Great Debate, they played a fundamental role in the discovery of the expanding Universe.

The first attempt at studying the relationship between distance and velocity was done by Lundmark in Lundmark (1925). There, he first finds that the distribution of velocities is uniform on the sky. The second thing he notes is that a "correlation is shown between apparent dimensions and radial velocity, in the sense that the smaller and presumably more distant spirals have the higher space-velocity." This finding was properly quantified by Hubble who used Cepheid variable stars to better determine the distances to the galaxies for which redshifts had been determined. The sources of these redshifts were primarily from Slipher (though no attribution is given), with the rest taken by Milton Humason who was working with Hubble at Mt. Wilson. These results revealed a linear relationship between a galaxy's distance and its recession

velocity (Hubble, 1929a). This result would later be called Hubble's law, and is represented with the equation

$$v = H_0 D \tag{1.1}$$

1.3.2 Galaxy Rotation Curves

Spectroscopy is useful not just for measuring the net motion of galaxies, but also the internal motions. Van Maanen's attempted measurement of the rotation of spiral galaxies, while ultimately flawed, was a natural attempt to fill a gap in the knowledge about the internal motions of these objects. With the realization that these objects were significantly larger than previously thought, it became clear that the best approach to understanding their motions was through spectroscopy. The first observations of rotation made from spectroscopy were due to Slipher in Slipher (1914). Here, he reports observing that a spectrum of NGC 4594 showed inclined nebular lines, a clear sign of rotation he was familiar with from previous spectroscopic studies of the rotation of planets. This was corroborated by reports from Pease in Pease (1916, 1918), who observed rotation in NGC 4594 of ± 330 km s^{-1} and in M31 of ± 100 km s^{-1} [6].

Work on galaxy rotation curves progressed slowly due to the large observing times required to make these observations and the associated systematic uncertainties. Some progress came mid-century when Horace Babcock made some of the first measurements of the extended rotation curve of the Andromeda Galaxy, finding that velocities seemed to increase rather than fall as would have been expected from the inverse square law (Babcock, 1939). This work was expanded somewhat by Nicholas Mayall who found similar results to Babcock but who was suspicious of the results (Mayall, 1951).

A key advancement came a few decades later with the development of image-tube spectrographs by Kent Ford during his time at Carnegie Institution of Washington (now Carnegie Science). These work by taking the observed light and directing them towards a photocathode which emits electrons that are then accelerated towards a photographic plate. This effectively amplifies the signal, allowing for faint objects to be observed much more quickly. During this time there was another young scientist at Carnegie named Vera Rubin. Rubin was interested in the motions of stars near

[6] He notes that 'the results seem to be inconsistent with a system involving planetary motion about a central nucleus, since this would require an increase of linear velocity toward the centre of the nebula.' This piece of evidence was not mentioned in the Great Debate though in retrospect it is a strong argument in favour of the island universe theory.

the outskirts of galaxies. She first investigated this with a study of the stars in the outskirts of the Milky Way, finding "For $R > 8.5$ kpc, the stellar curve is flat, and does not decrease as is expected for Keplerian orbits" (Rubin et al., 1962; Rubin, 1965). With Ford's new instrument, it was natural to ask the same question about external galaxies. There, she continued to find flat rotation curves, first for the Andromeda Galaxy (Rubin & Ford, 1970) and later for a larger group of galaxies in Rubin et al. (1978). This large sample of flat rotation curves provided some of the most definitive evidence for the existence of dark matter.

1.3.3 Elliptical Galaxies

Rubin was able to measure the wide-field rotation curves of spiral galaxies by looking for Hα emissions in their outskirts. This worked well for spirals which have a larger gas fraction but did not work well for ellipticals which tended to have very little gas. This meant that to measure the kinematics of elliptical galaxies required using stellar absorption features which were less bright than gas emissions.

The first rotation curve for an elliptical galaxy was measured in Bertola & Capaccioli (1975) which used an image-tube spectrograph to measure Ca II H and K lines. They found the unexpected result that the elliptical galaxy NGC 4697, despite appearing flattened, had between 5 and 30 times less angular momentum than the comparable spiral galaxy. This issue was taken up by James Binney in Binney (1976) where he ran a numerical simulation suggesting that elliptical galaxies could, as a result of their formation process, become flattened while maintaining low levels of angular momentum.

Soon after this, a further 13 elliptical galaxies were studied in Illingworth (1977) and another 12 in Schechter & Gunn (1979). These found that the dissipative model by Binney could not completely explain the kinematics of ellipticals. This prompted work showing that allowing a non-isotropic anisotropy could better match the observed $(V/\sigma, \varepsilon)$ relation (Binney, 1978), though the geometry of these galaxies could not be oblate. These observations were later extended to both galaxy bulges (Kormendy & Illingworth, 1982; Kormendy, 1982) and fainter elliptical galaxies (Davies & Illingworth, 1983; Davies et al., 1983). In both cases these objects were found to have higher-rotation than the bright ellipticals previously studied.

This dichotomy was further strengthened by two discoveries using photometry. The first was the discovery that ellipticals tend to come in two classes, those with disky isophotes and those with boxy isophotes (Bender et al., 1988). It was further shown that the galaxies with disky isophotes correspond to the higher-rotation faint

ellipticals, and the boxy isophotes corresponding to the brighter slower-rotating ellipticals (Bender, 1988). One way to make sense of disky isophotes is as stellar disks embedded in an otherwise elliptical galaxy. However, disky isophotes would only be able to be observed if the inclination angle of the disk was close to edge-on (Rix & White, 1990; Rix et al., 1999).

The second discovery using photometry came from space-based observations from the Hubble Space Telescope. Until the development of adaptive optics, typical seeing from ground-based observatories was limited to around 1 arcsecond. This meant that the centres of galaxies could not be probed beyond this resolution. This changed when the Hubble Space Telescope, named in honor of Edwin Hubble, was launched in April of 1990. Being space-based meant that there was no impact from atmospheric seeing, allowing for well-resolved images down to fractions of an arcsecond to be made. Observations of elliptical galaxies revealed two types of characteristic surface brightness profiles in the central region: one which features a cuspy power law shape down to the resolution limit of the Hubble Space Telescope and one which abruptly flattens in the centre of the galaxy (Ferrarese et al., 1994; Lauer et al., 1995a). The power law galaxies corresponded to those with disky isophotes and the core galaxies corresponded to those with boxy isophotes.

1.4 The first Dynamical Models

1.4.1 The Jeans Equations

Spectroscopic methods were measuring the motions of stars and galaxies. But to make the most use of this data, dynamical models needed to be devised to take these measurements and turn them into insights about the mass and motion of galaxies. James Jeans was interested in understanding the motions of stars in the Milky Way. He started by taking the collisionless Boltzmann equation (Binney & Tremaine, 2011, Eq. 4.11)

$$\frac{\partial f}{\partial t} + \sum_{i=1}^{3} \left(v_i \frac{\partial f}{\partial x_i} - \frac{\partial \Phi}{\partial x_i} \frac{\partial f}{\partial v_i} \right) = 0 \tag{1.2}$$

Where t is time, x_i is position and v_i is velocity, Φ is the gravitational potential, and f is the 6-dimensional distribution function describing the probability of finding a star at a given position x_i and velocity v_i. Assuming that the system is stationary gives $\partial f / \partial t = 0$. This equation is also sometimes referred to as the Vlasov equation after Anatoly Vlasov who independently used the equation during his study of plasmas (Vlasov, 1938). In its current form, the collisionless Boltzmann equation is difficult

to work with due to the number of unknowns. Jeans' insight was to write equations not for the unknown probability distribution of the galaxy, but for the second velocity moments of the distribution. Converting to cylindrical coordinates and integrating over the velocities gives (see chapter 5 for a full description) (Jeans, 1922; Binney & Tremaine, 2011, Eqs. 4.222a-b)

$$\frac{\nu\overline{v_R^2} - \nu\overline{v_\phi^2}}{R} + \frac{\partial(\nu\overline{v_R^2})}{\partial z} + \frac{\partial(\nu\overline{v_R v_z})}{\partial z} = -\nu\frac{\partial\Phi}{\partial R} \tag{1.3}$$

$$\frac{\nu\overline{v_R v_z}}{R} + \frac{\partial(\nu\overline{v_z^2})}{\partial z} + \frac{\partial(\nu\overline{v_R v_z})}{\partial R} = -\nu\frac{\partial\Phi}{\partial z} \tag{1.4}$$

These are now called the Jeans equations. He described these as 'hydrodynamical equations of motion for the stars.' While Jeans originally intended to use these equations to study the motion of stars in the Milky Way (in 1922 it wasn't agree on that other galaxies even existed), they ultimately found use describing external galaxies (see Binney (1982)). However, there were some limitations with this approach. At this time it was assumed that $\sigma_R = \sigma_z$ (that is, the z and R components of the velocity ellipsoid were equal) and $\overline{v_R v_Z} = 0$. However, van der Marel (1991) showed for a group of 37 galaxies that in general it is not true that $\sigma_R = \sigma_z$, therefore motivating the development of more general dynamical modeling methods.

1.4.2 Schwarzschild's Orbit Superposition Method

The suggestion in Binney (1978) that some elliptical galaxies might not be oblate warranted further investigation. One outstanding question was if it was possible to produce a stellar system with triaxial geometry in equilibrium. This question was taken up by Martin Schwarzschild in Schwarzschild (1979). His approach was to leverage computation to show that, using individual orbits as basis functions, it was possible to reproduce a stable triaxial geometry. His exact setup was to

- Select a 3D stellar distribution
- Calculate the gravitational potential using the Poisson equation
- Integrate a number of stellar orbits in the potential
- Find the non-negative linear combination of orbits best able to reproduce the stellar distribution

This setup was natural to apply to questions in dynamics. This was first done by Douglas Richstone who measured line of sight velocity dispersions predicted from a galaxy generated using Schwarzschild's method (Richstone, 1980). He then turned the question around, asking how measurements of kinematics can constrain information about the gravitational potential, namely the M/L, if one measurement of the dispersion is available at a given radius (Richstone & Tremaine, 1984). This ultimately led to the modeling of full spatially resolved line of sight velocity distribution data in Rix et al. (1997) in spherical geometry, and later in axisymmetric and triaxial geometries (van der Marel et al., 1998; van den Bosch et al., 2008). These models were then applied to a variety of systems using long-slit data (e.g. Gebhardt et al. (2000a); Cappellari et al. (2002); Valluri et al. (2004)).

1.5 Integral Field Spectroscopy

1.5.1 Motivation and Development

The combination of long-slit spectroscopic data and anisotropic dynamical models of elliptical galaxies produced significant modelling challenges. Binney & Mamon (1982) found that the mass and anisotropy were degenerate, making it difficult to reliably recover the mass distribution in ellipticals. It was also noted in Bacon et al. (1983); Bacon (1985) that the lack of two-dimensional kinematic data made it impossible to constrain the shape of the anisotropy profile (see also Valluri et al. (2004)).

The key was the lack of two-dimensional resolved data. There were some possible solutions to this. Multiple long-slit observations can be taken in series, to provide more spatial information on the kinematics for testing dynamical models, though this requires regular careful alignment of the slit and makes the PSF determination challenging (Wilkinson et al., 1986). Another approach known at the time was the use of a scanning Fabry-Perot interferometer which produces a two-dimensional image and scans the wavelength dimension over time (Boulesteix et al., 1984). Bacon attempted to use this to measure stellar kinematics from absorption features in 1986 but encountered a number of issues, including low signal-to-noise and poor accuracy in the wavelength dimension due to the time scanning (Bacon, 2024). This motivated the development of a new instrument better optimized to produce two-dimensional kinematics.

A natural approach was first considered by George Courtès, who proposed using a micro-lens array which, when properly aligned, would place spectra in parallel but not overlapping on a CCD (see Figure 1.3 for an example) (Courtès, 1980). A similar

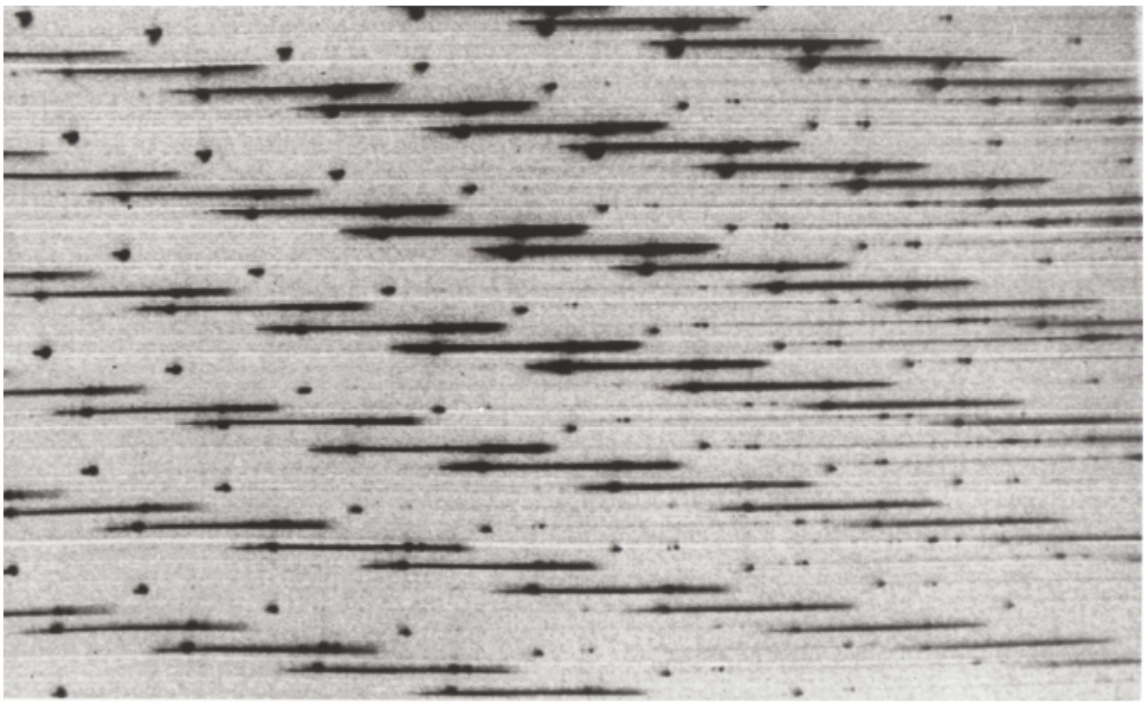

Figure 1.3: Raw image of the nucleus of M51 taken with the TIGER spectrograph. This was the first science exposure taken by the instrument in 1987 (taken from Bacon (2024)). The importance of the invention of CCDs is immediately apparent as it is relatively straightforward to project each spectrum on a different part of the CCD.

approach was taken in Vanderriest (1980) with the difference that optical fiber bundles were used rather than a CCD. This, however, came with the disadvantages of fiber bundles, namely throughput losses and low transmission in the blue. A prototype using the micro-lens and CCD approach, nicknamed TIGER, was developed and tested at the Canada-France-Hawaii Telescope (CFHT) (Bacon et al., 1988; Courtes et al., 1988). It was used for a broad range of science use cases, from studying gravitational lensed systems and quasars, to the surface of Mars (Bacon et al., 1995). With regard to the stellar kinematics of galaxies, this culminated in the first publication of 2D velocity and dispersion maps, done for the nucleus of the Andromeda Galaxy in Bacon et al. (1994). This marked the birth of integral field spectroscopy (IFS).

1.5.2 First Generation IFUs

The success of the TIGER integral field unit (IFU) motivated a more user-friendly instrument to better serve the community. The result was the Optically Adaptive System for Imaging Spectroscopy instrument (OASIS). OASIS was the first IFU designed to take advantage of adaptive optics. It was originally hosted at the CFHT

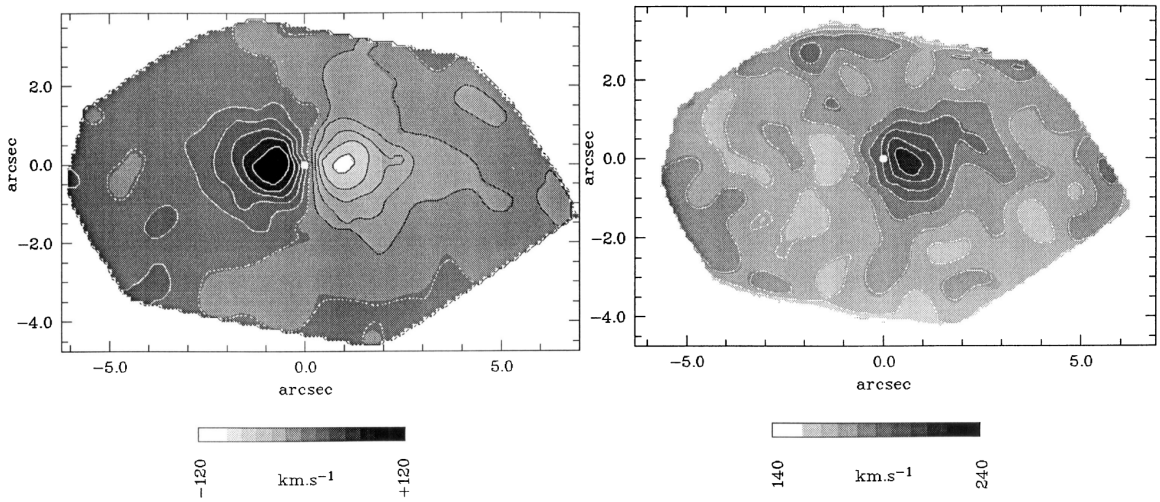

Figure 1.4: Velocity and dispersion maps of the Andromeda Galaxy determined using the TIGER spectrograph and aligned with the major axis of the nucleus. The location of the photometric centre of the nucleus is marked with a white dot. (taken from Bacon et al. (1994)). The dispersion is off-center from the photometric center of the galaxy due to the presence of a double nucleus (Light et al., 1974).

before being relocated to the William Herschel Telescope on La Palma. OASIS had a small field of view[7] of $8'' \times 10''$ with a wavelength range of 4760–5558Å. This was carefully chosen to surround the Mgb absorption feature, a useful deep absorption feature for determining stellar kinematics. It was realised towards the end of its development that OASIS would not be able to deliver the sorts of discoveries IFS was capable of. This provided a strong incentive to develop an instrument that would be well-optimised to investigate the wide-field two-dimensional kinematics of early-type galaxies, the very question which originally motivated the development of IFS.

This led to the development of the SAURON instrument Bacon et al. (2001). SAURON had a field of view of $33'' \times 41''$ with a $0.94''$ spatial scale, significantly larger than OASIS and allowing many nearby galaxies to be studied out to one effective radius. The wavelength range of the instrument was 4800–5380 Å with a 4.2 Å spectral resolution (corresponding to an instrumental dispersion of $\sigma_{\rm inst} = 108$ km s^{-1}). The relatively large instrumental dispersion meant that the ideal target for this instrument would be local massive galaxies.

The SAURON IFU was used first for the SAURON survey of 48 nearby early-type galaxies (de Zeeuw et al., 2002) and later expanded to the Atlas^{3D} survey of 260 galaxies (Cappellari et al., 2011a). This was a volume-limited (1.16×10^5 Mpc^3) sample of

[7]As I'll show in chapter 2, it is even small for some black hole studies.

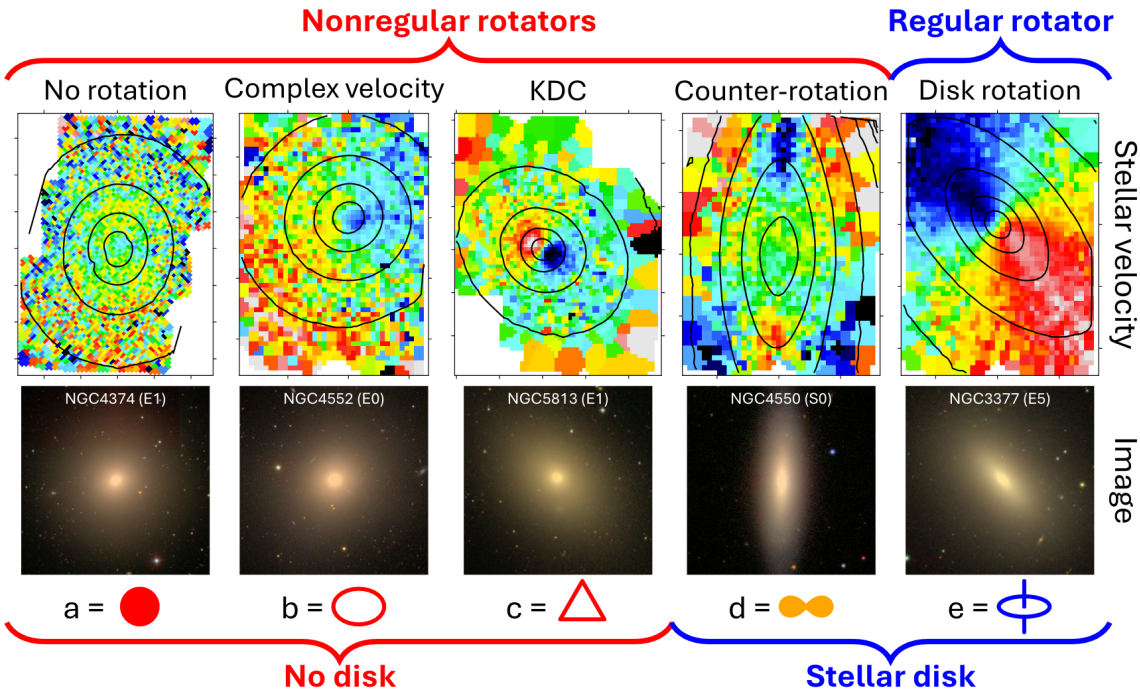

Figure 1.5: Kinematic classification of galaxies from Krajnović et al. (2011). Taken from Cappellari (2026). The difference between different maps is immediately quite stark, motivating the fast and slow rotator classification.

nearby early-type galaxies with the advantage of being statistically representative of nearby galaxies without the selection effects of the SAURON survey.

1.5.3 Kinematic Morphology

One key result immediately evident from the SAURON maps is the clear presence of distinct morphological types in the stellar kinematics (Figure 1.5). These include galaxies with no rotation, complex rotation, counter-rotation, kinematically decoupled cores, and regular rotation. Regular and counter-rotation are well explained by the presence of a disk, while no rotation, complex velocity, and kinematically decoupled cores fit more into the anisotropic, non-oblate models of Binney. This initially provided strong evidence for a dichotomy of early-type galaxies, with the regular and counter-rotating galaxies the "fast rotators", and the remaining the "slow rotators." However, one outstanding question remained. A decrease in rotation would also be observed if a regular rotator galaxy was viewed close to face-on. To establish the physical significance of these two classes would require showing that slow rotators were not just near face-on fast rotators.

1.5.4 Fast and Slow Rotators

Solving the inclination problem came in several steps. First, Cappellari et al. (2006) performed Jeans dynamical modelling to directly measure the inclination of the available galaxies. This revealed that many objects on the border of being classified as a slow rotator were in fact near face-on fast rotators. Additionally, the distribution of fast rotators on key diagnostic diagrams such as the $(V/\sigma, \epsilon)$ or (λ_r, ϵ) diagram (where ϵ is the observed ellipticity of the galaxy) could be reproduced assuming they follow simple relationships (Cappellari et al., 2007, Fig. C3).

The suggestion that there are two classes of early-type galaxies separated by their rotation demands a new parameter to allow their quantitative study. Emsellem et al. (2007) introduced the parameter λ_r, given definitionally and for IFS as

$$\lambda_r \equiv \frac{\langle R|V|\rangle}{\langle R\sqrt{V^2+\sigma^2}\rangle} \tag{1.5}$$

$$= \frac{\sum_{i=1}^{N_p} F_i R_i |V_i|}{\sum_{i=1}^{N_p} F_i R_i \sqrt{V_i^2+\sigma_i^2}} \tag{1.6}$$

where F_i is the flux in the ith bin, R_i the distance to the centre of the galaxy, and V_i and σ_i the corresponding velocity and dispersion of the LOSVD. The interpretation of this parameter is as the stellar projected specific angular momentum. The key is to note that the numerator is the projected angular momentum while the denominator $\sqrt{V^2+\sigma^2}$ is the second moment and proportional to the mass.

There are several parameters which show the bimodality of early-type galaxies. Slow rotators tend to show a pronounced misalignment between the photometric major axis and axis of rotation. This is a result of the fact that they tend as a class to, on average, be close to spherical ($\varepsilon < 0.4$), making measurements of the photometric major axis unreliable. Furthermore, triaxial galaxies can have twists in their photometry, resulting in a kinematic-photometric misalignment (Figure 1.6). The presence of two distinct populations of elliptical galaxies has been independently reproduced by other groups working with larger sets of galaxies (van de Sande et al., 2021).

The exact cutoff used to differentiate between fast and slow rotators has been updated several times over the years. Emsellem et al. (2007) describes galaxies with $\lambda_r < 0.1$ as slow rotators and galaxies with $\lambda_r > 0.1$ as fast rotators. This was updated in Emsellem et al. (2011) to slow rotators satisfying $\lambda_r < 0.31\sqrt{\epsilon}$ and fast rotators greater than that. This has been refined most recently to defining slow rotators as satisfying $\lambda_r < 0.08 + \epsilon/4$ with $\epsilon < 0.4$ (Cappellari, 2016). The role of the limit on ϵ

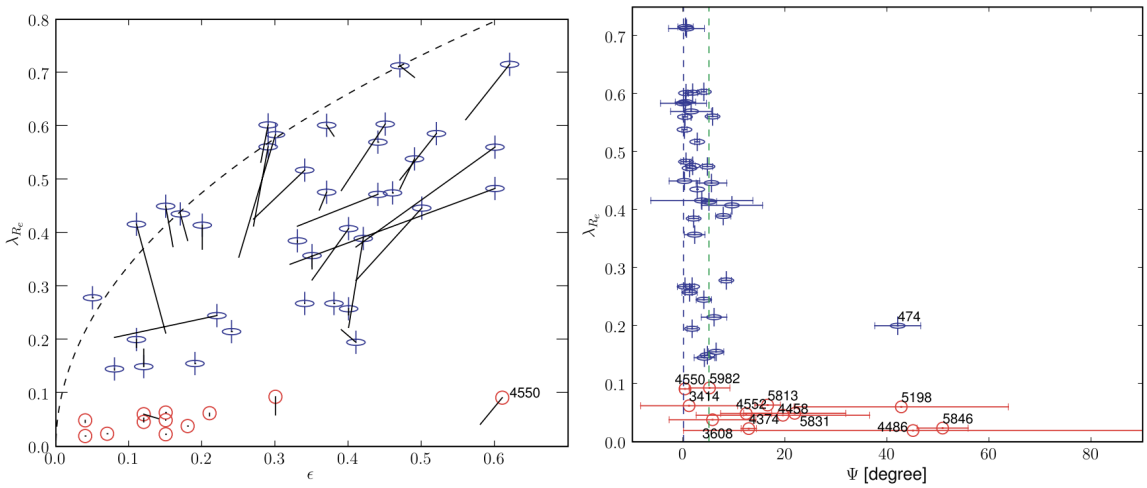

Figure 1.6: Plots showing the parameter λ_r and galaxy ellipticity and misalignment angle Ψ, defined as the absolute value of the difference between the photometric major axis and the kinematic major axis. Taken from Emsellem et al. (2007). The black dashed line in the left panel corresponds to the curve expected for isotropic oblate rotators viewed edge-on. It is especially apparent in the λ_{R_e} vs Ψ figure that there is a clear change in the distribution at low λ_{R_e}.

is to protect against contamination from galaxies with counter-rotating disks which can have a low λ_r but are often flattened.

1.5.5 Galaxy Classification

Hubble's tuning fork, as it pertains to elliptical galaxies, depends on the observed ellipticity of the galaxy. This suffers from the fact that the classification depends on the inclination, rather than reflecting the intrinsic physical properties of the galaxies. As a result, a number of modifications to the Hubble tuning fork have been proposed over the years (see Sandage (2005) for a review). One contribution of importance was that of the 'anemic spirals' in van den Bergh (1976). Van den Bergh correctly pointed out that there is a smooth continuum of galaxy properties between spirals and lenticular galaxies. As such, he proposed including an intermediate class of galaxy, dubbed the anemic spiral, to better emphasize the transition between the two regimes.

Another update came as a result of photometry revealing two classes of elliptical galaxy photometry, some with boxy isophotes and others with disky isophotes. Kormendy & Bender (1996) introduces the E(d) and E(b) classification for ellipticals with disky and boxy isophotes, respectively, with the E(d) class intermediate between E(b) and the lenticulars. While this is an improvement over Hubble's ellipticity system, it

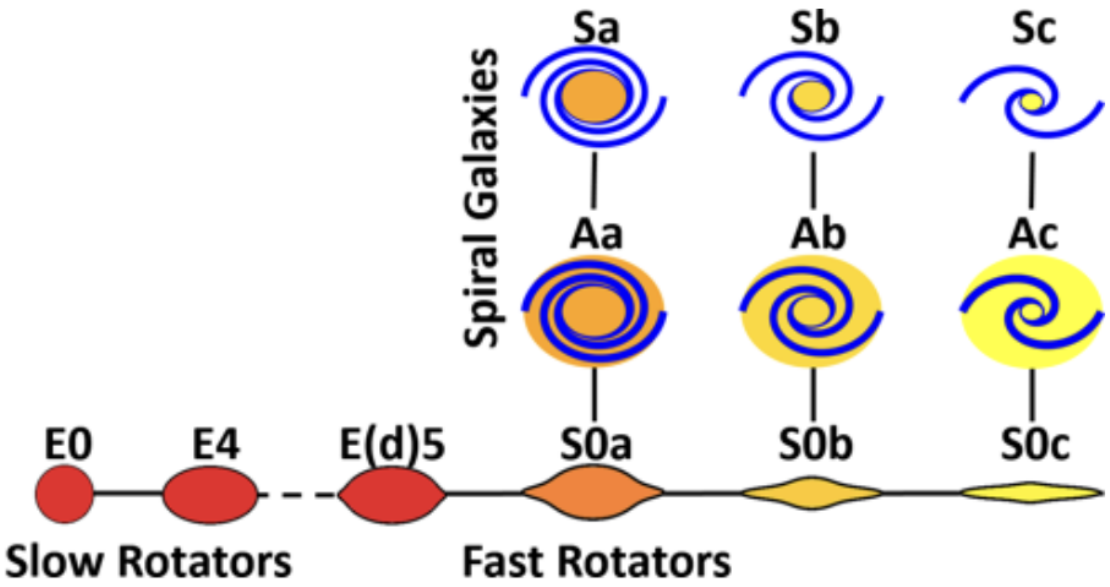

Figure 1.7: The Atlas3D tuning fork. The updated scheme clearly separates the slow rotators with nearly spherical shape from the fast rotators. Unlike the Hubble tuning fork, however, this diagram has a comb structure. Taken from Cappellari et al. (2011b).

still suffers from a projection effect as disky isophotes are not apparent when a galaxy is viewed face-on (Rix et al. (1999), see Cappellari (2026) for a discussion).

The solution to this comes immediately from IFS, as face-on stellar disks can be identified with integral field data, even though they can't be identified by photometry (Cappellari, 2016). The resulting classification system, using the fast and slow rotator classification since it is free of projection effects, is shown in Figure 1.7 (Cappellari et al., 2011b). This combines the van den Bergh 'anemic spiral' intermediate class, with the fast and slow rotator classification for ellipticals, an improvement over the E(b) and E(d) classes from Kormendy & Bender (1996). Lastly, comparing the fast-slow rotator dichotomy to the Hubble classification, we see that one third of slow rotators would be classed as ellipticals under Hubble's classification scheme Emsellem et al. (2011). This confirms that Hubble's classification, while well motivated at the time, is not physically meaningful, and demonstrates the demand for an updated classification system.

1.6 Thesis Outline

Understanding the kinematics of galaxies has played a fundamental role in transforming our understanding of the Universe over the past century. Many major discoveries, from the existence of 'island Universes' to dark matter and the expanding Universe have all been strongly influenced by the use of kinematics.

This thesis is primarily concerned with elliptical galaxies and the techniques used to study their kinematics. In this introduction, I have summarized some of the history of extragalactic astronomy and the approaches used to study external galaxies, particularly ellipticals and lenticulars, using spectroscopy and especially integral field spectroscopy. In this thesis, I investigate in greater detail some of the techniques used in integral field spectroscopy.

The research I have conducted is presented in four chapters. A summary of these is provided here:

- In chapter 2 I use integral field data from MUSE, OASIS, and SAURON to measure the supermassive black hole in the massive elliptical galaxy M87. In doing so, we pioneer a new approach for measuring the PSF in active galaxies, as well as measuring the central stellar surface brightness obscured by the AGN. We also develop a new modelling approach with JAM using varying radial anisotropy.

- In chapter 3 I take a closer look at the extraction of stellar kinematics from integral field data, focusing on the systematic impacts of using a Gauss-Hermite parameterisation for the line of sight velocity distribution.

- In real observations, however, the primary source of systematic uncertainty is not just the choice of the function form of the fitted LOSVD, but also uncertainties due to the use of stellar template spectra. We discuss this in chapter 4.

- In chapter 5 I give a detailed description of JAM and Schwarzschild modelling, and study the robustness of black hole mass measurements using both JAM and Schwarzschild modelling.

- I investigate foundational questions for a machine-learning-based approach to dynamical modelling in chapter 6.

- Lastly, I look to the future of this field and discuss promising directions for future research in chapter 7.

Chapter 2

The Supermassive Black Hole in M87

2.1 Summary

The massive elliptical galaxy M87 has been the subject of several supermassive black hole mass measurements from stellar dynamics, gas dynamics, and recently the black hole shadow by the Event Horizon Telescope (EHT). This uniquely positions M87 as a benchmark for alternative black hole mass determination methods. Here we use stellar kinematics extracted from integral-field spectroscopy observations with Adaptive Optics (AO) using MUSE and OASIS. We exploit our high-resolution integral field spectroscopy to spectrally decompose the central AGN from the stars. We derive an accurate inner stellar-density profile and find it is flatter than previously assumed. We also use the spectrally-extracted AGN as a reference to accurately determine the observed MUSE and OASIS AO PSF. We then perform Jeans Anisotropic Modelling (JAM), with a new flexible spatially-variable anisotropy, and measure the anisotropy profile, stellar mass-to-light variations, inner dark matter fraction, and black hole mass. Our preferred black hole mass is $M_{\rm BH} = (8.7 \pm 1.2[\text{random}] \pm 1.3[\text{systematic}]) \times 10^9\ M_\odot$. However, using the inner stellar density from previous studies, we find a preferred black hole mass of $M_{\rm BH} = (5.5^{+0.5}_{-0.3}) \times 10^9\ M_\odot$, consistent with previous work. We find that this is the primary cause of the difference between our results and previous work, in addition to smaller contributions due to kinematics and modelling method. We conduct numerous systematic tests of the kinematics and model assumptions and conclude that uncertainties in the black hole mass of M87 from previous determinations may have been underestimated and further analyses are needed.

2.2 Background

Supermassive black holes play an important role in galaxy evolution. This is shown through empirical relations between black hole mass and luminosity (Kormendy & Richstone, 1995; Magorrian et al., 1998), as well as black hole mass and stellar velocity dispersion (Ferrarese & Merritt, 2000; Gebhardt et al., 2000b). The reliability of these relationships depends on accurate black hole mass measurements.

M87 is one of the most massive galaxies of the Virgo cluster and sits at the centre of the main sub-cluster (e.g. Cappellari et al., 2011b, fig. 7). It is a prototypical massive slow rotator early-type galaxy, with a large Sersic (Sérsic, 1963) index and a core in the nuclear surface brightness profile (Kormendy et al., 2009), and fits all characteristics for having assembled most of its mass by dry mergers (see review by Cappellari, 2016). Like other galaxies of its type, M87 contains a supermassive black hole at its centre (Kormendy & Ho, 2013) whose event horizon has the largest angular size of any known black hole outside of the Milky Way, making it a valuable target for black hole studies. The measurement of the black hole shadow by the Event Horizon Telescope (EHT) (Event Horizon Telescope Collaboration et al., 2019) makes M87 the first galaxy for which direct imaging of the supermassive black hole has taken place. This has the potential to serve as a powerful test of the general theory of relativity: but only if we can confirm through independent measurements that the black hole mass recovered assuming general relativity is correct. The measurement of the black hole shadow from the EHT (Event Horizon Telescope Collaboration et al., 2019) assuming general relativity determined the black hole mass to be $(6.5 \pm 0.7) \times 10^9 \mathrm{M}_\odot$. In the case of M87 there are two other such classes of measurement. Gas dynamical measurements by Harms et al. (1994), Macchetto et al. (1997), and Walsh et al. (2013) have measured the masses $(2.7 \pm 0.8) \times 10^9 \mathrm{M}_\odot$, $(3.6 \pm 1.0) \times 10^9 \mathrm{M}_\odot$, and $(3.3^{+0.8}_{-0.7}) \times 10^9 \mathrm{M}_\odot$, respectively. Stellar dynamical measurements (Gebhardt & Thomas, 2009; Gebhardt et al., 2011; Liepold et al., 2023a) using the orbital superposition method of Schwarzschild (Schwarzschild, 1979) have been made which found a black hole masses of $(6.0 \pm 0.5) \times 10^9 \mathrm{M}_\odot$, $(6.2 \pm 0.4) \times 10^9 \mathrm{M}_\odot$, and $(5.37^{+0.37}_{-0.25}) \times 10^9 \mathrm{M}_\odot$ respectively. There is thus a discrepancy in the recovered black hole masses by a factor of two depending on the method used. Jeter et al. (2019); Jeter & Broderick (2021) propose that this may be due to unrealistic assumptions made in the gas modelling. They find that more detailed accounting of the radial motion of the gas as well as allowing for a thick gas disk can alleviate the discrepancy (though it should be noted that their models are not fit to data). Recently, Osorno et al. (2023) has produced detailed ionized gas

maps of M87 using the same MUSE data as in this paper, revealing a complex multi-component gas structure. They suggest that the cause of the discrepancy is likely due to incorrect assumptions about the ionized gas disk inclination, though they comment that it is challenging to make an independent black hole mass measurement with this data. The agreement between stellar dynamical measurements and the measurement of the black hole shadow is reassuring, but it is still important to continue testing and independently verifying stellar dynamical models in order to fully understand possible systematics and to improve the robustness of the measurement. In this paper we derive new independent measurements of the black hole mass using stellar dynamics with two different high-resolution integral-field datasets from two different telescopes and a different dynamical modelling approach than previously used.

This paper is laid out as follows: in section 2.3 we introduce the integral field data and photometric data used in this study. In section 2.4 we describe our spectral fitting and discuss a number of tests we performed and several methods of extracting the kinematics that we use. In section 2.5 we use a combination of IFS data with photometry to accurately measure the stellar density profile for M87 down to the region dominated by the AGN. In section 2.6 and section 2.7 we describe the details of our Jeans modelling and present our black hole mass constraints. We compare these with previous observations and discuss a number of systematic uncertainties. Lastly, in section 2.8 we summarize our results and comment on the future landscape for studies of M87.

We take the distance to M87 to be 16.8 Mpc (Event Horizon Telescope Collaboration et al., 2019). All black hole masses quoted are scaled to this distance. This corresponds to a spatial scale of 81.1 pc per 1 arcsecond.

2.3 Data

2.3.1 Integral Field Spectroscopy

We use integral field observations from the Optically Adaptive System for Imaging Spectroscopy (OASIS) spectrograph made on the Canada-France-Hawaii Telescope (CFHT) (McDermid et al., 2006) and the Multi Unit Spectroscopic Explorer (MUSE) (Bacon et al., 2010) in narrow field mode (NFM) on the Very Large Telescope (VLT) with adaptive optics (AO). This gives us two independent views of the central kinematics of M87. To add kinematic information at larger distances, we also use observations from SAURON Bacon et al. (2001).

The OASIS integral field spectrograph (IFS) has a $10'' \times 8''$ field of view with a $0\farcs27 \times 0\farcs27$ pixel scale. For this measurement the spectrograph was configured to cover the wavelength range of 4760-5558 Å with a resolution of 5.4 Å FWHM (corresponding to an instrumental dispersion of $\sigma_{\rm inst} \approx 134$ km s^{-1}) sampled at 1.95 Å per pixel. Three observations were made for 2700 seconds each, which were then combined to form the final image (see McDermid et al. (2006) for a detailed description of the observations and data reduction).

The MUSE observations were made as part of program 0103.B-0581 (PI: N. Nagar) in NFM. The NFM covers a field of view of 7.5"x7.5" with a pixel scale of 0.025"/pix and is used together with the the adaptive optics facility GALACSI (Arsenault et al., 2008; Ströbele et al., 2012), providing laser tomographic AO corrections. The observations were carried out on 20 February 2021 and consist of nine dithered exposures of 700s, resulting in a total exposure time of 6300s. The data were reduced with the standard MUSE data reduction pipeline (Weilbacher et al., 2020) in the ESO reflex environment (Freudling et al., 2013) using the dedicated offset sky exposures for sky subtraction with the standard pipeline parameters. The parameters for source detection and image alignment were optimized to align the individual exposures based on a combination of point sources and the knots of the jet of M87. MUSE covers a wavelength range from 4650-9300 Å with a gap between 5780-6050 Å due to a Na notch filter blocking the light from the four laser guide stars facility (4LGSF). The spectrum is sampled at 1.25 Å per pixel with a resolution of about 2.6 Å FWHM, corresponding to an instrumental dispersion $\sigma_{\rm inst} \approx 63$ km s^{-1}. We show a plot of a spectrum in Figure 2.1. We see that there are still significant skylines starting from 7000Å. We tested whether the remaining sky residuals could be removed using the Zurich Atmospheric Purge (ZAP, Soto et al. (2016)) on sky residual cubes produced by the pipeline, but the improvement was not significant so we proceeded with the original cube. A significant question is how to treat the region with skylines. Curiously, the region with the Calcium triplet appears minimally impacted by skylines. We tested extracting the kinematics in this region using the CAT stellar library (Cenarro et al., 2001). We found that fitting the spectra using a single stellar template fit to the mean spectrum resulted in some of the central fits failing to describe the wavelength region between 0.86μm and 0.87μm. We show our fits in Figure 2.2. This is a significant issue as large residuals in the fits suggest that there will be large biases in the kinematics, precisely in the centermost spaxels which the supermassive black hole mass is most sensitive to. The failure to accurately fit this region suggests an issue with either the assumption of a single stellar population in the center, that

there is another spectral source, possibly the AGN, contributing to this region, or that there is some other issue pertaining to the skylines that is contributing to this region. Given these issues, we choose to exclude the region with skylines from the analysis. The remaining range to the right of the notch filter (6050 Å - 7000 Å) does not have deep stellar features that provide information for stellar kinematics. Furthermore, this region has several gas emission features with multiple kinematic components. One way of dealing with this would be to mask all of the gas. However, this would conceal some kinematic features to the left of the notch filter and would leave only a few disconnected regions of unmasked spectrum to the right of the notch filter. We thus choose to proceed by omiting the region to the right of the notch filter and fit all of the gas to the left of the notch filter. We observe some slight flux calibration issues at the ends of the spectrum which we clip for our analysis.

The SAURON observation of M87 was first described in Emsellem et al. (2004) and later reanalysed as a part of Cappellari et al. (2011a). The field of view of SAURON is $33'' \times 41''$ with a pixel size of $0\farcs94 \times 0\farcs94$. The wavelength range covered is 4800-5380 Å at 4.2 Å spectral resolution (corresponding to an instrumental dispersion of $\sigma_{\rm inst} = 108$ km s^{-1}) sampled at 1.1 Å per pixel. Four observations were made for 1800 seconds each. The data reduction was performed with XSAURON (Bacon et al., 2001). Further details are available in Emsellem et al. (2004); Cappellari et al. (2011a).

2.3.2 Photometry

We use HST imaging from Côté et al. (2004) (HST proposal 9401) to construct our stellar surface brightness model while carefully removing the AGN. This is an F850LP ACS/WFC observation covering a field of view of approximately $211'' \times 212''$ with a pixel scale of $0\farcs05$. The exposure was made for 90 seconds, guaranteeing that the nucleus does not become saturated. We also use an r-band SDSS mosaic generated with the software MONTAGE[1] to constrain the stellar surface brightness at larger radii. The SDSS image covers a spatial scale of approximately $713'' \times 713''$ with a pixel scale of $0\farcs396$.

[1]Available from `http://montage.ipac.caltech.edu/`

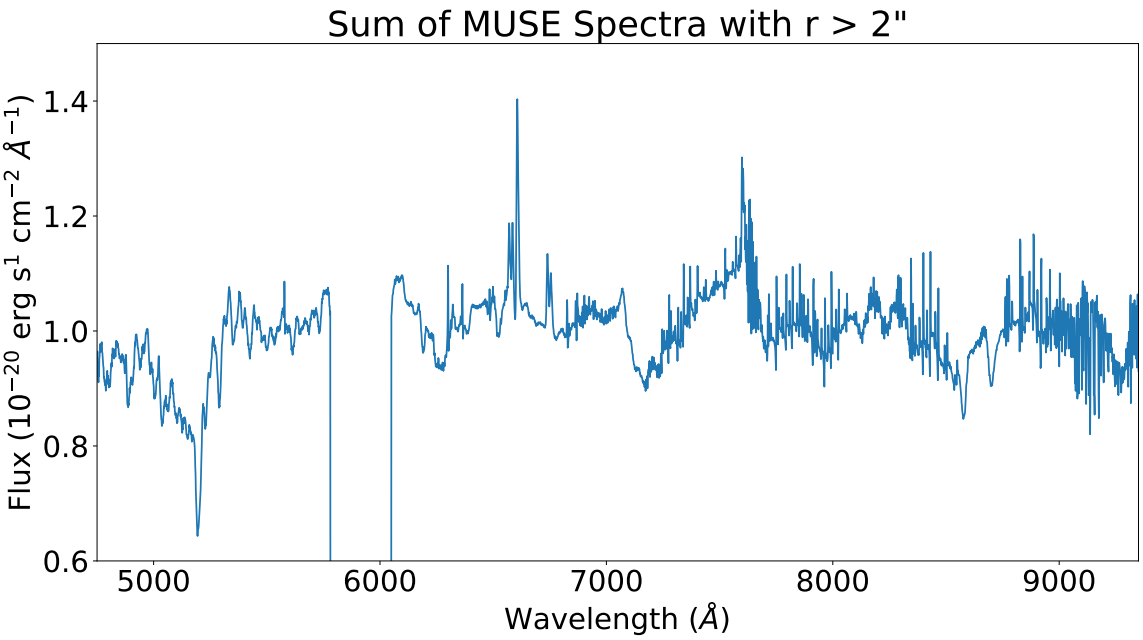

Figure 2.1: Plot of the sum of all MUSE spectra with $r > 2''$. The spectrum has significant skylines beyond around 7000 Å. The gap around 6000 Å is due to the use of sodium laser guide stars that produce an emission around 6000 Å. This is blocked with a notch filter to avoid contaminating the spectrum.

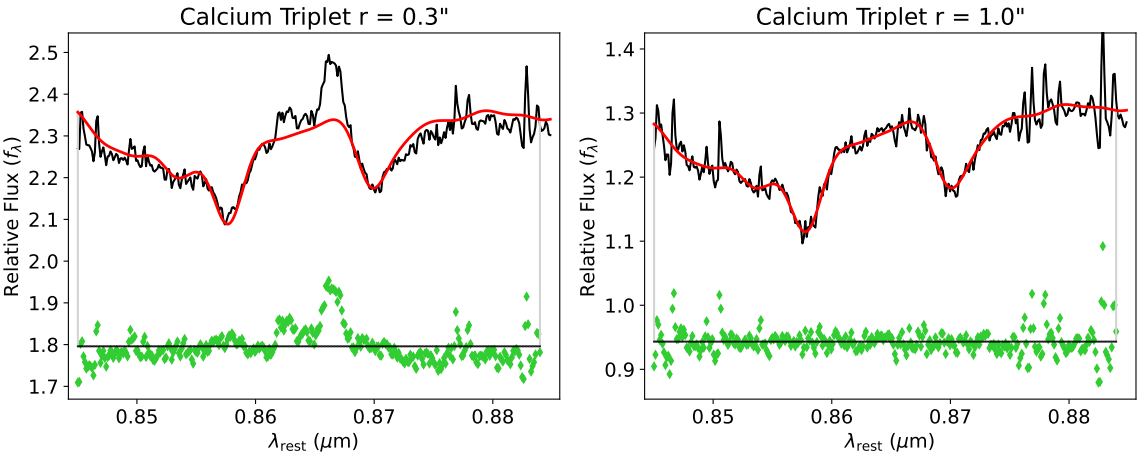

Figure 2.2: Fits to the calcium triplet region of the MUSE datacube using a single stellar template. We see that the innermost spectrum is not well described by the template spectrum between 0.86μm and 0.87μm.

2.4 Kinematic Extraction

2.4.1 Spectral Fitting

We bin the galaxy spectra spatially using the Voronoi tesselation algorithm and VorBin software package[2] described in Cappellari & Copin (2003). This algorithm takes the x and y coordinates of a set of data along with the assigned signal and noise and bins neighboring points to a target signal to noise ratio. We define the signal to be the median spectral flux and the noise to be the median error (this is done before logarithmically rebinning the spectra). For the OASIS data, we bin to a target signal to noise ratio of 50 per 1.95 Å spectral pixel. This leaves all of the spaxels in the innermost arcsecond unbinned, allowing for maximum spatial resolution. MUSE has a much higher spatial resolution, requiring some spatial binning in the innermost parts of the galaxy to have sufficient signal for kinematic fitting. We start by masking the innermost $0\farcs1$ as these spectra are entirely dominated by the AGN[3]. We tested binning with a target signal to noise ratio of 50, 40, 30, 20, 15, and 10 per 1.25 Å spectral pixel. The recovered black hole mass is consistent for 50, 40, 30, and 20, but increases sharply for lower signal to noise. This is due to the fact that with this level of noise, a flat fit to the spectra is allowed in the innermost regions, resulting in many central spectra having anomalously large dispersion. For the rest of this work, we use the case with a signal to noise ratio of 50 per 1.25 Å spectral pixel. For both datasets, we remove all data points for which the fraction of the flux due to stars is less than 50%. The motivation for this is that we find in our tests that it is around this value that the recovered kinematics become highly unreliable due to the spectra features being washed out by the AGN. This is determined after running the fits by comparing the average flux in the Legendre polynomials to the stars (more on this later).

We logarithmically resample the spectra to a velocity scale $\Delta V = c\Delta \ln \lambda$ of 105 and 66 km s^{-1} for OASIS and MUSE respectively and fit the binned spectra using the penalized Pixel Fitting method pPXF software package[4] of Cappellari & Emsellem (2004) and Cappellari (2017, 2023). This method allows for the simultaneous fitting of template stellar spectra, template gas spectra, and continuum contributions/template mismatch with the addition of additive polynomials. We can write an observed galaxy

[2]Available from `https://pypi.org/project/vorbin/`

[3]Note that we do not mask the jet. This is because the jet contributes a much smaller flux to the total spectrum and can be fit by additive Legendre polynomials

[4]Available from `https://pypi.org/project/ppxf/`

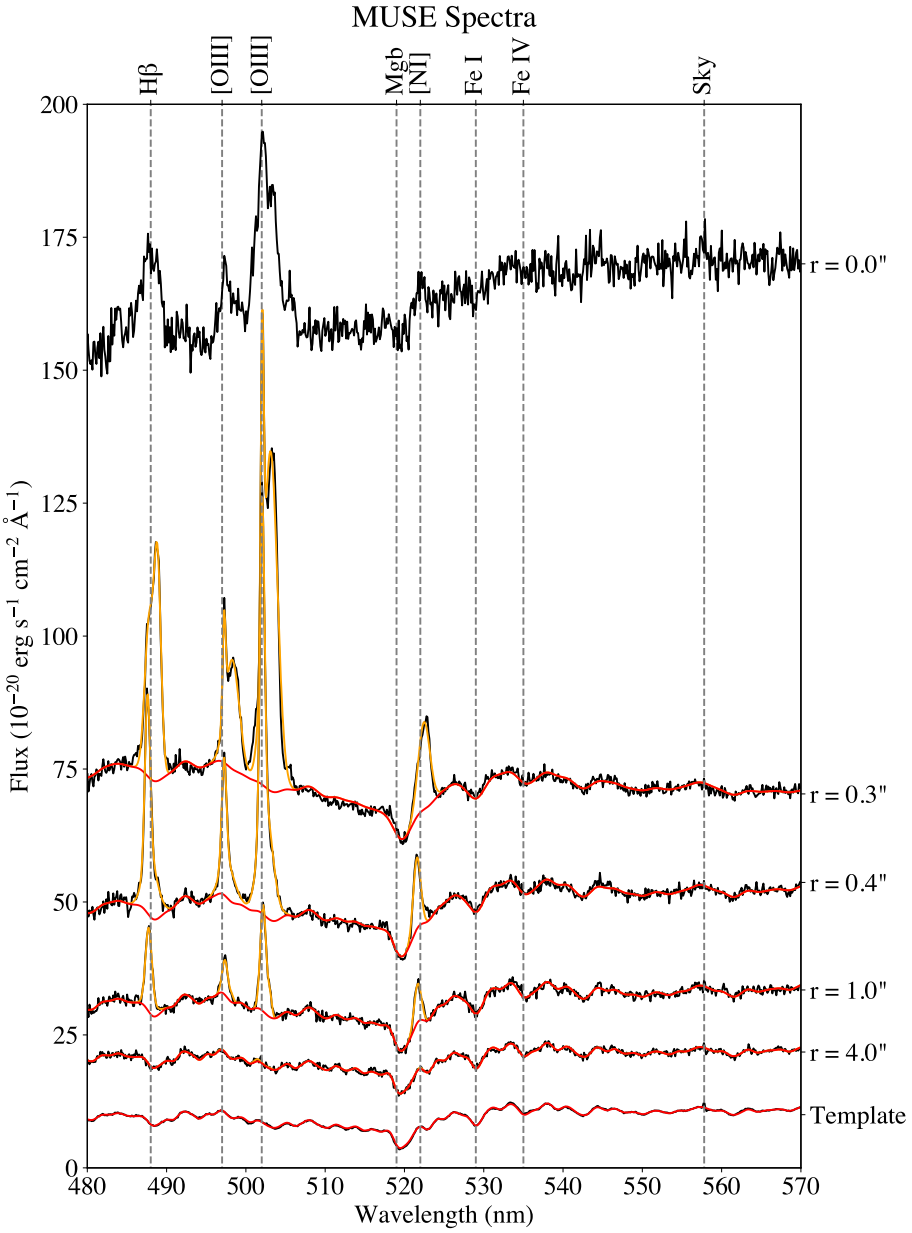

Figure 2.3: Plot of the MUSE spectra as a function of radius. The black curve is the data and the red curve is the best pPXF fit to the stars and the orange curve is the best fit to the gas. The $r = 0''$ spectra is the combined innermost $0''.1$ of MUSE. This is not fit as we do not include it in the final analysis. The double peaked structure of the gas appears intermittently at a variety of radii. The bottom spectra shows the single stellar template that we use to fit the stellar kinematics for all of the MUSE spectra, as well as the sum of the gas free spaxels that we use to fit for the single stellar template. We artificially shift and scale this spectrum for readibility. All other spectra are at their true absolute flux except for the $r = 0''$ spectrum which is significantly shifted down and compressed as it otherwise has significantly more flux than the other spectra. The mean luminosity weighted radius of the gas free spectra is $3''.3$.

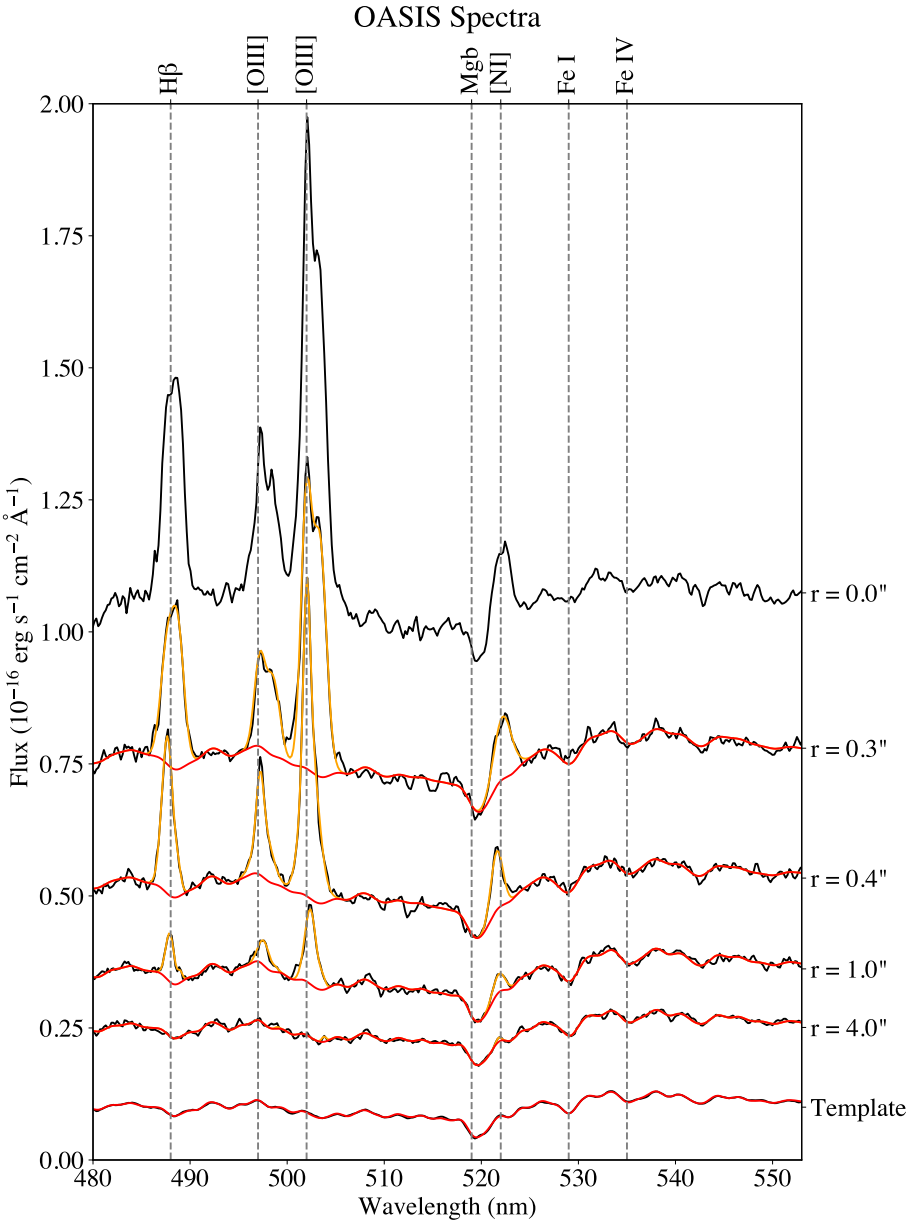

Figure 2.4: Plots of the OASIS spectra as a function of radius. The black curve is the data and the red curve is the best pPXF fit to the stars and the orange curve is the best fit to the gas. The $r = 0''$ spectra is the central spaxel for OASIS. This is not fit as we do not include it in the final analysis. The double peaked structure of the gas appears intermittently at a variety of radii. The bottom spectra shows the single stellar template that we use to fit the stellar kinematics for all of the OASIS spectra, as well as the sum of the gas free spaxels that we use to fit for the single stellar template. We artificially shift and scale this spectrum for readibility. All other spectra are at their true absolute flux. The mean luminosity weighted radius of the gas free spectra is $3\farcs7$.

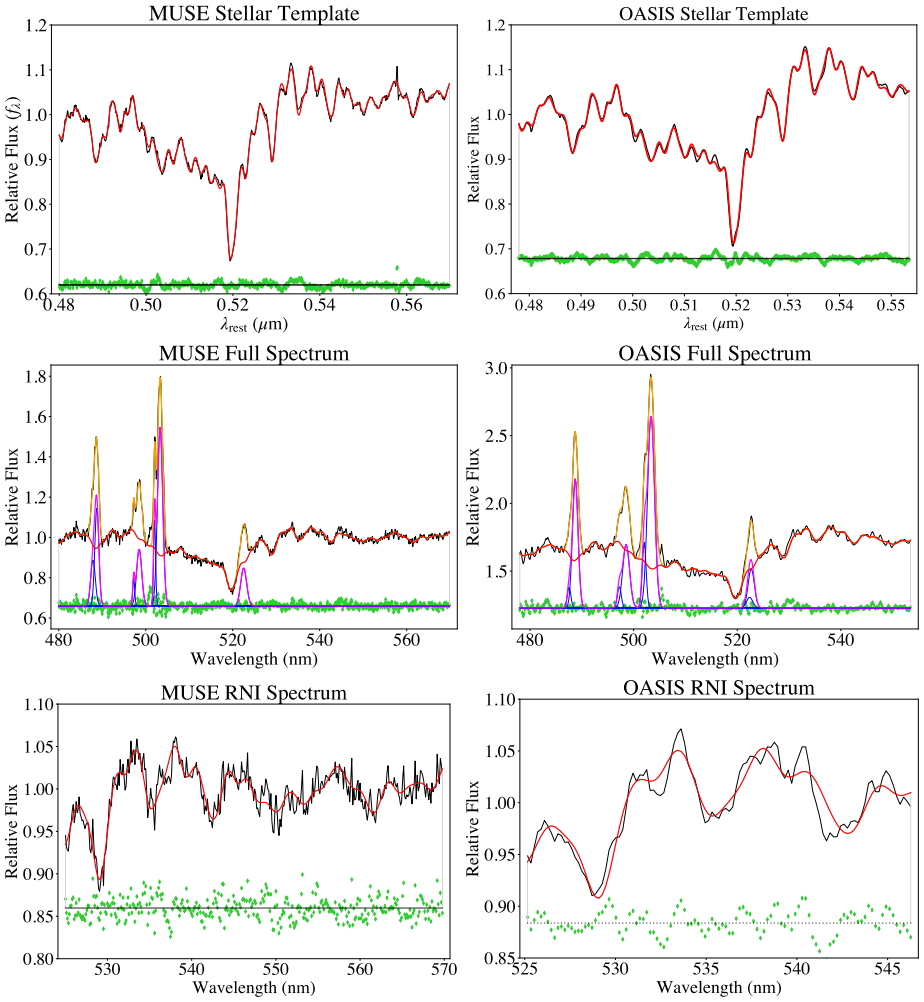

Figure 2.5: The top panels show a fit of the MILES stellar spectra to the M87 spectra at large radii without gas emissions. The black curve is the original spectrum, the red curve is the best fit stellar component, the orange curve is the best fit gas component, the pink curves are the pure gas components, and the green points are the residuals between the original spectrum and the overall bestfit model. This is done separately for MUSE and OASIS. This fitted spectra is then used to fit the stellar kinematic component for each spaxel. The middle panels show a spectrum from the centre of M87 (r = 0.6″) that has been fitted with the above stellar template, as well as with multiple gas templates. This shows strong gas emissions with a double peaked profile for Hβ and [OIII]. These must be carefully accounted for in order to produce a reliable kinematic fit. The bottom panels show a fit of the stellar template to a spectrum in the central parts of M87 (r = 0.7) where the spectra has been restricted to start at the right of [NI] (denoted RNI spectra). Here there is no gas, simplifying the fitting, but the noise in the spectrum increases.

spectrum as[5]

$$G = \sum_i w_i[T_i^s(x) * \mathcal{L}_i^s(cx)] + \sum_j u_j[T_j^g(x) * \mathcal{L}_j^g(cx)] + \sum_k b_k \mathcal{P}_k(x) \qquad (2.1)$$

Here T^s represents the stellar templates, T^g the gas templates, $\mathcal{L}$ the corresponding line of sight velocity distribution, and $\mathcal{P}$ additive polynomials (in this case taken to be Legendre polynomials).

The strongest absorption feature in the OASIS and MUSE spectra that contributes to the fits of the stellar kinematics is due to Mgb around 5200 Å. However, in the innermost arcsecond of the galaxy this feature is contaminated by gas emission from [NI]$\lambda\lambda$5197,5200 (see Figure 2.3, Figure 2.4, and Figure 2.5). When it comes to fitting the absorption feature from Mgb then, there is a degeneracy between the stellar dispersion and the gas kinematics. This is compounded by the fact that it is precisely in the innermost regions that the relative flux in each spectra due to the AGN increases, further increasing the uncertainty in the kinematic extraction. In order to test how the treatment of this affects the extracted kinematics, we consider two separate scenarios. In the first scenario we simply fit the full spectrum. This has the advantage that outside of the innermost arcsecond the kinematic extraction is very reliable with the disadvantage of the innermost arcsecond being less reliable. In the second scenario we restrict the spectra to begin at 5250 Å (these spectra are referred to as spectra to the right of [NI] or RNI spectra) so that we exclude the contaminated region. This has the benefit of avoiding any uncertainty from the fit to Mgb at the cost of further restricting the spectral range. This is done for both OASIS and MUSE. The details of the kinematic extraction for each of these scenarios is significantly different.

Spectral fitting over the full wavelength range is challenging to perform due to the presence of strong gas emission lines from Hβ, [OIII]$\lambda\lambda$4959,5007, and [NI]$\lambda\lambda$5197,5200. Furthermore, we observed multiple gas kinematic components for Hβ and [OIII] in the inner most part of the galaxy (see Figure 2.5). Additionally, the presence of the central AGN dilutes the stellar features of the central spectra making a reliable extraction of the kinematics without an assumption of the spectra shape difficult. Furthermore, since we fit the gas, there is a degeneracy between the stellar flux and the flux from the gas. In order to increase the robustness of the extracted kinematics, we allow for only one single stellar component in our fits (i.e. we fit one linear combination of template spectra to the data, as opposed to many variably weighted

[5]This ignores sky, attenuation, and multiplicative polynomials. See the full expression in eq. 13 of Cappellari (2023)

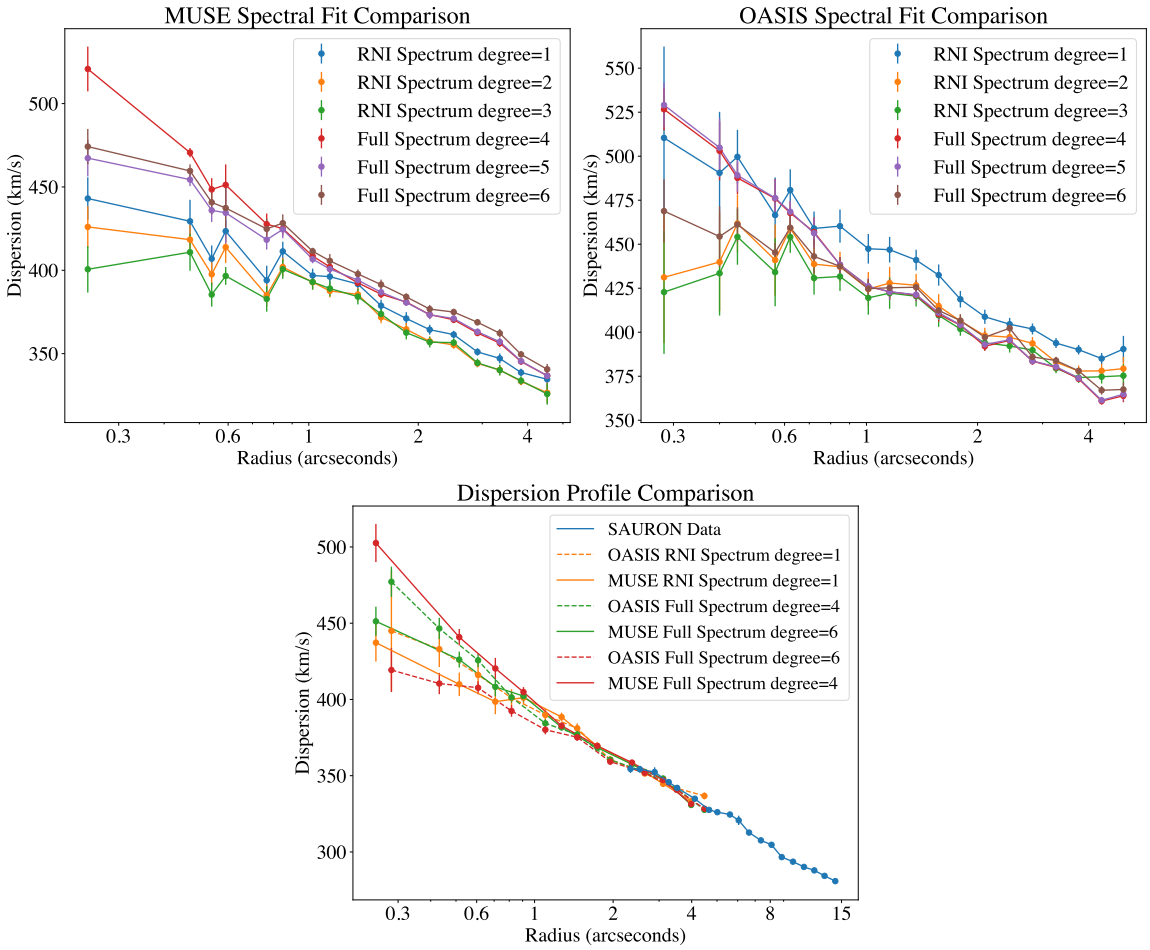

Figure 2.6: The top left and top right panels show the extracted dispersion for M87 under a number of different assumptions for MUSE and OASIS, respectively. Note that in each case the dispersion profile is consistent up to a constant scaling outside of one arcsecond. Within one arcsecond, however, there is a large deviation in the extracted profiles. This is due to the uncertainty in the kinematic extraction caused by contamination from the AGN and the degeneracy in the fits between the fit to [NI] and the Mgb absorption. The bottom panel shows the dispersion profiles used in the final analysis scaled to match the SAURON profile. Data shown in the same color are combined in the final analysis.

template spectra). We determine this single stellar template by co-adding the spectra with no gas emission lines and fitting this with stars from the MILES stellar library[6] (Sánchez-Blázquez et al., 2006; Falcón-Barroso et al., 2011). Note that one single stellar template does not mean one constant stellar population across the field of view since the Legendre polynomials allow for significant variations in the line strength of the spectral features of the stellar population (note, we know from Sarzi et al. (2018) that there are no sharp changes to the stellar population in the innermost parts of M87). We use the MILES stellar library consisting of nearly 1000 individual stellar spectra. We also tried this using SSPs from the MILES stellar library (Vazdekis et al., 2010) including synthetic spectra with alpha enhancement (Vazdekis et al., 2015) allowing for maximum possible variation in the parameters, as well as the MUSE stellar library (Ivanov et al., 2019), consisting of 35 individual stellar spectra, but we ultimately find the best fit using the MILES stars.

The choice to fit only a single stellar template as opposed to the full spectral library is very important for galaxies with an AGN. Silge et al. (2005) measured the supermassive black hole mass of Centaurus A using Gemini NIFS data, allowing for a varying stellar template and found a steeply increasing dispersion profile in the center of the galaxy and measured a black hole mass between (1.5-2.4)$\times 10^8$ $M_\odot$. This was a significant outlier of the $M-\sigma$ relation and did not agree with later measurements of the black hole mass using gas (Neumayer et al., 2007). Later, Cappellari et al. (2009) performed the same measurement with SINFONI data using a fixed single stellar template and found a black hole mass equal to 5.5×10^7 $M_\odot$. This was no longer a significant outlier of the $M-\sigma$ relationship and is in excellent agreement with the determination of the black hole mass from gas (Neumayer et al., 2007). The case where there is no AGN was tested in Westfall et al. (2019). There, the authors show in figure 13 that the difference in the extracted kinematics for high mass red galaxies between those using a single or varying template is at most 3 per cent. This is in contrast to the case of blue galaxies for which there is a much stronger deviation, likely due to the fact that these galaxies are undergoing star formation and thus have stronger population gradients. We also note that, while we independently determine the single stellar template for both MUSE and OASIS, the two agree quite closely (less than one percent RMS deviation over the relevant wavelength range). Using the OASIS template to fit the MUSE data and vice versa returns effectively the same kinematics as just using the OASIS template for OASIS and the MUSE template for MUSE.

[6]Available from `http://miles.iac.es/`

The OASIS spectral resolution is much lower than that of the stars in the MILES stellar library which have an instrumental resolution of 2.51 Å FWHM corresponding to $\sigma_{\rm instr} \approx 61$ km s^{-1} at 520nm. We account for this by degrading the resolution of the template spectra (before logarithmically rebinning) with a gaussian to a constant resolution per angstrom so that the spectral resolution for the two are the same. The spectral resolution for MUSE over the wavelength range we use is nearly the same as that of the MILES stars, so we do not apply a correction. It is worth emphasising that we treat this line spread function (LSF) as constant over the whole detector. This is generally not the case, as different instruments can have significant non-gaussian shapes for the LSF, as well as variations across the detector (Hau, 2017, Fig.22-25). The LSF adds to the true dispersion of a spectrum in quadrature, i.e. $\sigma_{\rm new} = \sqrt{\sigma^2_{\rm kin} + \sigma^2_{\rm LSF}}$. Given that M87 is a very massive galaxy and, across our field of view, has a large dispersion typically greather than 300 km s^{-1}, we can estimate that a factor of 2 uncertainty in the LSF will at most lead to a 6 per-cent uncertainty in the dispersion. This is smaller than our other sources of systematic error (choice of additive Legendre polynomial degree), and thus we proceed assuming a constant LSF. We determine the single stellar template separately for both OASIS and MUSE. We then allow gas templates in the pPXF fits for $H\beta$, [OIII] and [NI], but allow Hβ and [OIII] to have two distinct kinematic components each. Given the number of templates (1 stellar spectrum + 2 Hβ + 2 [OIII] + 2 [NI] spectra) and the fact that the gas could be challenging to fit due to the possibility of there being multiple local best fits, we experimented with a number of constraints, such as treating the gas components of Hβ and [OIII] as a part of the same kinematic component by fixing their velocity and dispersion to be equal. Ultimately, we found that the most reliable fit to the stellar spectra is comes from allowing maximum freedom in the gas fit. That is, treating each gas template as having its own velocity and dispersion. This is because even slight offsets in the gas fits for central spectra result in large residuals that end up driving the stellar fit. This means having a total of six kinematic components: one for the stars, one for each set of Hβ and [OIII], and one for [NI].

We assume that the line of sight distribution can be treated as a Gaussian without the inclusion of higher order Gauss-Hermite moments. We made this assumption because (i) Cappellari et al. (2007, sec. 2) found using synthetic galaxy models that the sigma obtained from a Gaussian fit (moments=2) with pPXF provides a better approximation to the second velocity moments than computing of the second moment by integrating the LOSVD from a fit which includes higher Gauss-Hermite moments (e.g. moments=4); (ii) making this assumption one is able to produce $V_{\rm rms}$ maps with

JAM that visually provide an excellent match to the observed $V_{\rm rms}$ of hundreds of real galaxies ((e.g. Cappellari et al., 2015, fig. 1); and (iii) in the specific case of the stellar kinematics of M87, previous studies have found that, within the range of our data, h_4 is within ± 0.03 (Liepold et al., 2023a, fig. D2), implying that the LOSVD is essentially Gaussian. To test this, we run pPXF for our MUSE data where we include Gauss-Hermite moments up to h_4. We find evidence for some offset in h_3 (median between 0.015 and 0.018 depending on the choice of Legendre Polynomial) which suggests that there is some small template mismatch. We find that h_4 is randomly scattered around zero, suggesting that the non-zero values of h_4 are a manifestation of the underlying noise in the fits rather than tracing a real trend in the data (see Figure 2.7). This is consistent with the results of Liepold et al. (2023a).This holds across the choice of Legendre Polynomials. Additionally, our results from chapter 4 suggest that in general for recovering the second moment it is optimal to not include higher order GH moments due to the sensitivity of the second moment to the wings of the LOSVD. Thus we feel confident not including the Gauss-Hermite moments in our analysis.

One issue that we faced was some of the fits returning large stellar velocities in a couple of the central spaxels of the galaxy. M87 is well known to be a slow rotator (Emsellem et al., 2007, 2014), so any large deviations in the velocity suggest an error in the fitting. We handle this by fixing the velocity across the field for both datasets to equal the recession velocity[7]. This serves as a realistic prior that helps decrease the noise in our kinematic extraction. In order to confirm that this does not impact the extraction of the dispersion, we compared the dispersion measured before and after fixing the velocity. We found that, for both MUSE and OASIS, this did not meaningfully impact the dispersion over the field of view (less than one percent RMS deviation between the two) and the difference for the central spaxels is less than $\sim$2 per cent. It is worth noting that the SAURON data we use does not fix the velocity at any point over the field of view. Given that we exclude SAURON data from the central regions and we know that fixing the velocity in the center does not significantly impact the dispersion at large radii, we believe that it is consistent to do this.

Previous work has not run into this issue due to differences in the spectral range considered. Gebhardt et al. (2011) uses data covering the CO band head, which features several deep absorption structures ideal for determining stellar kinematics. Emsellem et al. (2014) uses MUSE data in wide field mode without AO which does not

[7]Note that Emsellem et al. (2014) finds evidence for a kinematically decoupled core with rotation velocity ± 5 km s^{-1}. This is within the errors of this analysis.

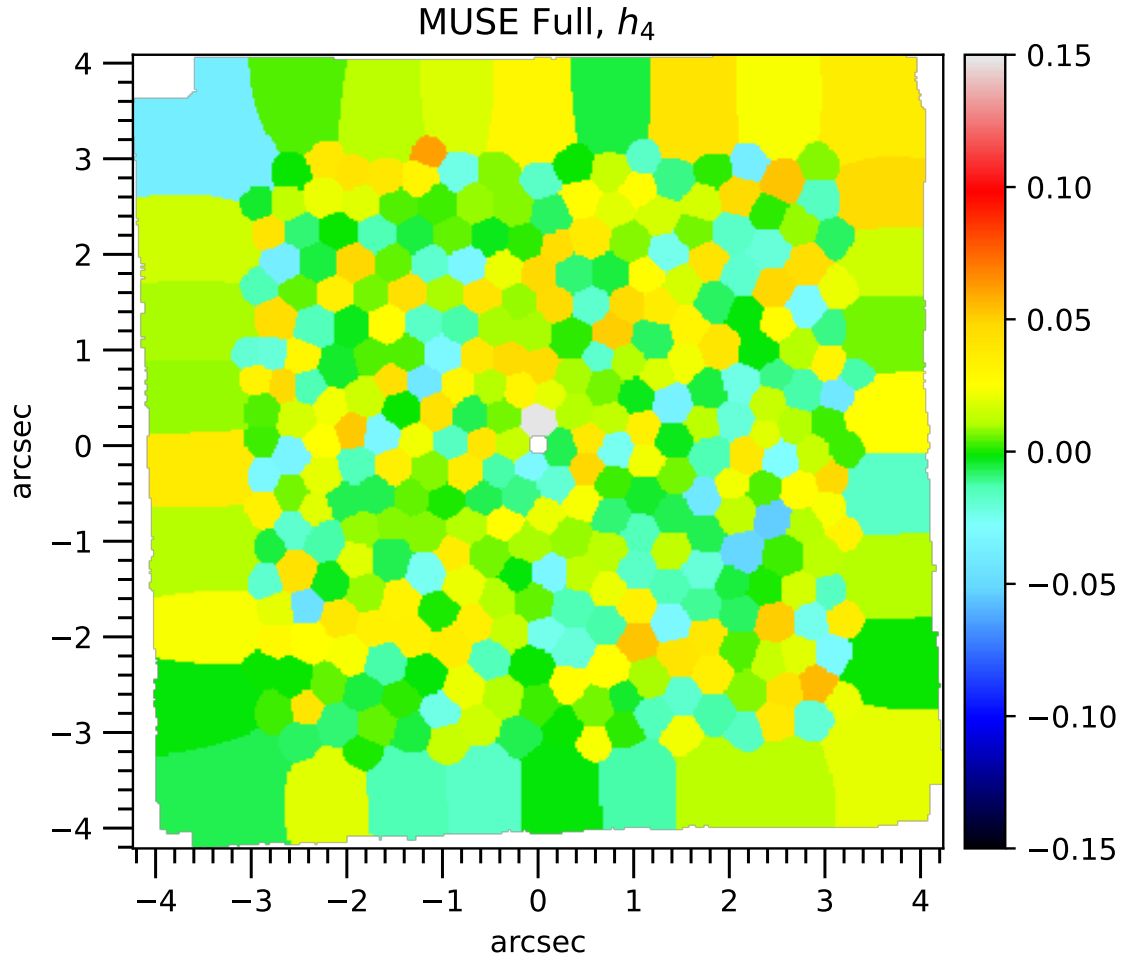

Figure 2.7: Map of h_4 measured by running pPXF on the Full MUSE spectra with Legendre polynomial degree of 4 binned to a signal to noise ratio of 50. The values are centered on zero with some scatter. The results are unchanged under different Legendre Polynomial assumptions.

feature the notch filter designed to block light from laser guide stars used in AO and thus can see the absorption due to the sodium doublet around 5900 Å. Additionally, Liepold et al. (2023a) uses much bluer data with spectral features such as CA II H, CA II K in addition to Mgb that add valuable kinematic information.

We calculate errors in the dispersion using wild bootstrapping (Davidson & Flachaire, 2008). Wild bootstrapping works by first generating a model spectrum by performing a pPXF fit to the initial spectrum, and then randomly generating mock spectra by randomly adding (or subtracting) the residuals between the inital model and spectrum at each spectral pixel in the model spectrum (this is in contrast to standard bootstrapping where the residuals are shuffled). Each of these mock spectra is then fit and the standard deviation of the recovered dispersion is treated as the uncertainty. We generate 100 mock spectra for each spaxel to determine the bootstrapped error in the dispersion. Lastly, we perform this analysis three separate times using Legendre polynomials of degrees 4, 5, and 6. We find that the extracted dispersion in the centre most region depends on the choice of Legendre polynomial (see Figure 2.6). This is likely due to the degeneracy between [NI] and Mgb absorption. Different degrees of Legendre polynomial, especially those with very high degree given the total wavelength range, can go beyond accounting for template mismatch and can start reproducing parts of the stellar spectrum.

Kinematic extraction redwards of [NI] is done using the same single stellar template as in the case for the full spectra, no gas templates, and additive Legendre polynomials of degree 1, 2, and 3. As in the previous case, we also fix the velocity of each spectra to the recession velocity, which we determined as the median over the field of a free fit. A dispersion map of the extracted kinematics in the case where we fit the full spectrum with additive Legendre polynomials of degree 4 is shown in Figure 2.8. The map appears symmetric, as we expect since the core of M87 is highly spherical. As a result, we plot the dispersion profile for the remaining scenarios as a function of radius in Figure 2.6.

The supermassive black hole in M87 has the largest angular size from Earth of any known black hole outside of the Milky Way. We define the sphere of influence $r_{\rm BH}$ as

$$M_{\rm BH} = M^*(< r_{\rm BH}) \tag{2.2}$$

i.e. the radius such that the black hole mass equals the mass in stars. Assuming the range of black hole masses and M/L values determined in this work, we determine the $r_{\rm BH}$ is between $\sim$5 and 6″and possibly even larger. Thus the black hole sphere of influence dominates much of the field of view for both OASIS and MUSE. This makes

measuring parameters such as the stellar mass to light ratio challenging as, within this field of view, there is a large degeneracy between the black hole mass and the stellar mass. We can break this degeneracy by adding larger field kinematic data that is more sensitive to the mass of the stars. We do this by including SAURON data (Emsellem et al., 2004) for M87 as reanalysed for the ATLAS^{3D} project[8] by Cappellari et al. (2011a). Here the spaxels were binned to a target signal to noise ratio of 40 (as opposed to 60 in the original SAURON reduction). The gas was masked during the fits and the degree of additive Legendre polynomials was set to 4. Similar to this work, the stellar kinematics were fit using the MILES stellar library with a single optimal linear combination of templates.

When considering large field data, it is important to be conscious of the fact that any subsequent parameter studies will be heavily influenced by information provided at larger radii because there is a greater volume of data points at large radius than at small radius. To account for this we only consider SAURON data out to a radius of $15''$. To avoid accounting for uncertainties in the SAURON PSF, we do not include any SAURON data within the innermost $2''$of the galaxy. We observe an offset between our extracted dispersions and the SAURON dispersion. This is expected due to systematic uncertainties in the kinematic extraction, especially due to template mismatch. To account for this, we apply a multiplicative scale to the MUSE/OASIS dispersions so that they match the SAURON data between $2''$ and $4''$. Scaling the velocity axis by a factor $\sqrt{\Upsilon}$ is equivalent to changing the overall mass normalization of the model by a factor Υ. For MUSE, the factor required to scale the dispersion to match the SAURON data varies between 0.95 and 1 (given the choice of Legendre polynomial). For OASIS, the range is between 0.87 and 0.9. If one wanted to scale each data set to either the OASIS or MUSE data, this would correspond to an increase in the black hole mass of between $\sim 20 - 25$ per-cent for OASIS and up to ~ 10 per-cent for MUSE.

We see that the recovered kinematics appear to depend strongly on the additive Legendre polynomial degree used. There is no strong criterion to determine which of these are reliable and which are unreliable. As a result, we choose to study a subset of them that samples the reasonable range of different recovered shapes in the dispersion profile. We rule out using the degree 2 and 3 RNI spectra as we expect that using such a high degree of Legendre polynomial over such a small wavelength range will lead to an unreliable kinematic extraction (indeed, for both MUSE and OASIS, we see the dispersion profile either completely flattening out or decreasing in

[8]Available from `https://purl.org/atlas3d`

the centre). For the degree 4,5,6 polynomials note that the degree 5 polynomial for MUSE closely resembles the degree 6, and for OASIS the degree 5 dispersion profile closely resembles the degree 4 dispersion profile. As such, we choose to throw out the degree 5 profile from each data set and are left with 6 dispersion profiles as shown in the bottom panel of Figure 2.6. In the final analysis, we combine different data sets in order to draw a reliable result. We choose the combinations:

1. SAURON + OASIS degree 6 + MUSE degree 4
2. SAURON + OASIS degree 4 + MUSE degree 6
3. SAURON + OASIS RNI degree 1 + MUSE RNI degree 1

In the latter case we refer to the kinematics as the RNI spectra.

2.4.2 PSF from AGN Spectrum

The PSF from IFS data has often been determined by comparing convolved HST photometry to the flux from the IFS cube (e.g. McDermid et al., 2006; Krajnović et al., 2018, fig.B1). This method is most reliable when the centre of the galaxy has a cusp since the observed profile will be more strongly affected by the PSF. Since the nuclear region of M87 has a core, we expect this method to produce a less reliable PSF. However, we can circumvent this by noting that M87 has a bright AGN that can be assumed to be unresolved and can be used as a point source to infer the PSF directly. The profile of the PSF can thus be measured if we can extract out the component of each spectrum due to the AGN. This can be done using the additive polynomials in our spectral fitting. The AGN likely has a flat non-thermal continuum that can be well approximated by polynomials while the underlying galaxy does not. We can then integrate the flux from the AGN, parameterised as these additive Legendre polynomials, across the full wavelength range and thus determine the shape of the unresolved AGN.

We extract the shape of the PSF by performing a pPXF fit to the unbinned central spaxels in each dataset over the largest possible wavelength range (4800-5700Å for MUSE and for 4760-5558Å OASIS) using the fixed stellar template determined before from the gas-free and AGN-free spectra with multiple gas components and additive Legendre polynomials. We also tried restricting the wavelength range to that of the RNI data but found that we are unable to obtain reliable results due to the spectral range lacking distinct features in the center to anchor the total weight of the stars compared to the AGN. Additionally, we experimented with fixing the kinematics

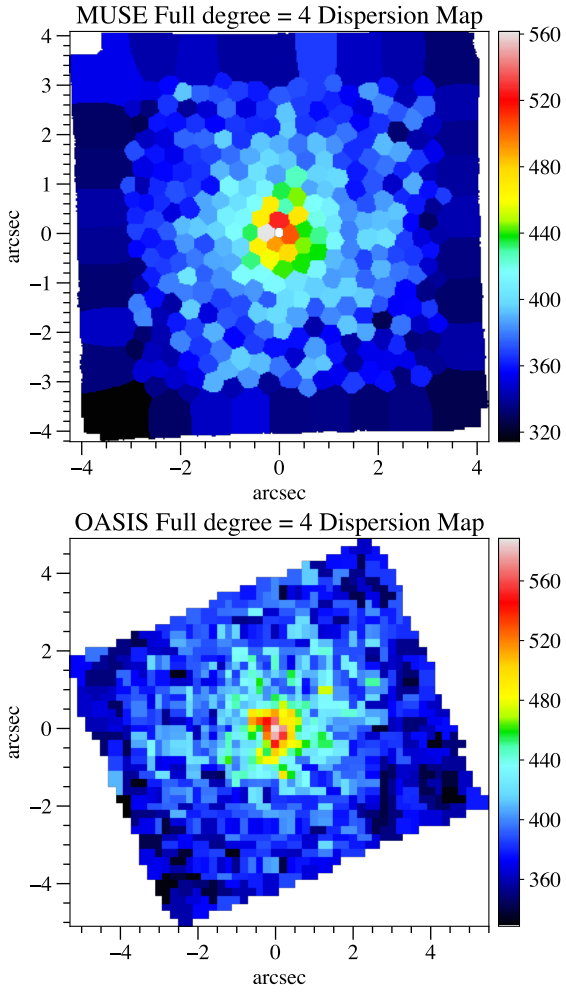

Figure 2.8: Top panel: MUSE kinematics extracted over the full spectrum with additive Legendre polynomial degree of 4. The centremost 0.1″ is masked (this is the circle in white at the origin). Bottom panel: OASIS kinematics extracted over the full spectrum with additive Legendre polynomial degree of 4. Both are oriented so that north is facing up and east is facing to the left.

though we found that this did not impact the results. We allow the degree of the additive Legendre polynomials to vary from 1 to 6. The FWHM of the PSF varies between 0.″042 to 0.″061 for MUSE. For OASIS this range is 0.″527 to 0.″586. In this case we do not mask any of the central spaxels. We also test the effects of masking the jet but find only small diferences to the extracted FWHM. We adopt the PSF in the degree 4 case for both the MUSE and OASIS data. It is worth noting that, in the case of MUSE, we do not expect the uncertainty in the PSF to impact the final results since we mask an inner region that is about the same size as the PSF. In Figure 2.9, we show the results for MUSE and OASIS using degree 4 Legendre polynomials, as well as a plot of the one dimensional profile as a function of radius to the left or right of the origin.

As we see from Figure 2.9, the shape of the PSF for both MUSE and OASIS visually appears close to axisymmetric. We parametrize the PSF using a multi-gaussian expansion in the form of

$$\mathrm{PSF(R)} = \sum_{\mathrm{i=1}}^{\mathrm{Q}} \frac{\mathrm{G_i}}{2\pi\sigma_\mathrm{i}^2} \exp\left(\frac{-\mathrm{R}^2}{2\sigma_\mathrm{i}^2}\right) \tag{2.3}$$

with R the radius, σ_i the standard deviation of gaussian i, and G_i the normalization of each gaussian satisfying $\sum_{i=1}^{Q} G_i = 1$. The choice of an MGE parameterisation is very convenient as it allows parts of the dynamical modelling to be performed analytically. Additionally, the freedom of the MGE parameterisation means it can describe a wide range of axisymmetric shapes, including circular Airy or Moffat profiles. We model the observed AGN profile by integrating the PSF over the spaxel size of OASIS/MUSE. We do this by first noting that a PSF convolved observable $S_{\mathrm{obs}}(x, y)$ can be written as (e.g. Qian et al., 1995, appendix. D)

$$S_{\mathrm{obs}}(x, y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} S(x, y)K(x - x', y - y')dxdy \tag{2.4}$$

with

$$\begin{aligned} K(x, y) = \sum_{i=1}^{Q} \frac{G_i}{4} &\left[\mathrm{erf}\left(\frac{\mathrm{L_x}/2 - \mathrm{x}}{\sqrt{2}\sigma_\mathrm{i}}\right) + \mathrm{erf}\left(\frac{\mathrm{L_x}/2 + \mathrm{x}}{\sqrt{2}\sigma_\mathrm{i}}\right)\right] \\ &\times \left[\mathrm{erf}\left(\frac{\mathrm{L_y}/2 - \mathrm{y}}{\sqrt{2}\sigma_\mathrm{i}}\right) + \mathrm{erf}\left(\frac{\mathrm{L_y}/2 + \mathrm{y}}{\sqrt{2}\sigma_\mathrm{i}}\right)\right] \end{aligned} \tag{2.5}$$

Note that K is the analytic expression of a gaussian integrated over a spaxel of size L_x by L_y centred on the point (x, y). As our observable is unresolved, we can treat it

Table 2.1: Table of fitted MGE parameters for OASIS, MUSE, and the HST F850LP TinyTim PSF, respectively. The FWHM for OASIS is 0.″561, for MUSE is 0.″049, and for HST is 0.″063. The HST PSF is larger than the MUSE PSF because the wavelength observed in the HST observation is longer than that in the MUSE observation.

Number	G_i	σ_i (arcsec)
1	0.515	0.215
2	0.485	0.423
1	0.00769	0.00793
2	0.20493	0.03595
3	0.59847	0.14573
4	0.18891	0.79289
1	0.34266	0.02454
2	0.36807	0.08022
3	0.13923	0.09772
4	0.04820	0.21524
5	0.10184	0.47409

as a delta function. Substituting that into Equation 2.4 gives that the model of the PSF is simply $K(x, y)$, with the lenslet size substituted for that of MUSE or OASIS.

We then fit the PSF parameters by matching this model to the observed spectrally determined AGN by employing the optimize.least_squares function of SciPy (Virtanen et al., 2020). We assume the errors are constant except in the centre, where we set them to be small so as to force a good fit at both large and small radii. We use the PSF extracted with degree 4 Legendre polynomials for the rest of the analysis. The measured parameters for each PSF is given in Table 2.1. The FWHM for OASIS is 0.″561, which is consistent with the seeing on the night of the observation (McDermid et al., 2006). The FWHM for MUSE is 0.″049, which is close to diffraction limited. We see in Figure 2.9 that the PSF fit with a MGE provides a visually good fit to the data.

2.5 Photometry and Mass Modeling

2.5.1 Stellar Tracer Distribution

Modelling the stellar tracer distribution of M87 is challenging due to non-stellar contributions in the photometric data, namely the AGN and jet. It is sufficient to mask the jet, but masking the AGN would mean masking the location where the supermassive black hole is. This is the most important region to model for black hole studies, implying that we must take another approach. The way that we circumvent

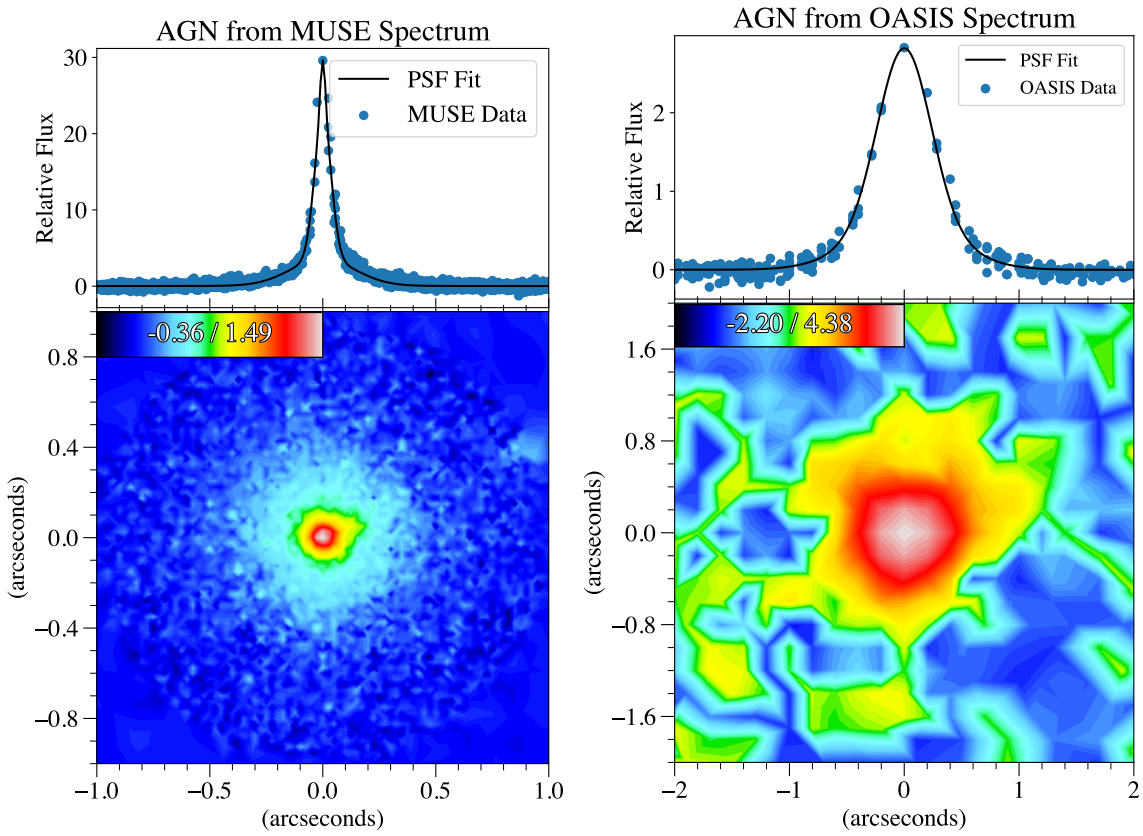

Figure 2.9: The top left and right figures show a plot of the spectroscopic data of the AGN with the best fit multi-gaussian expansion of the PSF after being integrated over the MUSE/OASIS spaxel. The data is measured by integrating the additive Legendre polynomials (our proxy for the AGN continuum) over the full spectral range. This is presented as a one dimensional function where the x-axis is the radius of each data point, with those lying to the left of the origin having $x < 0$ and those lying to the right having $x > 0$. The colormap shows the 2D image of the same data but on a symlog scale with linear threshold equal to 1 for MUSE and .1 for OASIS. These results show the reliability of using the AGN contribution to the spectra as measurement of the PSF.

this is by using the information from our spectral fits. Spectral fitting determines the contribution of stars, gas, and the AGN (through additive polynomials) such that one can extract out the pure stellar contribution.

To do this, we measure the flux due to the stars in the pPXF fit to the MUSE data from section 3.2. We do this by subtracting from each spectra the best fit gas lines and additive Legendre polynomials (which approximate the AGN spectrum). We test this using Legendre polynomial degrees between 1 and 6. The main difficulty is in the central spaxels where some of the fits of the stellar spectrum become so diluted that they are completely degenerate with the Legendre polynomials. In order to test this, we ran one set of fits for each polynomial degree and with the kinematics fixed to the largest value in the bottom panel of Figure 2.6, as well as the smallest. We show the results of this in Figure 2.10. Here you can see that, while some of the choices of polynomial degree lead to decreasing stellar densities in the center of the galaxy, they all agree with the profile of Kormendy et al. (2009) beyond 1 arcsecond, and the results among the different choices of Legendre polynomials are roughly consistent down to 0.2 arcseconds. Beyond this there is more scatter as the degeneracy between the stellar component and additive Legendre polynomials becomes very pronounced due to the high dispersion of the stars. The profiles approach an upper bound which we take to be the true stellar density. We ignore those profiles with decreasing inner stellar density as this behavior is not observed in other core galaxies either observationally or in simulations. Additionally, the profiles with the largest surface brightness in the center appear to naturally interpolate from the region around 0.2 arcseconds to the center. We fit the profile over the region of the MUSE data using a double power law and match it to the radial surface brightness from Hubble outside of the innermost $0.5''$. From this we create a modified Hubble image which has the innermost arcsecond replaced with our fitted AGN-free profile.

We parametrize the galaxy surface brightness using the Multi-Gaussian Expansion method (Emsellem et al., 1994; Cappellari, 2002). In order to model the full extent of the galaxy, we match an SDSS r-band mosiac of M87 to the Hubble image and fit them simultaneously using the robust MGE fitting algorithm and the MgeFit[9] software package of Cappellari (2002). This algorithm fits the projected surface brightness using a multi-gaussian expansion of the form

$$\Sigma(x', y') = \sum_{j=1}^{N} I_j \exp\left[-\frac{1}{2\sigma_j'^2}\left(x_j'^2 + \frac{y'^2}{q_j'^2}\right)\right] \tag{2.6}$$

[9]Available from `https://pypi.org/project/mgefit/`

Where x' and y' are the projected locations on the sky, and q'_j, I_j, and σ'_j are the axial ratio, the maximum surface brightness, and the standard deviation of the jth gaussian, respectively. We mask the jet and gap between the detectors in the HST image, and we mask a prominent star in the SDSS image. We also provide MgeFit with the Hubble ACS/WFC PSF in order to obtain the PSF-deconvolved stellar distribution as opposed to the observed distribution. We generate this Hubble PSF using the tool TinyTim (Krist et al., 2010). We record our MGE expansion for this in Table 2.1. The MGE/galaxy contours are shown in Figure 2.11 and the parameters we fit are shown in Table 2.2. We express the MGE in the AB photometric system and the F850LP band of HST where we have used the absolute magnitude of the sun $M_{\odot,\mathrm{F850LP}}$=4.50 mag from Willmer (2018) and the galactic extinction $A = 0.029$ mag from Schlafly & Finkbeiner (2011). The zero point at the time of the observation was $ZP = 24.873$.

Our spectral-decomposition approach allows for a much more reliable measurement of the stellar surface brightness near the black hole than was possible before. The most recent stellar dynamical determinations of the black hole mass in M87 (Gebhardt & Thomas, 2009; Gebhardt et al., 2011; Liepold et al., 2023a) have used the surface density profile provided in Kormendy et al. (2009). This work determines the surface density profile of M87 by combining observations across a number of different distance scales and photometric bands. The data determining the central profile of M87 comes from Lauer et al. (1992), in which the authors fit a power law starlight model, central nonthermal point source, and optical counterparts for the jet knots in order to determine the shape of the stellar distribution. Kormendy et al. (2009) notes that there are some gradients and discontinuities in the ellipticity and principal axis inside the core of M87 which may be due to issues with the AGN treatment in Lauer et al. (1992). We find a much flatter core in the innermost region which, holding everything else constant, will result in a larger black hole mass. This is discussed in more detail later.

2.5.2 Mass Modelling with M/L Gradients

In order to perform dynamical modelling, one has to parametrize the gravitational potential. Sarzi et al. (2018) used stellar population models to measure gradients in the stellar M^*/L in M87 by allowing for both ages, metallicity and stellar initial mass function (IMF) variations. Although these measurements are assumption dependent and quite uncertain, they provide an estimate for a possible stellar M^*/L variation within the innermost 30-40″ of M87, implying that it is not sufficient to assume that

mass follows light without testing the alternative. This is further corroborated by Oldham & Auger (2018) which also finds evidence for a radially decreasing M^*/L. We allow for variations away from mass follows light in this region by allowing for a radially dependent M^*/L profile given by

$$\left(\frac{M^*}{L}\right)(r) = \begin{cases} (M/L)_1 & r < 1'' \\ \left[\left(\frac{M}{L}\right)_1 - \left(\frac{M}{L}\right)_2\right] \frac{\lg(r/30)}{\lg(1/30)} + \left(\frac{M}{L}\right)_2 & 1'' \leq r \leq 30'' \\ (M/L)_2 & r > 30'' \end{cases} \tag{2.7}$$

This functional form is motivated by the fact that, within the error bars, figure 11 of Sarzi et al. (2018) is well fit by this function. The observed nearly linear variation in M^*/L with $\lg(r)$ cannot represent the true M^*/L variation since it is unbounded at 0 and infinity, so we set it to be constant outside of the regions constrained by the data. We implement this within the MGE formalism in the following way: we evaluate $(M^*/L)(r)$ at each σ_j (where σ_j is the standard deviation of each gaussian in the MGE) and multiply the surface brightness $I_{F850LP,j}$ by this value. In principle, $(M^*/L)_1$ and $(M^*/L)_2$ are free parameters, but we can determine possible upper and lower bounds by studying the effect of varying the IMF. Sarzi et al. (2018) finds that the largest possible M/L ratio assuming a Kroupa IMF is close to 5.0 in the r-band. From figure 2 of Cappellari et al. (2012), we see that, empirically in the galaxy population, the largest M^*/L increase one can expect due to the IMF normalization is a factor 2.6 heavier than the value corresponding to a Kroupa IMF. Taking the r-band M^*/L with Kroupa IMF as reference, the heaviest M^*/L one can realistically expect for M87, allowing for extreme IMF gradients, correspond to M^*/L=13 in the r-band. Lastly, we can convert this to the SDSS z-band (which we take to be approximately the same as the ACS/WFC F850LP band) using the conversion formula

$$M/L_z = M/L_r \times 10^{0.4([z-r]_{\rm M87} - [z-r]_\odot)} \tag{2.8}$$

Using M87 colors $z = 9.92$ and $r = 10.70$ from SDSS DR7, and solar magnitudes $z_\odot = 4.50$ and $r_\odot = 4.65$ from Willmer (2018), we find that the upper bound is $(M^*/L)_z \lesssim 7.23$. This becomes important later in the analysis where we introduce this cut off when our modelling would otherwise prefer an unphysically large M^*/L ratio.

This choice of parametrization restricts the steepest possible M^*/L variation. One could further increase the freedom of this parametrization by allowing the innermost radius at which the M^*/L ratio becomes constant to vary. We discuss the impact of this in subsection 2.7.2.

2.5.3 Dark Matter

Several studies of the dark matter halo of M87 have been previously made using globular clusters Murphy et al. (2011); Agnello et al. (2014); Oldham & Auger (2016); Li et al. (2020). Oldham & Auger (2016) finds a preference for a somewhat cored halo, whereas Murphy et al. (2011); Li et al. (2020) does not find a preference for either a cored profile or a cuspy halo (though Murphy et al. (2011) has a slight preference for a cored halo). Agnello et al. (2014) finds a preference for a very steep cusp, but this may be due to choices in their modelling (See Oldham & Auger (2016) for an extensive discussion). Our data only covers up to $15''$, which is about a fifth of a half light radius, so we cannot make any reliable inference about the dark matter contribution. We include dark matter and vary M/L in our study to account for the possible range of shapes of the total density profiles, which leads to a more conservative estimate of the black hole mass. As such, we proceed with a NFW dark matter halo (Navarro et al., 1997)

$$\rho(r) = \frac{\rho_0}{(r/r_s)(1 + r/r_s)^2} \tag{2.9}$$

The range of our data is much less than r_s so we set it arbitrarily to 20kpc because its precise value is irrelevant for the modelling results. We parametrize the overall magnitude as the fraction of matter within a sphere of one half light radius consisting of dark matter, $f_{\rm dm}$. We calculate the enclosed masses from the MGE analytically using the routine MGE_RADIAL_MASS from JAMPY[10]. We calculate the half light radius from the MGE of Table 2.2 using the routine MGE_HALF_LIGHT_RADIUS in JAMPY and find a value of $70''$.

2.6 Dynamical Modeling

2.6.1 Jeans Modelling

The Jeans Anisotropic Modelling (JAM) method (Cappellari, 2008, 2020) has been used to model the stellar dynamics of galaxies and study their stellar mass-to-light ratios and dark matter content in large integral-field spectroscopic surveys such as ATLAS^{3D} (Cappellari et al., 2013), SAMI (Scott et al., 2015), MaNGA (Li et al., 2018a) as well as surveys at high redshift e.g. for the LEGA-C survey (van Houdt et al., 2021). It was applied to the study of galaxies' total density profiles out to large radii (Cappellari et al., 2015) and in several studies of smaller galaxy samples. More recently, JAM was employed to accurately predict Gaia kinematics of the Milky Way

[10]Available from `https://pypi.org/project/jampy/`

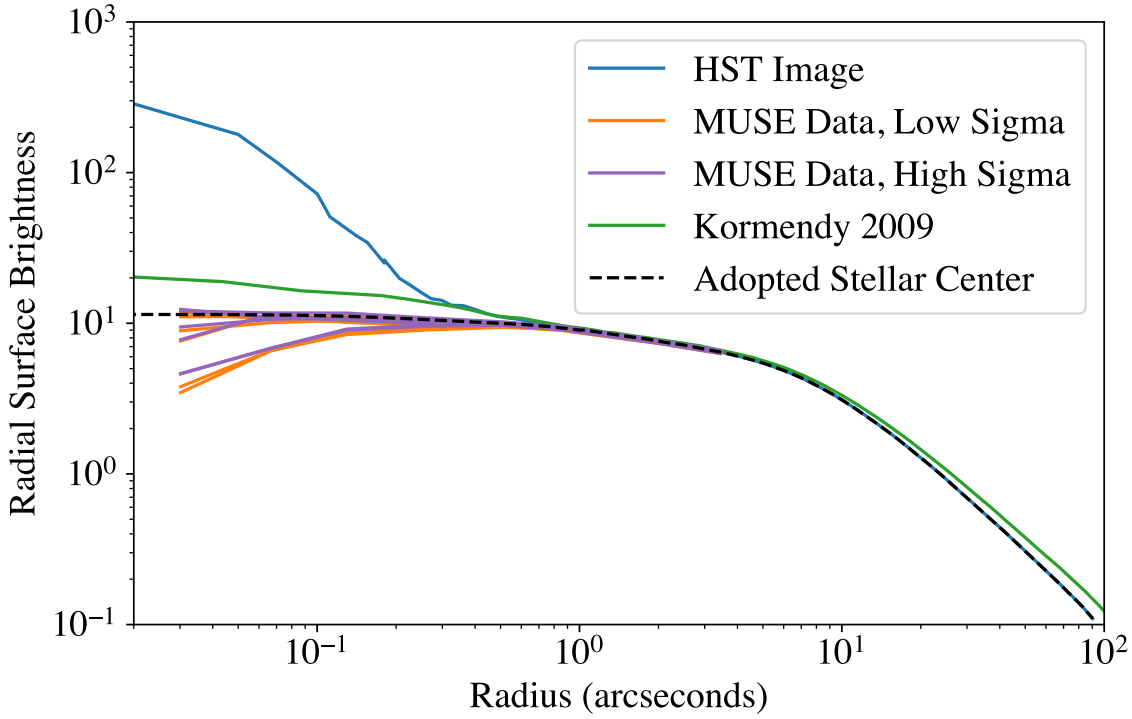

Figure 2.10: Comparison of a scaled HST radial profile with a profile where the centre region is set by the stellar profile measured spectroscopically from MUSE. The purple and orange curves show the spectroscopically extracted stellar profile from the MUSE data under different assumptions. The orange curve fixes the stellar kinematics to the lowest values seen in the bottom panel of Figure 2.6, and similarly for the purple curve with the largest values. The different lines of the same color show different choices of Legendre polynomial degree. The black dashed curve shows the profile we adopt. The green curve shows the profile determined by combining different datasets across different scales in Kormendy et al. (2009). This profile was used in the previous stellar dynamical determinations of Gebhardt & Thomas (2009); Gebhardt et al. (2011); Liepold et al. (2023a). Their profile is a factor of 2 larger than ours in the centre.

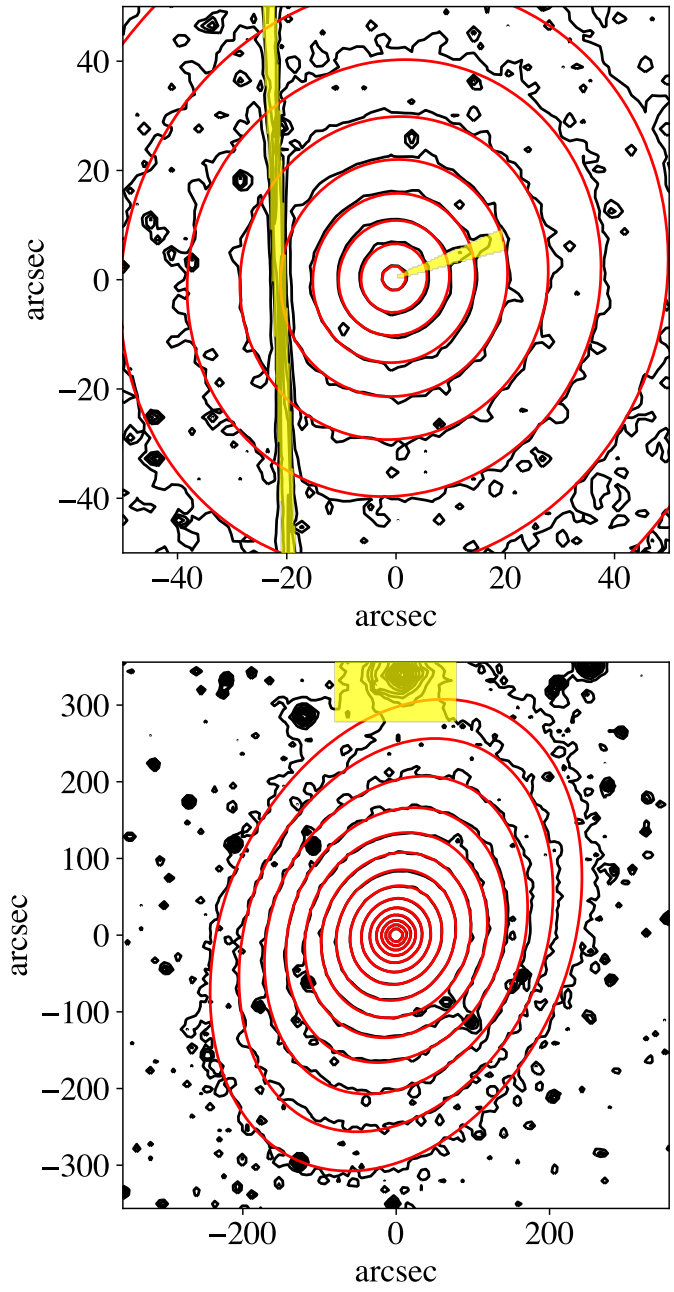

Figure 2.11: Top panel: best fit MGE contours in the innermost 100 arcseconds of the HST image. The regions in yellow are the masked jet and gap between the detectors. Bottom panel: best fit MGE contours over the full SDSS field. We mask a prominent star at the top of the field. You can see that the MGE fit covers the full shape of the galaxy and provides an excellent visual fit in the innermost region where our data is the most sensitive. The galaxy is oriented so that north is at the top and east is to the left.

Table 2.2: MGE parameters for the deconvolved ACS/WFC F850LP surface brightness. This corresponds to the orange profile in Figure 2.10 and the red contours in Figure 2.11. The principal photometric axis measured is 156.7 degrees east of north.

j	$\lg(I_j)$ ($L_\odot$ pc^{-2})	$\lg(\sigma_j)$ (arcsec)	q_j
1	3.070	-0.834	0.957
2	3.372	-0.147	0.957
3	3.354	0.246	0.957
4	3.366	0.624	0.957
5	3.435	0.820	0.957
6	3.310	1.056	0.957
7	3.001	1.314	0.957
8	2.389	1.583	0.847
9	2.380	1.745	0.936
10	1.894	2.018	0.743
11	1.488	2.317	0.743

using all six-dimensional components of the stellar phase space (Nitschai et al., 2020, 2021).

Tests of JAM using high-resolution N-body simulations (Lablanche et al., 2012) and lower-resolution but more extensive cosmological hydrodynamical simulations (Li et al., 2016) have shown that, with high-S/N data, JAM recovers accurate total density profiles with negligible bias. More recently, JAM was compared in detail against the Schwarzschild (1979) method using samples of both observed galaxies, with circular velocities from interferometric observations of the CO gas, and numerical simulations respectively. Both studies consistently found that the JAM method produces even more accurate (smaller scatter vs the true values) density profiles (Leung et al., 2018, fig. 8) and enclosed masses (Jin et al., 2019, fig. 4) than the more general Schwarzschild models. More quantitatively, between 0.8-1.6 effective radii, where the gas is well-resolved and the V_c is better determined, Leung et al. (2018) reports a mean 1 σ error 1.7× smaller for JAM over the equivalent Schwarzschild model. Similarly, when considering all 45 model fits to the N-body simulations by Jin et al. (2019), the 68th percentile deviation (1 σ error) is 1.6× smaller for JAM than the equivalent Schwarzschild model. The increased accuracy of JAM in extracting density distributions may be due to JAM assumptions acting as an empirically-motivated prior and reducing the degeneracies of the dynamical inversion.

For supermassive black hole studies, JAM was found to accurately recover the "known" mass of the two most accurate benchmark black holes in NGC4258 (Drehmer

et al., 2015) and the Milky Way (Feldmeier-Krause et al., 2017, sec. 4.1.2). Moreover, extensive tests of a few tens of galaxies have found that JAM and Schwarzschild methods recover black hole masses that are generally consistent with one another (Cappellari et al., 2010; Seth et al., 2014; Thater et al., 2017, 2022a; Krajnović et al., 2018). However, not all BH measurements from different methods agree within the uncertainties and further comparisons between different approaches are still needed.

There are two implementations of JAM: one where the velocity ellipsoid is assumed to be aligned with the cylindrical-polar coordinate system (Cappellari, 2008), and one where the velocity ellipsoid is assumed to be aligned with the spherical-polar coordinate system (Cappellari, 2020). The choice of which implementation to use depends on the galaxy's intrinsic shape. M87 is a slow rotator early type galaxy (Emsellem et al., 2011) and slow rotators as a class are weakly triaxial, or nearly spherical, inside the half-light radius, becoming more triaxial at larger radii (see review by Cappellari (2016)). M87 has a specific angular momentum $\lambda_{\mathrm{R_e}} \approx 0$ within the uncertainties (Emsellem et al., 2011). It showed some barely detectable misaligned stellar rotation from SAURON data (Emsellem et al., 2004), which became more clearly visible from high-S/N MUSE data (Emsellem et al., 2014). The ellipticity of M87 increases at large radii (see Figure 2.11), where misaligned stellar rotation and triaxiality become evident (Liepold et al., 2023a). Within the slow-rotators class, M87 was classified as a non-rotator (Krajnović et al., 2011). As a class, these are massive early type galaxies generally found at the centre of clusters, which tend to be rounder than $\epsilon \lesssim 0.15$ in projection (Emsellem et al., 2011, fig.6), indicating that, although triaxial, they must be intrinsically close to spherical with a ratio between the minor to major axes of the ellipsoidal density $c/a \gtrsim 0.85$. This excludes the possibility that M87, which is nearly round in projection, may appear as such due to a special viewing angle. Instead, M87 must be intrinsically close to spherical in the region where it shows circular isophotes.

JAM models, unlike the Schwarzschild models, do not require large-radii kinematics to constrain the models because they are nearly insensitive to the mass distribution at radii outside the region where kinematics are available. For this reason, one can expect axisymmetric JAM models with the velocity ellipsoid aligned with spherical-polar coordinates, to provide an accurate description of the inner dynamics of M87, even though the galaxy, like all slow rotators, becomes more strongly triaxial at much larger radii than those we model.

2.6.2 Priors on the Anisotropy

JAM takes as parameters the galaxy inclination, anisotropy profile, stellar tracer distribution, and mass distribution, including the black hole mass. The kinematic axis is poorly constrained due to M87 being dominated by unordered motion in the central region, (Emsellem et al., 2014; Sarzi et al., 2018) so we fix the kinematic axis to align with the galaxy photometric axis as given in Krajnović et al. (2011). The value reported there is 151.3 degrees east or north, which is consistent with our value of 156.7 degrees east of north. The stellar tracer and mass distributions are described in subsection 2.5.1. As the inner regions of M87 are nearly spherically symmetric with little ordered motion, we fix the inclination to 90 degrees. Changing this has a negligible effect on the measured black hole masses, because a nearly spherical model appears spherical from any inclination. We can exclude M87 being an intrinsically flat but nearly face-on disk, because of the general shape distribution of slow rotators (Cappellari, 2016; Li et al., 2018b).

The last thing to specify is the anisotropy profile. It is well known that for a spherical system, there is a degeneracy between the anisotropy and the density profile. This so-called mass-anisotropy degeneracy implies that for a range of assumed density profiles, one can adopt a corresponding anisotropy profile in such a way that the model reproduces the same profile of second velocity moments (Binney & Mamon, 1982; Gerhard, 1993). The degeneracy, however, is not complete, and the range of allowed profiles depends on the specific situation because the anisotropy is limited by the two extreme cases where the orbital distribution is fully radial or fully tangential respectively. For this reason, without further assumptions, one would generally expect large uncertainties in BH masses from spherical models based on the Jeans equations.

The situation, however, has improved dramatically from the days when the mass-anisotropy degeneracy was first discovered. Since then, many studies have modelled the inner dynamics of galaxies using general models that allow one to account for the full shape of the line-of-sight velocity distribution, rather than the moments alone. We think we now even have a good understanding of the underlying physics of the orbital distributions we have measured. In particular, we have found that massive slow-rotator galaxies with a core in their surface brightness profile, like M87, are consistently characterized by a nearly isotropic, or just slightly radially anisotropic orbital distribution outside the break radius, while orbits start becoming tangentially biased inside that radius, reaching the peak tangential anisotropy well inside the BH sphere of influence (Gebhardt et al., 2003 fig.10; Cappellari et al., 2008 fig.2; Thomas et al., 2014). The observations are quantitatively well reproduced by models in which

both the cores in the surface brightness and the tangentially biased orbits are due to gas-free mergers of galaxies with supermassive black holes in their centres. The black holes sink towards the centre of the gravitational potential via dynamical friction, while ejecting stars on radial orbits (e.g. Milosavljević & Merritt, 2001; Milosavljević et al., 2002; Rantala et al., 2019; Frigo et al., 2021).

In the case of M87, due to its very flat inner core, one can place constraints on its orbital anisotropy even from theoretical arguments alone. The cusp-slope vs central anisotropy theorem by An & Evans (2006) states that for a spherical power-law tracer population $\rho \propto r^{-\gamma}$ in a Keplerian potential, there is a relation between the anisotropy $\beta = 1 - \sigma_t^2/\sigma_r^2$ with σ_t and σ_r the tangential and radial dispersion, respectively, and the logarithmic slope γ of the tracer, such that $\beta < \gamma - 1/2$[11]. The inner slope of M87 varies from nearly flat in the centre ($\gamma \approx 0$ see Figure 2.10) to $\gamma \approx 0.27$ between 1-5″(Lauer et al., 2007). We can thus conservatively conclude that the inner anisotropy has an upper limit of $\sigma_r/\sigma_t \lesssim 0.9$ and possibly even $\sigma_r/\sigma_t \lesssim 0.8$ for $\gamma = 0$. The assumptions of the theorem are satisfied well inside the sphere of influence of the BH of M87 and for this reason the theorem provides additional support for the expected significant tangential anisotropy near the BH of M87.

For all these theoretical and empirical reasons, nowadays, it does not make sense to assume complete freedom in the orbital anisotropy of Jeans models as done in the past. Instead, the knowledge we accumulated on the galaxies anisotropy can be used as a Bayesian prior, which is easy to enforce to our JAM models. The ability to place priors is an important feature of stellar dynamical codes. All Schwarzschild codes use regularization to enforce smoothness in the orbital distribution (e.g. Richstone & Tremaine (1988); van der Marel et al. (1998); Gebhardt et al. (2000a); Cappellari et al. (2002); Valluri et al. (2004); Thomas et al. (2005)). A recent study suggested that the fine tuning of regularization is essential for accurate results (Neureiter et al., 2023).This regularisation is mathematically equivalent to a prior. This is because a prior modifies the likelihood by a multiplication via Bayes theorem. Since the χ^2 is the log of the likelihood, adding another term $\chi^2_{\rm new}$ is the same as multipliying the likelihood by the factor $e^{-\chi^2_{\rm new}}$.

In the next section we describe a new way of specifying the anisotropy variations in JAM models, which is ideally suited to enforce anisotropy priors.

[11]In fact, it is possible to generate JAM models that violate this relation. The second column in Figure 2.12 has a β of 0.75 while the limit from the slope-anisotropy theorem is 0.5. It is important to ensure when doing JAM modelling of real galaxies that the relation is enforced, either through priors or otherwise.

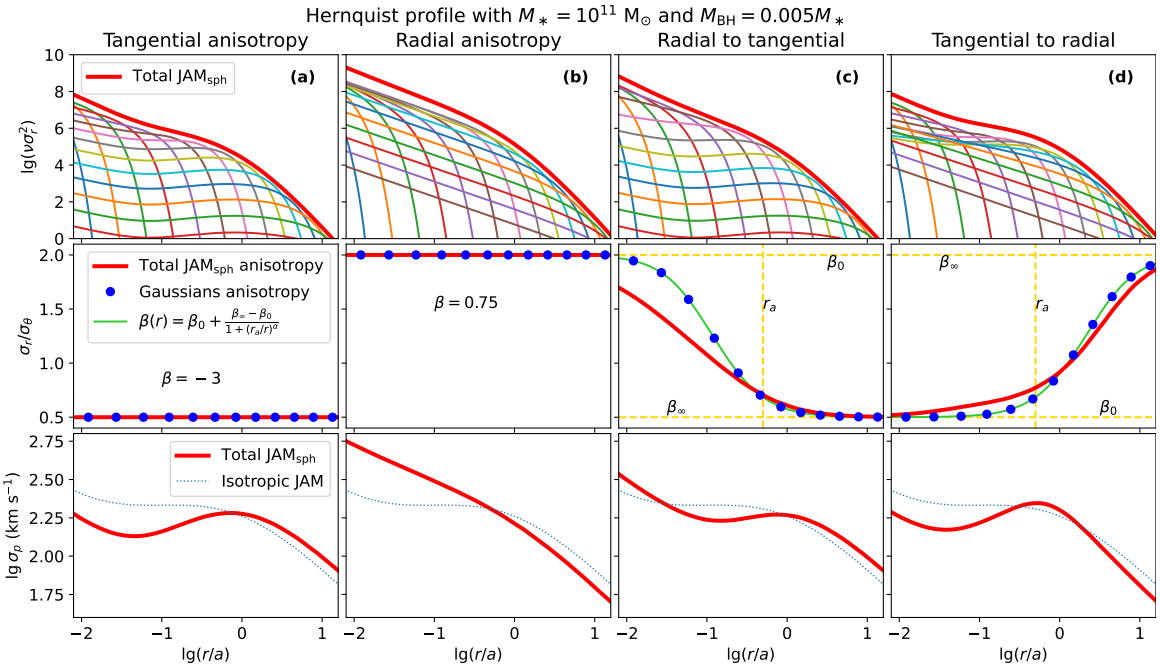

Figure 2.12: Top row: the contributions to the total luminosity weighted second moment from different MGE components. Middle row: intrinsic anisotropy ratio compared with the anisotropy assigned to each gaussian. Bottom row: projected velocity dispersion. The columns correspond to different choices of the total anisotropy, ranging from constant to radially varying. This figure was provided by Michele Cappellari.

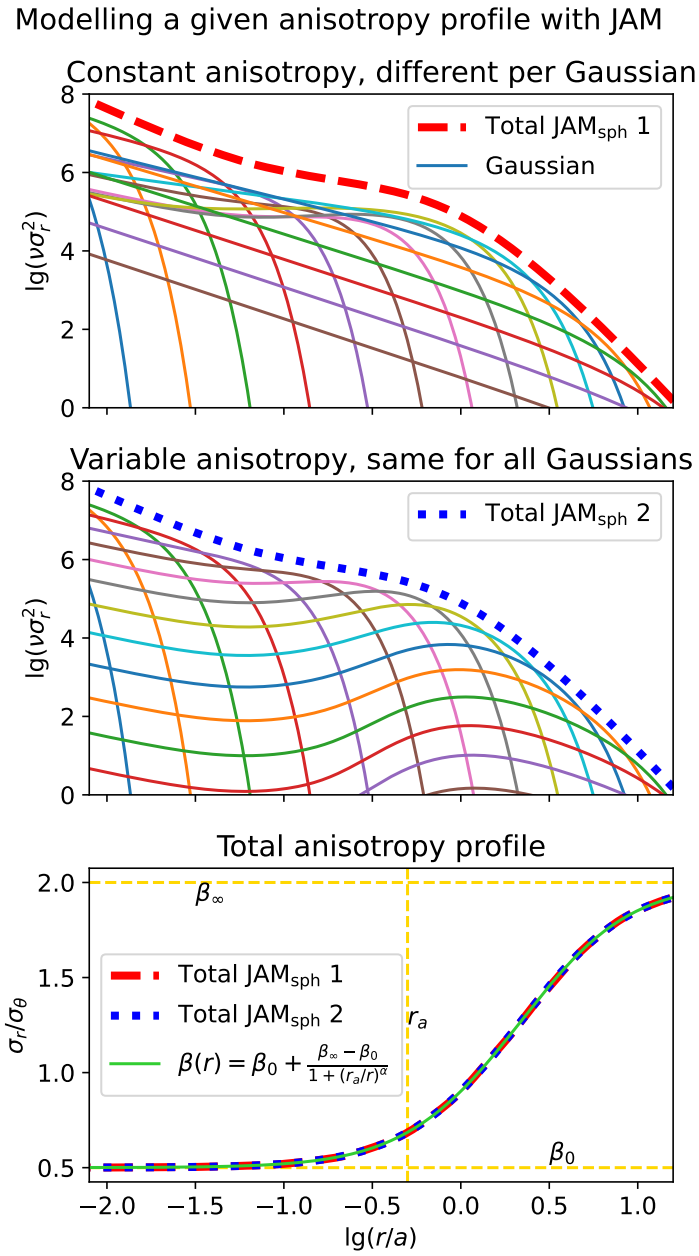

Figure 2.13: The top panel shows the contributions to the radial velocity dispersion where the desired anisotropy profile is reached by fitting the anisotropy for each gaussian in the MGE. The center panel shows the same thing using the analytic implementation described in Equation 2.12a–Equation 2.14. The bottom panel shows the intrinsic anisotropy ratio and confirms that the result using the two methods is the same, the key difference being that the top one requires fitting the anisotropy for each gaussian while the center one requires a small modification to the analytic solution. This figure was provided by Michele Cappellari.

2.6.3 Fitting a given anisotropy profile with JAM

Let's consider for simplicity a spherical non-rotating JAM model. The intrinsic stellar dispersion of the model is given by the luminosity-weighted sum of the dispersion of the individual Gaussians making up the MGE

$$\nu\sigma_r^2 = \sum_k [\nu\sigma_r^2]_k. \tag{2.10}$$

Where ν is the deprojected MGE oblate axisymmetric luminous density (see equation 13 of Cappellari (2008)). Figure 2.12 shows the contribution of the individual $[\nu\sigma_r^2]_k$ for different anisotropies, for a Hernquist (1990) model with mass $M_* = 10^{11}$ $M_\odot$ containing a typical nuclear supermassive black hole of mass 0.5% that of the stellar mass (Kormendy & Ho, 2013, eq. 11). When the Gaussians in a JAM model are isotropic or have tangential anisotropy, each Gaussian essentially contributes to the total $\nu\sigma_r^2$ only near a radius close to its dispersion $r \approx \sigma_k$ (Figure 2.12a). In these cases, one can construct a desired anisotropy profile $\beta(r)$ by simply assigning the anisotropy $\beta_k = \beta(\sigma_k)$ to the Gaussians with dispersion σ_k, as pointed out in Cappellari (2008, sec. 3.2.2). Figure 2.12 shows that this approximation works quite well in general. However, when the Gaussians are significantly tangential anisotropic, or near the central supermassive black hole, the total $\nu\sigma_r^2$ rises steeply at small radii and a single Gaussian does not contribute to the total JAM model only around $r \approx \sigma_k$ (Figure 2.12c,d). In those situations, there is no simple precise relation between the anisotropy of a given Gaussian and the total anisotropy of the JAM model at $r \approx \sigma_k$. This is generally not a problem, if one is not interested in the fitted anisotropy. However, if one wants to quantitatively reproduce a specific total anisotropy profile one has to numerically fit for the anisotropies β_k of the different Gaussians.

In this paper we parametrize the anisotropy using a rather flexible logistic function of logarithmic radius

$$\beta(r) = \beta_0 + \frac{\Delta\beta}{1 + (r_a/r)^\alpha}, \tag{2.11}$$

with $\Delta\beta = \beta_\infty - \beta_0$. This anisotropy function was also used by Baes & van Hese (2007, eq. 30). For $\beta_0 = 0$, $\beta_\infty = 1$ and $\alpha = 2$ it reduces to the Osipkov-Merritt special form (Osipkov, 1979; Merritt, 1985). For $\alpha = 1$ it specializes to the homographic anisotropy function used by Bacon (1985).

In the top panel of Figure 2.13 we show how one can reproduce our logistic anisotropy profile with JAM. In the figure, we adopted a rather extreme anisotropy variation, with inner tangential anisotropy $\sigma_r/\sigma_\theta = 1/2$ ($\beta_0 = -3$), outer radial

anisotropy $\sigma_r/\sigma_\theta = 2$ ($\beta_\infty = 0.75$), anisotropy radius $r_a = a/2$, where a is the break radius of the Hernquist (1990) profile, and the sharpness of the transition is $\alpha = 1.5$. One can see that by fine-tuning the anisotropy of the different Gaussians, one can reproduce quite general anisotropy variations. The problem with this approach is that the anisotropy β_k of the individual Gaussians has to be fitted non-linearly in a least-square sense, while repeatedly computing the intrinsic velocity moments of the model, to obtain the total anisotropy for a given choice of β_k parameters.

An alternative is to solve the original Jeans equations for an axisymmetric model with spherically-aligned velocity ellipsoid, in Cappellari (2020, eq. 8) by relaxing the assumption of a constant anisotropy per Gaussian. This is only possible analytically for special choices of the anisotropy function $\beta(r, \theta)$. We found that, adopting the parametrization of Equation 2.11 for the anisotropy, the solution of Cappellari (2020, eq. 10), using the same notations, generalizes to

$$\nu\overline{v_r^2}(r,\theta) = \int_r^\infty \left(\frac{r'}{r}\right)^{2\beta_0} \left[\frac{1+(r'/r_a)^\alpha}{1+(r/r_a)^\alpha}\right]^{\frac{2\Delta\beta}{\alpha}} \Psi(r',\theta')\,\mathrm{d}r' \tag{2.12a}$$

$$\theta' = \arcsin\left\{\left(\frac{r'}{r}\right)^{\beta_0-1} \left[\frac{1+(r'/r_a)^\alpha}{1+(r/r_a)^\alpha}\right]^{\frac{\Delta\beta}{\alpha}} \sin\theta\right\}. \tag{2.12b}$$

In the special case $\alpha = 1$, this solution reduces the one using homographic functions in Bacon (1985, eq. 3)[12].

The application of a generic varying anisotropy function to the axisymmetric cylindrically-aligned Jeans solution of Cappellari (2008, eqs. 8, 9), is straightforward if one make the cylindrical anisotropy a function $\beta_z(|z|)$ of the modulus of the cylindrical coordinate z. One simply has to replace the constant axial anisotropy β_z, with the corresponding varying expressions $\beta_z(z)$ parametrized by the function of Equation 2.11. In the case of the tangential anisotropy γ, the situation is identical regardless of the alignment of the velocity ellipsoid, and one can just replace the constant with an arbitrary function of the coordinates $\gamma(R, z)$, without having to change anything else.

When applying the axisymmetric solution to a model with both the tracer distribution and the total density described by an MGE (Emsellem et al., 1994; Cappellari, 2002), following the steps outlined in Cappellari (2020, sec. 5.1), the resulting solution requires only minimal changes and the numerical algorithm can be left unchanged. One only needs to replace the following two expressions with the corresponding one

[12]The published expression has a typo, with an extra density $\nu(\rho, \alpha)$.

indicated by the arrows, in all the expressions of Cappellari (2020, eqs. 46–53)

$$\beta_k \to \beta_0 + \frac{\Delta\beta}{1 + (r_a/r)^\alpha} \tag{2.13}$$

$$\left(\frac{r'}{r}\right)^{2\beta_k} \to \left(\frac{r'}{r}\right)^{2\beta_0} \left[\frac{1 + (r'/r_a)^\alpha}{1 + (r/r_a)^\alpha}\right]^{\frac{2\Delta\beta}{\alpha}}. \tag{2.14}$$

In the spherical limit, the expression for the intrinsic radial velocity dispersion in Cappellari (2020, eq. B2) generalizes to

$$\nu\overline{v_r^2}(r) = \int_r^\infty \left(\frac{r'}{r}\right)^{2\beta_0} \left[\frac{1 + (r'/r_a)^\alpha}{1 + (r/r_a)^\alpha}\right]^{\frac{2\Delta\beta}{\alpha}} \frac{\nu(r')M(r')}{r'^2}\, \mathrm{d}r'. \tag{2.15}$$

Unlike the constant-anisotropy case, when projecting this model along the line-of-sight one cannot remove one of the two resulting integrals, except for some special cases (Mamon & Łokas, 2005), which are not very useful for practical applications.

We implemented these changes in v7.0 of the Python JamPy package[13], which now allows one to compute axisymmetric or spherical models with the logistic radial anisotropy variation, for both the spherically-aligned and cylindrically-aligned solutions. An application in the spherical limit is shown in the middle panel of Figure 2.13. As expected the JAM model with the same variable-anisotropy for all Gaussians produces the same dispersion profile as the model with constant anisotropy for each individual Gaussian, as they both follow by design the same given anisotropy profile.

2.6.4 MCMC Analysis

For each of the models and dispersion profiles we perform an MCMC analysis to carefully assess the influence of the modelling and systematic effects of the kinematic extraction on the recovered BH mass. We define the χ^2 to be

$$\chi^2 = \sum_{\ell=1}^{N} \left(\frac{\sigma_m^\ell - \sigma_d^\ell}{\Delta\sigma^\ell}\right)^2 \tag{2.16}$$

where σ_d^ℓ is the extracted dispersion in binned spaxel ℓ, σ_m^ℓ is the model dispersion, and $\Delta\sigma^\ell$ is the bootstrapped uncertainty in the extracted dispersion. We thus express the χ^2 for the combined data sets as

$$\chi^2 = \chi^2_{\mathrm{MUSE}} + \chi^2_{\mathrm{OASIS}} + \chi^2_{\mathrm{SAURON}} \tag{2.17}$$

From here, we perform an MCMC analysis using the code emcee of Foreman-Mackey et al. (2013). For each of the three combinations of different kinematic

[13] `https://pypi.org/project/jampy`

extractions described in subsection 2.4.1, we test four models, for a total of 12 different combinations of data and models. In each model we adopt as free parameters the four anisotropy parameters $(\beta_0, \beta_\infty, r_a, \alpha)$ described in subsection 2.6.2, the black hole mass $M_{\rm BH}$, and make the following assumptions:

- Constant stellar M/L without NFW dark matter. This adds an extra parameter $(M/L)_{\rm tot}$ for a total of 6 model parameters.
- Constant stellar M/L with NFW dark matter. This add two extra free parameters, the stellar $M*/L$ and the normalization of the halo, quantified by the dark matter fraction within one effective radius $f_{\rm dm}(< R_{\rm e})$, for a total of 7 model parameters.
- Varying M/L without NFW dark matter. This adds two extra free parameters: $(M/L)_1$ and $(M/L)_2$ which parametrize the mass-to-light variation (see Equation 2.7), for a total of 7 parameters.
- Varying M/L with NFW dark matter. This combines both a NFW dark matter halo, parametrized by the dark matter fraction $f_{\rm dm}$, along with the parameters $(M/L)_1$ and $(M/L)_2$ which parametrize the mass-to-light variation, for a total of 8 parameters.

These varying assumptions allow us to make contact with previous work which have made various assumptions on the parametrization of the gravitational potential, and also allow us to test the impact of each model assumption on the final result. It is worth noting that we sample most parameters linearly, rather than in the logarithm. Different approaches have been taken in the literature (e.g. Liepold et al. (2023b) samples the black hole logarithmically and the M/L linearly). For this work we proceed with linear sampling for most parameters as we already have an idea of what range of values each parameter should take, but acknowledge that there may be other well motivated choices.

In order to best sample the space of possible parameters, we re-express the anisotropy parameters in terms that result in a more efficient and uniform sampling of the model posterior. Namely, we sample the anisotropy parameters defined such that

$$(\sigma_r/\sigma_t)_0 = \frac{1}{\sqrt{1-\beta_0}} \tag{2.18}$$

$$(\sigma_r/\sigma_t)_\infty = \frac{1}{\sqrt{1-\beta_\infty}} \tag{2.19}$$

$$a = \lg \alpha \tag{2.20}$$

Table 2.3: Table of free parameters and their permitted upper and lower bounds. The limits on $(\sigma_r/\sigma_t)_0$ and $(\sigma_r/\sigma_t)_\infty$ come from the large literature of observations and simulations of core galaxies. The limits on M/L come from considering the heaviest possible IMF combined with the results from Sarzi et al. (2018)

Parameter	Lower Bound	Upper Bound
$(\sigma_r/\sigma_t)_0$	0.5	1
$(\sigma_r/\sigma_t)_\infty$	1	1.3
r_a	0	∞
$\lg \alpha$	-0.3	0.6
$M_{\rm bh}$	0	∞
$(M/L)_1$	0	7.23
$(M/L)_2$	0	7.23
$f_{\rm dm}$	0	1

Given what we know about the anisotropy profile in M87, we restrict these parameters to the ranges $0.5 \leq (\sigma_r/\sigma_t)_0 \leq 1$ and $1 \leq (\sigma_r/\sigma_t)_\infty \leq 1.3$. The bound at 1 comes from the enforced condition that the anisotropy becomes tangential in the center and radial at large radii. The bounds at 0.5 and 1.3 represent the largest range of anisotropy values reliably observed in early type galaxies Gebhardt et al. (2003); Cappellari et al. (2008, 2009); McConnell et al. (2012); Thomas et al. (2014); Krajnović et al. (2018), as well as in simulations of slow rotators Rantala et al. (2019); Frigo et al. (2021). This range also encompasses the range of anisotropy profiles previously determined for M87 (Cappellari & McDermid, 2005; Gebhardt et al., 2011). Furthermore, this range is guaranteed to satisfy the slope-anisotropy theorem (An & Evans, 2006) as σ_r/σ_t=1 corresponds to $\beta = 0$ and the inner slope of M87 is approximately flat. In Table 2.3, we list all of the parameters along with the corresponding bounds on their values. We also restrict the parameter $\lg \alpha$ to be between -0.3 and 0.6. This is due to the fact that small values of α correspond to no anisotropy transition, which we want to exclude, and large values of α give rise to an infinitely large parameter space where the spatial transition of the anisotropy takes place nearly instantaneously. In practice, the preferred range of parameter space almost always lies between -0.3 and 0.6 so this does not impact our final results.

Running JAM for the required number of steps necessary to generate reliable posteriors for all of these combinations of models and data is very computationally expensive. The final contours exhibit strong covariances that require a long burn in time for the walkers to sample and populate the posterior. The likelihood also has multiple local minima which further increases the run time if the chain is started further away from the global minimum. In order to speed up this process, we start by

Table 2.4: Table of best fit parameters for each combination of instrument, spectral range, and legendre polynomial degree. The values presented are the median of the posteriors, with the upper and lower bounds corresponding to the 1σ interval. The black hole masses are clustered between 8 and 10 billion solar masses. In the data column M stands for MUSE, O for OASIS, and S for SAURON.

Data	Model	$(\sigma_r/\sigma_t)_0$	$(\sigma_r/\sigma_t)_\infty$	r_a	$\lg \alpha$	$M_{\rm bh}$	$(M/L)_1$	$(M/L)_2$	$f_{\rm dm}$
M RNI + O RNI + S	Constant M/L	$0.55^{+0.08}_{-0.04}$	$1.24^{+0.03}_{-0.05}$	$1.4^{+0.5}_{-0.3}$	$0.08^{+0.04}_{-0.04}$	$(9.7^{+0.7}_{-0.9})\times10^9\ M_\odot$	$3.4^{+0.1}_{-0.1}$	N/A	N/A
M RNI + O RNI + S	NFW DM	$0.54^{+0.07}_{-0.03}$	$1.26^{+0.03}_{-0.05}$	$1.6^{+0.5}_{-0.4}$	$0.09^{+0.04}_{-0.04}$	$(10.2^{+0.8}_{-1.0})\times10^9\ M_\odot$	$3.2^{+0.1}_{-0.2}$	N/A	$0.06^{+0.07}_{-0.04}$
M RNI + O RNI + S	Varying M/L	$0.58^{+0.14}_{-0.06}$	$1.21^{+0.06}_{-0.06}$	$1.2^{+0.6}_{-0.5}$	$0.04^{+0.07}_{-0.08}$	$(8.2^{+1.7}_{-1.7})\times10^9\ M_\odot$	$4.7^{+0.8}_{-0.9}$	$3.1^{+0.2}_{-0.2}$	N/A
M RNI + O RNI + S	Varying M/L + NFW DM	$0.56^{+0.09}_{-0.05}$	$1.19^{+0.07}_{-0.07}$	$1.1^{+0.5}_{-0.5}$	$0.02^{+0.07}_{-0.08}$	$(8.7^{+1.2}_{-1.2})\times10^9\ M_\odot$	$6.5^{+0.5}_{-0.8}$	$1.5^{+0.6}_{-0.5}$	$0.36^{+0.14}_{-0.14}$
M d=4 + O d=6 + S	Constant M/L	$0.53^{+0.08}_{-0.02}$	$1.24^{+0.04}_{-0.04}$	$1.1^{+0.4}_{-0.3}$	$0.05^{+0.07}_{-0.04}$	$(9.2^{+0.7}_{-1.0})\times10^9\ M_\odot$	$3.4^{+0.1}_{-0.1}$	N/A	N/A
M d=4 + O d=6 + S	NFW DM	$0.53^{+0.05}_{-0.02}$	$1.26^{+0.03}_{-0.05}$	$1.2^{+0.3}_{-0.2}$	$0.05^{+0.03}_{-0.04}$	$(9.6^{+0.6}_{-0.7})\times10^9\ M_\odot$	$3.3^{+0.1}_{-0.1}$	N/A	$0.03^{+0.05}_{-0.02}$
M d=4 + O d=6 + S	Varying M/L	$0.59^{+0.15}_{-0.07}$	$1.19^{+0.07}_{-0.07}$	$0.9^{+0.6}_{-0.4}$	$0.02^{+0.1}_{-0.08}$	$(7.7^{+1.2}_{-1.6})\times10^9\ M_\odot$	$5.0^{+0.9}_{-0.8}$	$3.1^{+0.2}_{-0.2}$	N/A
M d=4 + O d=6 + S	Varying M/L + NFW DM	$0.57^{+0.11}_{-0.05}$	$1.14^{+0.07}_{-0.06}$	$0.9^{+0.4}_{-0.4}$	$0.03^{+0.09}_{-0.09}$	$(8.3^{+1.3}_{-1.6})\times10^9\ M_\odot$	$6.5^{+0.5}_{-0.9}$	$2.0^{+0.6}_{-0.6}$	$0.24^{+0.14}_{-0.12}$
M d=6 + O d=4 + S	Constant M/L	$0.67^{+0.17}_{-0.09}$	$1.22^{+0.06}_{-0.06}$	$2.2^{+1.5}_{-0.7}$	$0.12^{+0.07}_{-0.06}$	$(8.9^{+1.4}_{-1.4})\times10^9\ M_\odot$	$3.5^{+0.1}_{-0.1}$	N/A	N/A
M d=6 + O d=4 + S	NFW DM	$0.54^{+0.04}_{-0.03}$	$1.26^{+0.02}_{-0.04}$	$1.2^{+0.3}_{-0.2}$	$0.05^{+0.03}_{-0.04}$	$(9.6^{+0.6}_{-0.8})\times10^9\ M_\odot$	$3.3^{+0.1}_{-0.1}$	N/A	$0.03^{+0.05}_{-0.02}$
M d=6 + O d=4 + S	Varying M/L	$0.68^{+0.19}_{-0.12}$	$1.17^{+0.07}_{-0.07}$	$2.0^{+1.2}_{-0.7}$	$0.14^{+0.08}_{-0.09}$	$(8.7^{+1.6}_{-1.7})\times10^9\ M_\odot$	$4.5^{+0.9}_{-0.9}$	$3.2^{+0.2}_{-0.2}$	N/A
M d=6 + O d=4 + S	Varying M/L + NFW DM	$0.66^{+0.13}_{-0.11}$	$1.16^{+0.09}_{-0.07}$	$1.7^{+1.1}_{-0.5}$	$0.08^{+0.09}_{-0.1}$	$(8.9^{+1.4}_{-1.6})\times10^9\ M_\odot$	$6.2^{+0.8}_{-1.0}$	$1.7^{+0.8}_{-0.7}$	$0.33^{+0.16}_{-0.17}$
M RNI	Varying M/L + NFW DM	$0.58^{+0.14}_{-0.06}$	$1.24^{+0.04}_{-0.09}$	$1.7^{+1.23}_{-0.58}$	$0.00^{+0.08}_{-0.05}$	$(10.0^{+1.3}_{-1.7})\times10^9\ M_\odot$	$6.9^{+0.3}_{-0.5}$	$1.8^{+0.2}_{-0.3}$	$0.07^{+0.08}_{-0.05}$
O RNI	Varying M/L + NFW DM	$0.56^{+0.1}_{-0.04}$	$1.23^{+0.05}_{-0.07}$	$2.1^{+1.12}_{-0.54}$	$0.06^{+0.07}_{-0.06}$	$(11.1^{+0.8}_{-1.2})\times10^9\ M_\odot$	$6.9^{+0.3}_{-0.5}$	$1.6^{+0.3}_{-0.3}$	$0.06^{+0.08}_{-0.04}$

The values in this table are computed after the kinematics have been scaled to match the SAURON data (see subsection 2.4.1). In order to convert to the MUSE scaling, multiply the black hole mass, $(M/L)_1$, and $(M/L)_2$ by 1.018, 1.019, 1.025, and 1.007 for the the Full spectra degree 4, 5, 6 and RNI spectrum, respectively. To convert to the OASIS scaling, multiply by 1.051, 1.051, 1.057, and 1.053 for the Full spectra degree 4, 5, 6 and RNI spectrum, respectively. We mark our preferred data and model in bold.

running EMCEE for 300000 steps with 100 walkers for the JAM model of a spherical galaxy using the routine JAM_SPH_PROJ in JAMPY. [14] This is a good approximation to the true Jeans solution as M87 is highly spherical, especially in the range of our data. Once this step is complete, we have a good approximation of the posterior. From there, we run EMCEE for 50000 steps with 100 walkers for the JAM model assuming axisymmetry and a spherically aligned velocity ellipsoid using the routine JAM_AXI_PROJ. These final 50000 steps are what is used for the final analysis.

We present the posteriors for this in Figure 2.21, Figure 2.22, Figure 2.23, and Figure 2.24. We also show the posteriors assuming the most general model for the individual MUSE RNI, OASIS RNI, and SAURON data sets in Figure 2.14,Figure 2.15, and Figure 2.16. Given our choice of physically motivated priors, some of the posteriors are not symmetric and even run into the boundary. As such, we report the posterior median and left and right 1σ confidence interval for all of the combinations of models and combined data in Table 2.4. As we expect, the results for M87 change very little between using spherical JAM and axisymmetric JAM. We also show the model fits to the observed dispersion in Figure 2.18.

2.7 Discussion

2.7.1 Black Hole Results

In Figure 2.21, Figure 2.22, Figure 2.23, and Figure 2.24 we show the full posteriors using the RNI spectra for each of the four models. In Figure 2.17 we show 1σ and 3σ contours of the black hole mass and total M/L within one half light radius for the four different models using the RNI spectra. We choose to focus on the RNI spectra as this is the case where the OASIS and MUSE data have the closest agreement, suggesting that this choice of kinematics best reflects the true kinematics. The impact of the different choices of kinematics on the following results is some scatter, as opposed to single systemic shifts in the measured values. From Figure 2.14, Figure 2.15, and Figure 2.16 we can deduce the individual contributions of each data set to the final black hole masses. There we see that the OASIS and MUSE data are remarkably similar, with the SAURON data providing a unique contribution.

We find a large range of permissible black hole mass values, from nearly $6 \times 10^9 M_\odot$ to greater than $11 \times 10^9 M_\odot$ at one sigma uncertainty across different choices

[14] Note that this is not the same as JAM with a spherically aligned velocity ellipsoid. The final routine used is JAM_AXI_PROJ which solves the axisymmetric Jeans equation assuming the velocity ellipsoid is spherically aligned. Here we assume that the system truly is spherically symmetric, thus simplifying the computation required to solve the Jeans equation.

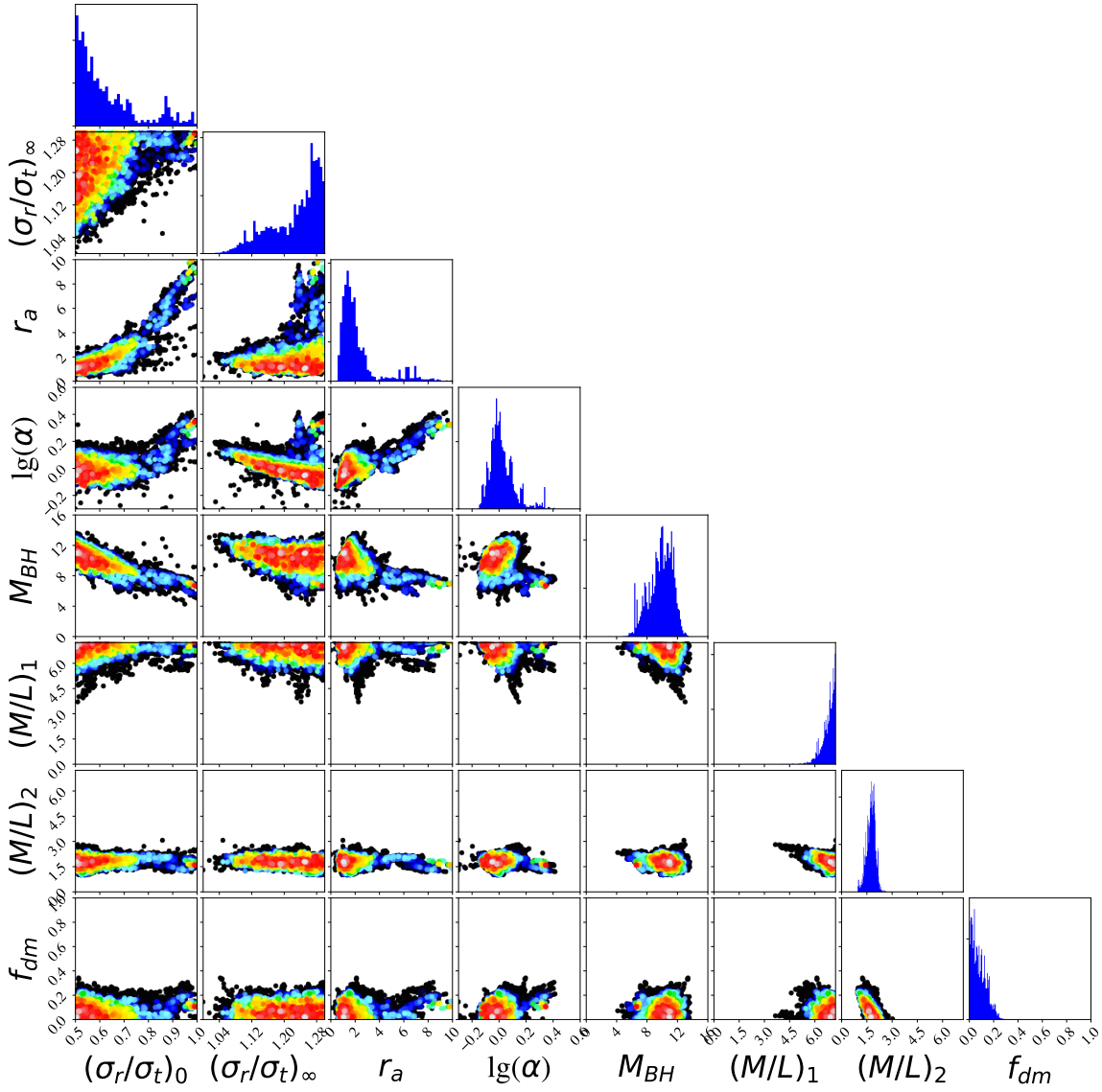

Figure 2.14: Corner plot for the individual MUSE data set using the RNI spectra. This features a secondary minimum corresponding to a near isotropic anisotropy profile in the centre of the galaxy with a lower black hole mass. Bestfit parameters and errors are shown in Table 2.4.

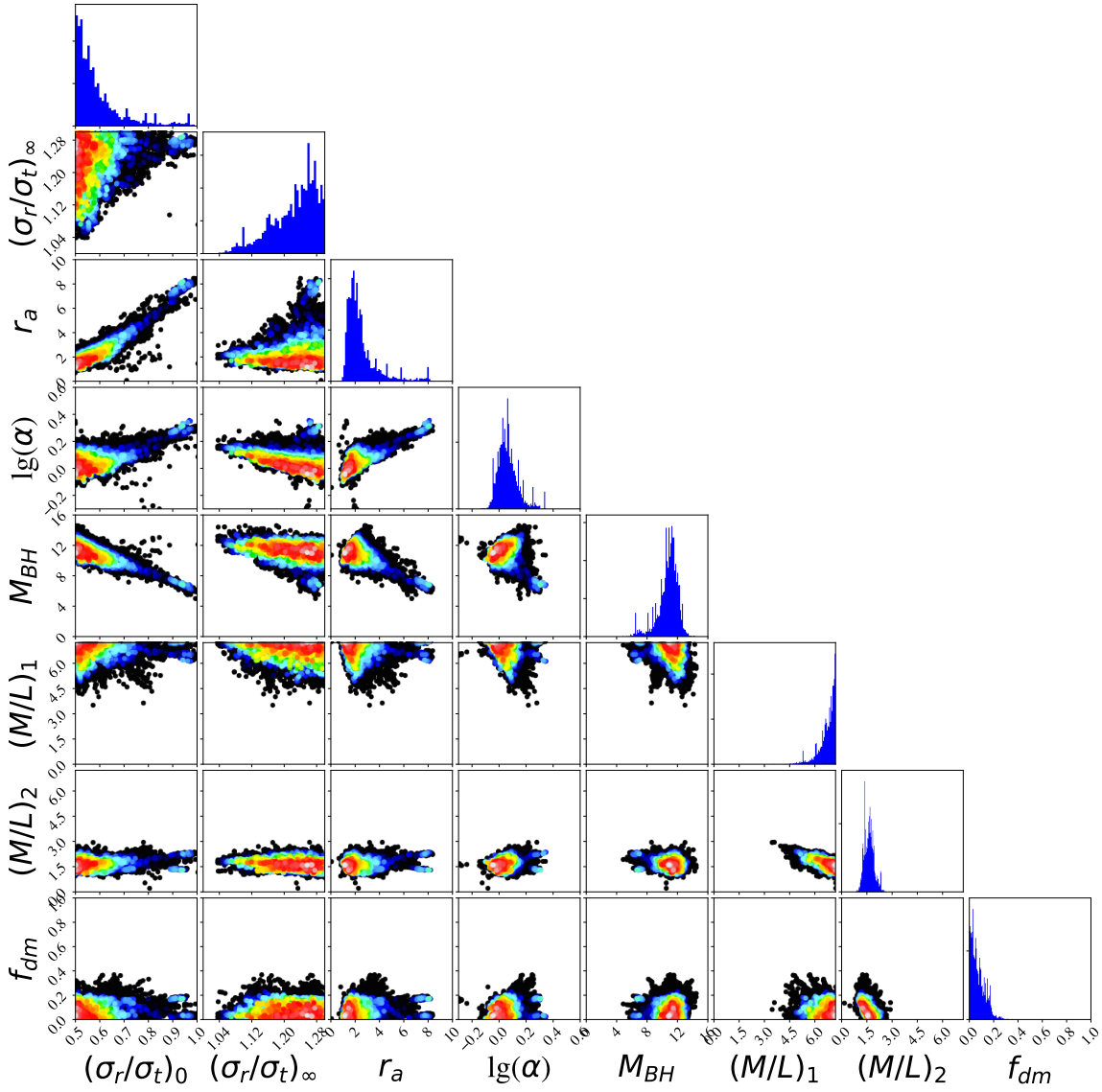

Figure 2.15: Corner plot for the individual OASIS data set using the RNI spectra. This features a secondary minimum corresponding to a near isotropic anisotropy profile in the centre of the galaxy with a lower black hole mass. Bestfit parameters and errors are shown in Table 2.4.

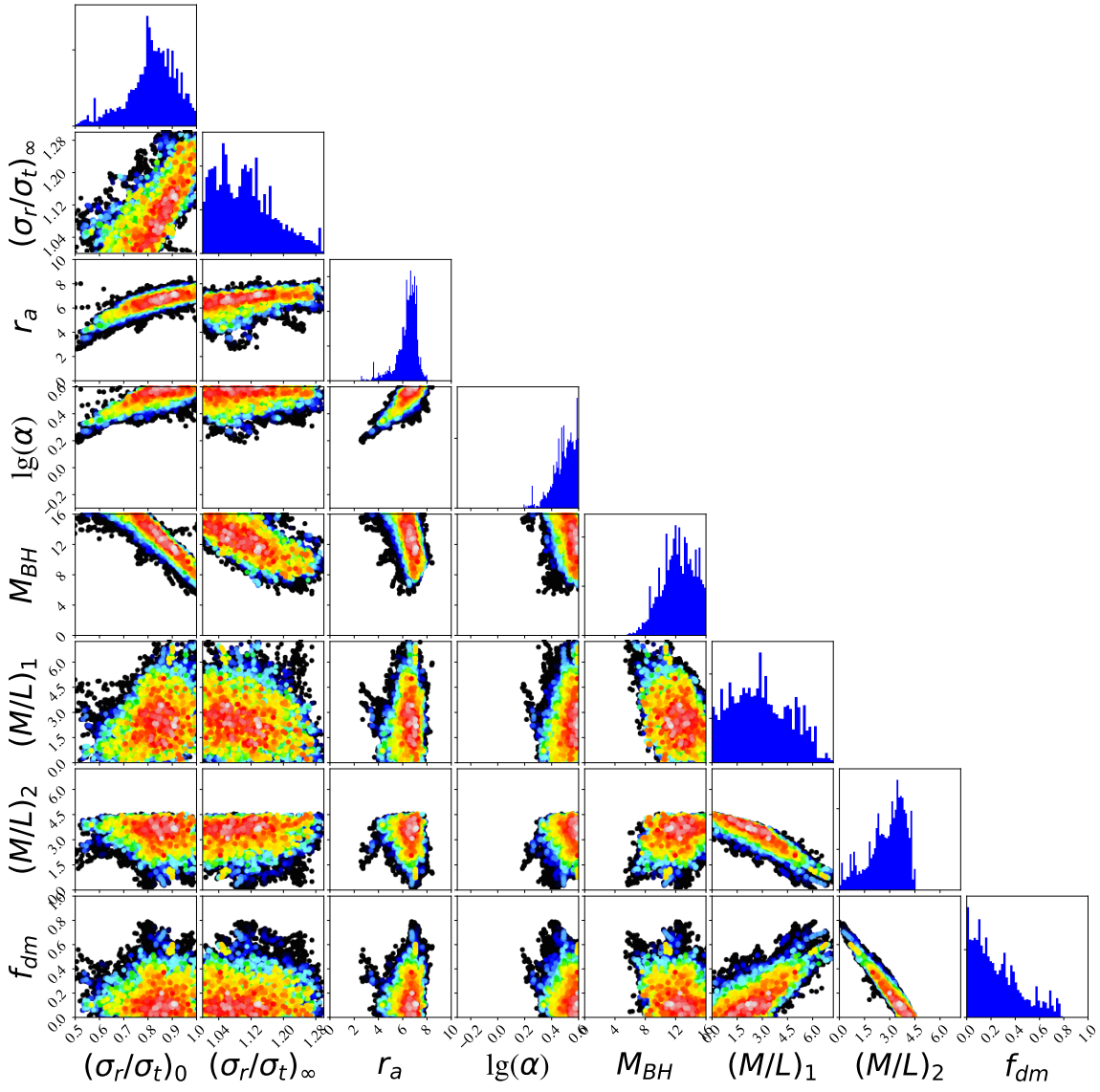

Figure 2.16: Corner plot for the individual SAURON data set. The range of allowed parameters differs substantially from the individual MUSE and OASIS datasets.

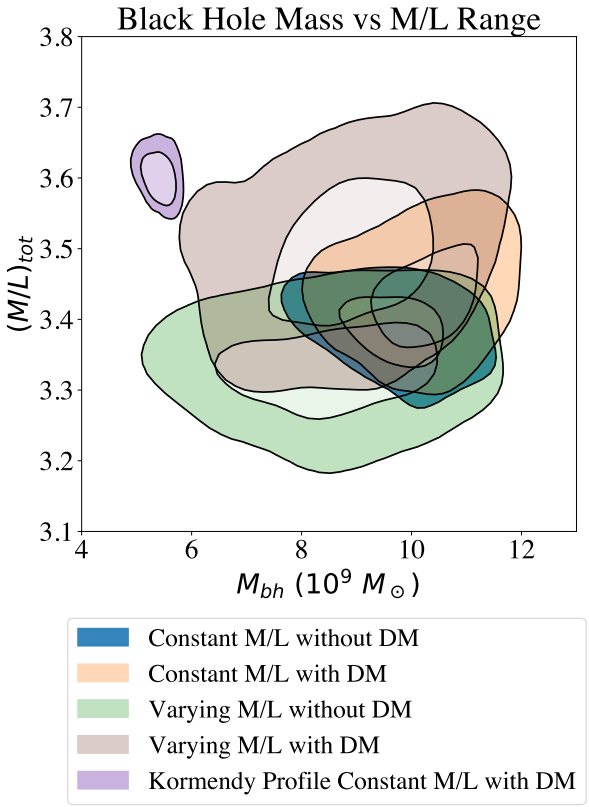

Figure 2.17: Plot of the 1 and 3σ confidence intervals for black hole mass and mass to light ratio within a sphere of one half light radius for each of the scenarios shown in Figure 2.21, Figure 2.22, Figure 2.23, and Figure 2.24.

of kinematics and models. The values obtained using the most general model (the median of the posterior) varies between $8.3-8.9\times10^9$ $M_\odot$. The errors are asymmetric and range in magnitude between 1.2 to 1.6. Using the kinematics derived from our preferred model with the RNI spectra gives $M_{\rm BH} = (8.7 \pm 1.2 \pm 1.3) \times 10^9$ $M_\odot$ where we calculate the second error to be half the difference between the largest and smallest black hole mass across all kinematics and models (i.e. the largest and smallest values in Table 2.4). Note that restricting this just to the most general model across each choice of kinematics decreases the modelling and kinematic systematic error to 0.3.

The final allowed range of black hole masses depend strongly on the model assumptions. The most simple model with only constant M^*/L has the smallest range of permitted black hole masses. Including dark matter slightly shifts this range to the right. This is because the SAURON data strongly constrains the kinematics at large radii, so there is a tight correlation between $f_{\rm dm}$ and M^*/L which allows one to interchange stars and dark matter. Decreasing stars at large radii in favor of dark matter, however, also results in a decrease in the stellar mass at the center of the galaxy which must be compensated with an increased black hole mass. Throughout this the anisotropy parameters remain fairly constant. This changes significantly with the introduction of varying M^*/L. In the model with a M^*/L variation and no dark matter, this results in a much larger range of viable black hole masses at the lower mass end. This is due to a correlation between $(M/L)_1$ and the black hole mass which effectively results in the mass of the black hole being exchanged for mass in stars. The results for the black hole mass are similar for the most general model featuring both varying M/L and DM.

The most interesting difference compared with previous studies is that the black hole mass we measure is more massive than that found in previous stellar dynamical studies. One key difference between our analysis and previous work is that we directly measure the stellar distribution within the influence of the AGN and find the stellar profile to be flatter than that used in all previous black hole studies of M87 using stellar dynamics. In order to determine if this could explain the increase in the black hole mass, we performed an MCMC run using the stellar profile of Kormendy et al. (2009) as this is used for the tracer distribution in previous studies (Gebhardt & Thomas, 2009; Gebhardt et al., 2011; Liepold et al., 2023a). We do our modelling with the RNI spectra and in the model with DM and constant M/L, as this most closely approximates the models used in Gebhardt & Thomas (2009) and Gebhardt et al. (2011). The 1 and 3σ posteriors for this are shown in Figure 2.17. We find a preferred black hole mass of $M_{\rm BH} = (5.5^{+0.5}_{-0.3}) \times10^9$ $M_\odot$, which agrees much more closely with

previous measurements which range between 5.4×10^9 $M_\odot$ (Liepold et al., 2023a) to 6.2×10^9 $M_\odot$ (Gebhardt et al., 2011). The equivalent model using our stellar distribution gives a black hole mass of $M_{\rm BH} = (10.2^{+0.8}_{-1.0}) \times 10^9$ $M_\odot$. Naively one might think that the difference in the black hole mass is purely due to the difference in the gravitational potential between the two models. On its own, our stellar profile should increase the measured black hole mass over previous determinations as it is exchanging the mass of the stars in the central regions of the galaxy with the mass of the black hole. However, one can calculate the decrease in stellar mass in the centre of the galaxy between our model and the model of Kormendy et al. (2009) and we find that, assuming a constant stellar M/L ratio of 3.4 without DM, that within $5''$(approximately the sphere of influence) the decrease in stellar mass is 5.6×10^7 $M_\odot$, or around 1 per cent of the total stellar mass within $5''$. This implies that the modification to the gravitational potential due to our model cannot be the sole cause of the difference between the two results.

One possible alternative is that this is a result of the very flat core. As pointed out by Kormendy & Ho (2013, sec. 3.1), measuring BHs with stellar dynamics in galaxies with flat cores is intrinsically less accurate than in galaxies with steep inner profiles, because of the increase importance of orbital anisotropy. Additionally, in centrally cuspy galaxies, the line of sight velocity distribution along the photometric centre of the galaxy receives its largest contribution from the three dimensional origin of the galaxy. In the case of a very flat core, the line of sight velocity distribution along the photometric centre receives an even contribution from a larger range of radii. This further increases the uncertainty in the kinematic modelling (note the much larger uncertainties using our profile over the Kormendy et al. (2009) profile) and has the potential to impact the extracted kinematics. Furthermore, a steeper profile implies more stars near the black hole, where the intrinsic sigma is higher. This leads to a higher observed sigma after projecting along the line-of-sight. In order to better demonstrate this effect, we show a plot of the model dispersion assuming our best fit parameters but using the stellar distribution from Kormendy et al. (2009) as opposed to our stellar distribution in Figure 2.18. We find that, given the same set of parameters, using the old stellar distribution leads to a sharp rise in the dispersion within $1''$. This suggests that in order to fit the data, previous studies required a smaller black hole mass with a more radially biased anisotropy closer to the black hole. This is exactly what we see in Figure 2.19 where we plot our anisotropy and that of Gebhardt et al. (2011). Further study of this will be required in future work.

2.7.2 Mass to Light Ratio Constraints

One important feature in this work is the inclusion of stellar M/L variations. In Figure 2.20 we show 1000 M/L profiles randomly sampled from the posterior of the most general model using the RNI data. We find a strong preference for an increasing M/L ratio in the central regions of the galaxy. Our best fit M/L variation agrees with the bottom end of figure 11 of Sarzi et al. (2018) assuming a Kroupa IMF. This result agrees with another analysis that dynamically measured the M/L variation in M87 and found it to be lower than that in Sarzi et al. (2018) (Parikh et al., 2024). In fact, a number of galaxies have been found to have stellar population determinations of their M/L that are systematically larger than those from dynamics (Parikh et al., 2024). The cause of this is poorly understood, but is likely to do with limitations of stellar population models, as the dynamical approach relies solely on gravitational physics and thus creates a smaller range of openings for systematic errors. Additionally, we also see that the right hand side of the figure has unrealistically small M/L. This suggests that the parameterisation of the M/L variation may not be as accurate in this region of the galaxy. Regardless, we do find that nearly every model prefers a radially increasing stellar M/L ratio. Previous studies have either failed to take M/L variations into account (Gebhardt & Thomas, 2009; Gebhardt et al., 2011), or directly used the profile measured in Sarzi et al. (2018) (Liepold et al., 2023a). Our result suggests that it is important to treat the profile as a free parameter.

One important limitation of these results is the fact that the M/L variation parametrization we use does not permit particularly steep variations due to it fixing the M/L ratio at 1″and 30″. To test the effect of this, we ran a model where the inner radius at which the M/L ratio becomes fixed is free to vary. We find a preference for a range of values between 1″and 4″but do not find significant changes in the distribution of M/L profiles in the range of the Sarzi et al. (2018) data. Our work thus establishes a preference for a M/L gradient at the lower end of what was reported in Sarzi et al. (2018). However, we caution that the recovered M_*/L profile is degenerate with the slope of the dark matter profile and the one we derive relies on a fixed NFW halo. Obviously no dynamical model can distinguish between a variation in the total density due to the stellar M_*/L or the dark matter without assumptions.

The total M/L ratio (defined as M/L within one $R_{\rm e}$) that we measure exhibits some model dependence. In Figure 2.17, we see that the total M/L ranges between 3.2 up to 3.7 depending on the model assumptions. This range of values is primarily due to the uncertainty in the anisotropy profile at large radius. The model with the largest total M/L is the model with varying M/L and DM. We see in Figure 2.22 that

the preferred value of $(\sigma_r/\sigma_t)_\infty$ is smaller than in the other models. This decrease in anisotropy at large radius must be made up by an increase in the mass. This further highlights the importance of the mass-anisotropy degeneracy when studying M87.

The total M/L ratio of M87 in the I-band has been previously measured in Cappellari et al. (2006) and in the r-band from Cappellari et al. (2013) (using mass follows light models in both cases). Converting these to the SDSS-z band (as a proxy for the ACS/WFC F850LP band) gives M/L ratios of 4.2/4.8[15], and 4.0. This is slightly larger than our range of M/L. In both previous determinations, the models were fit to the data out to R≈35″, while we only fit the data out to R=15″. The slightly larger total M/L of previous determinations may be explained as due to both their smaller adopted BH and an increase in the dark matter fraction between these radii. Gebhardt & Thomas (2009) and Murphy et al. (2011) also present measurements of the V-band stellar M/L ratio of 6.3 and 8.2. Converting these to SDSS-z band gives 2.8 and 3.7, respectively. These cannot be directly compared to the results in Figure 2.17 as we present the total M/L rather than just the stellar M/L. However, we can still conclude that a stellar M/L of 3.7, after including dark matter, will lead to a total M/L that is slightly above the range of what we have determined. Likewise, a stellar M/L of 2.8, combined with a dark matter fraction of ~ 15 will result in a total M/L of ~ 3.2, which is at the lower end of what we measure.

In this work we measure the stellar distribution within the influence of the AGN and find the stellar profile to be flatter than in previous work. The difference to the enclosed mass within 5″is close to 1%, implying that this does not significantly modify the gravitational potential. However, M87 is a large galaxy where the AGN covers only a small fraction of the stellar distribution relative to the size of the black hole. For other galaxies, such as NGC 4151, one of the key uncertainties in the black hole mass determination is uncertainty on the cuspiness of the inner stellar distribution (Roberts et al., 2021). Applying this technique to that case or similar cases could significantly reduce the uncertainties in the final black hole mass measurement.

2.7.3 Dark Matter

In this work we assume a NFW dark matter halo with break radius equal to 20kpc and find, in our most general models, a preference for a dark matter fraction within one effective radius of around $f_{\rm dm}(< R_{\rm e}) \approx 0.2$. This closely agrees with a result from Murphy et al. (2011) which determined the dark matter fraction within one

[15] $(M/L)_{\rm tot,z} = 4.2$ is determined from isotropic Jeans modelling whereas $(M/L)_{\rm tot,z} = 4.8$ is determined from Schwarzschild modelling

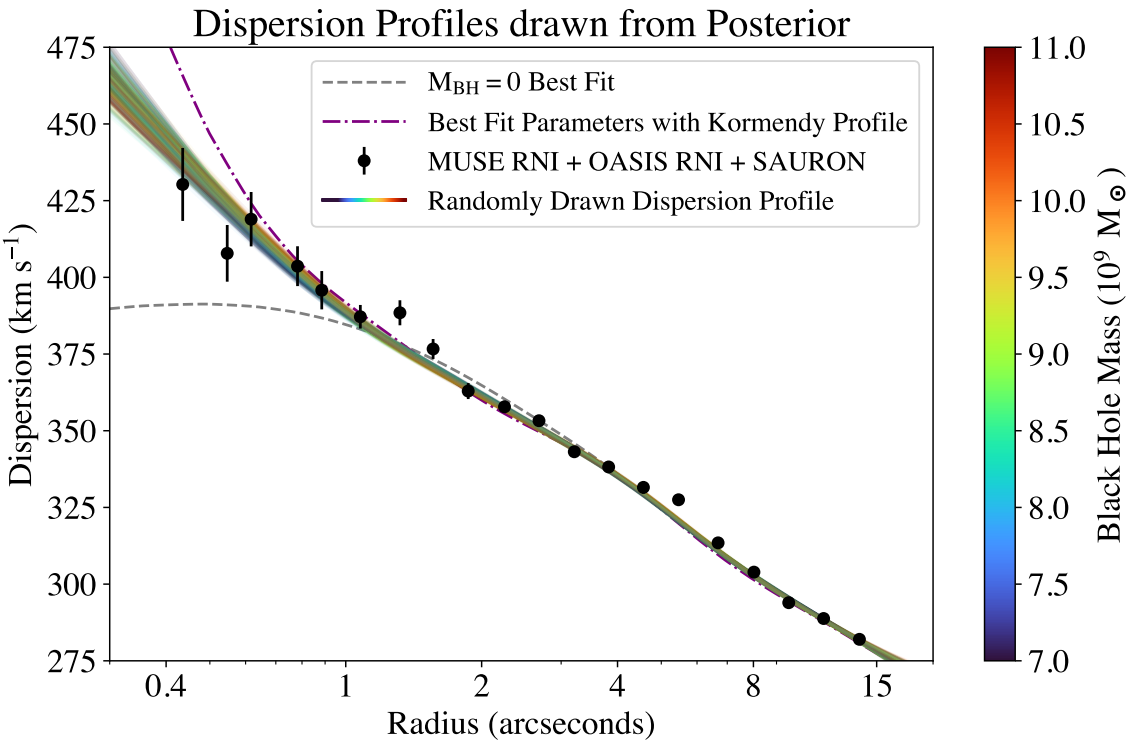

Figure 2.18: Plot of 1000 dispersion profiles randomly sampled from the MCMC chain with MUSE RNI + OASIS RNI + SAURON and colored according to their supermassive black hole mass. This is compared with a set of binned data points combining the MUSE RNI, OASIS RNI, and SAURON data. We also show a model with no black hole mass. It is technically possible to fit the data with no black hole mass if there is a highly radial anisotropy in the centre. To ensure this does not happen, we fix $r_a = 1$ and find the best fit model enforcing $M_{\rm BH} = 0$. The dotted and dashed purple line shows the dispersion after exchanging our stellar distribution with the one from Kormendy et al. (2009) but still using the parameters from the best fit model assuming our stellar distribution. You can see that the profile becomes much steeper in the center. The way to remedy this is by decreasing the black hole mass to the value determined in previous studies and adjusting the anisotropy correspondingly.

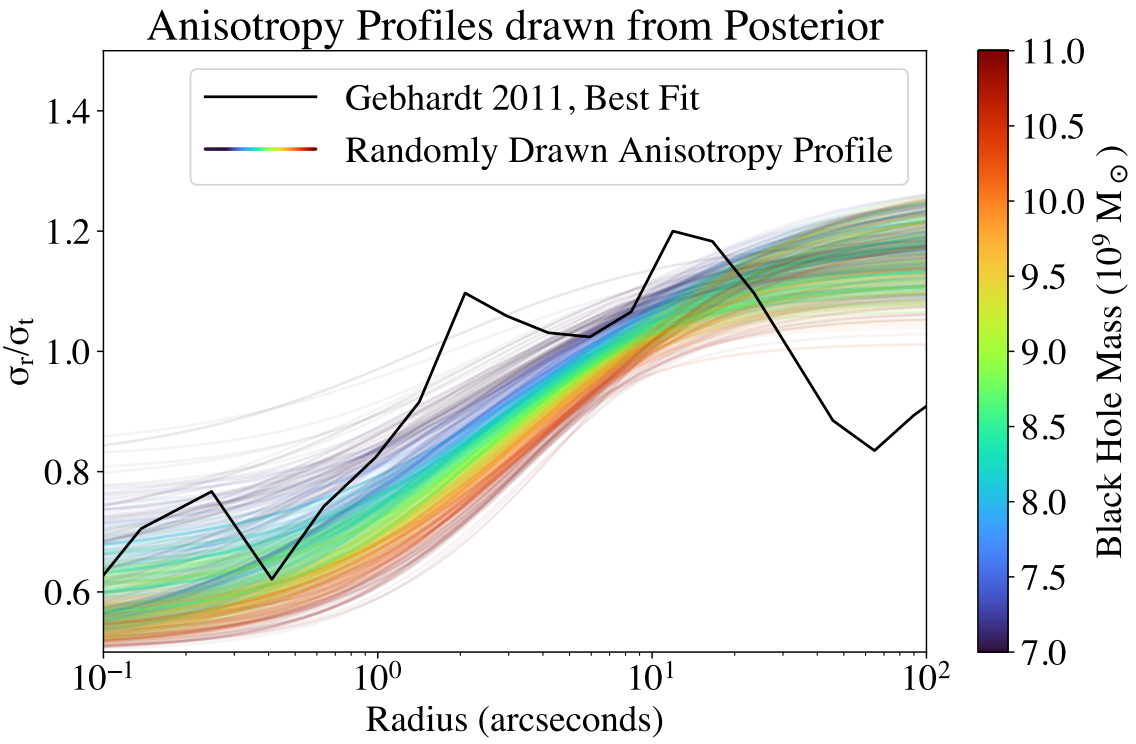

Figure 2.19: Plot of 1000 anisotropy profiles randomly sampled from the MCMC chain with MUSE RNI + OASIS RNI + SAURON and colored according to their supermassive black hole mass. The best fit anisotropy profile from Gebhardt et al. (2011) is shown in black. We find that all models prefer a radially increasing anisotropy ratio though there is considerable scatter due to the mass anisotropy degeneracy.

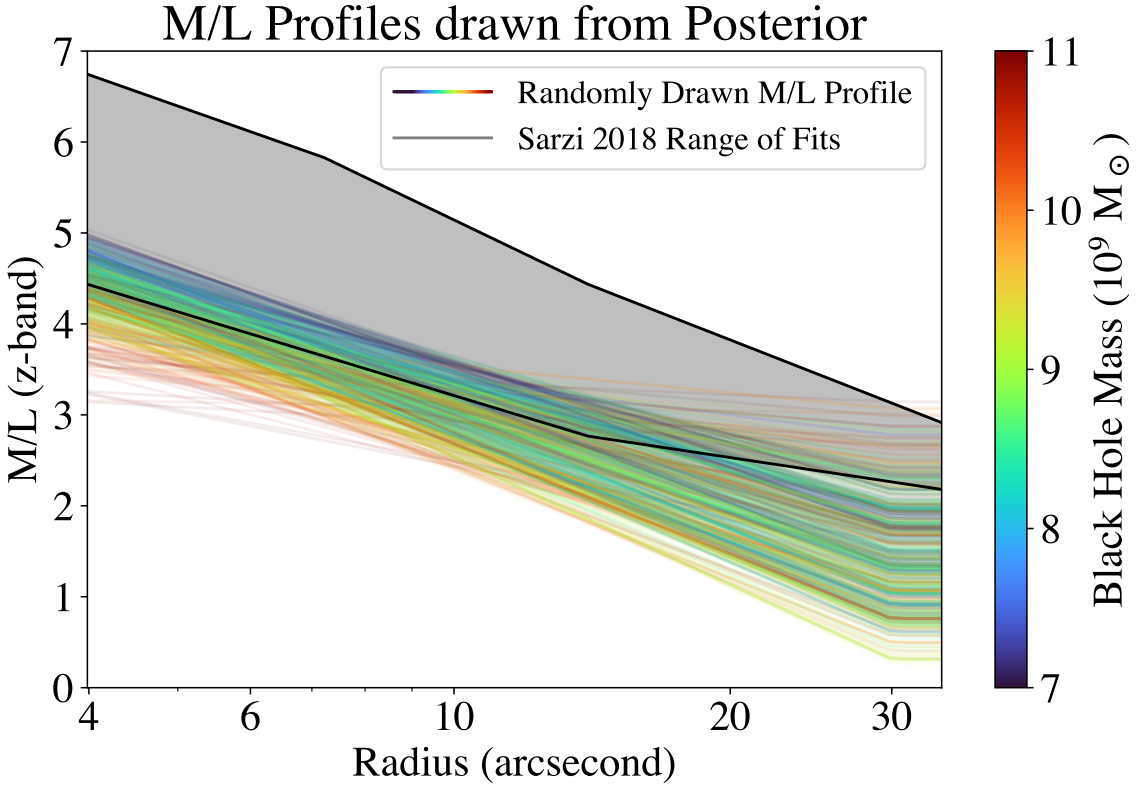

Figure 2.20: Plot of 1000 M/L profiles randomly sampled from the MCMC chain with MUSE RNI + OASIS RNI + SAURON and colored according to their supermassive black hole mass. This is compared with range of M/L variations assuming Kroupa IMF from Sarzi et al. (2018) which is shown in gray and outlined with black lines. Our data strongly prefers an increasing M/L variation towards the centre of the galaxy consistent with that in Sarzi et al. (2018).

effective radius to be approximately 17%. Our results, however, are strongly model and data dependent. In the less general models we consistently find a preference for no or very little dark matter (see Figure 2.23). Additionally, on their own, the MUSE, OASIS, and SAURON data do not have a preference for a dark matter halo (Figure 2.14,Figure 2.15, and Figure 2.16). This data only goes out to 1.2 kpc, so we do not expect to very strongly constrain the dark halo. This highlights the importance of including large scale kinematic data for constraining information on the dark matter halo.

2.7.4 Anisotropy Profile Constraints

In Figure 2.19 we show 1000 anisotropy profiles randomly chosen from the posterior of the NFW DM + Varying M/L model using the RNI spectra. We find a remarkable agreement with previous work. The profiles all display the radially increasing behavior expected of slow rotators. We also visually see that the profiles tend to transition from constant on the right hand side to lower values near $10''$. This is close to the size of the core of M87 ($5\farcs66$ according to Lauer et al. (2007)), and agrees with the results of previous studies demonstrating that the size of the core in slow rotators is close to the radius at which the velocity anisotropy ratio becomes tangential (Thomas et al., 2014). Another observation we should make is the strong correlation between the black hole mass and the anisotropy profile in Figure 2.19. This clearly demonstrates the strong role that the mass-anisotropy degeneracy plays in this analysis.

One important comment is that these results depend on our use of physically motivated priors. We see in Figure 2.21, Figure 2.22, Figure 2.23, and Figure 2.24 that in many cases, the posteriors for $(\sigma_r/\sigma_t)_0$ and $(\sigma_r/\sigma_t)_\infty$ run into the imposed boundaries and thus are unable to explore the full range of parameter space capable of reproducing the data.

2.8 Conclusion and Future Prospects

We have studied the galaxy M87 using stellar kinematics from SAURON, OASIS, and MUSE using the code JamPy and our primary conclusions are as follows:

- The stellar distribution of M87 can be measured directly within the influence of the AGN. This is done by directly measuring the fraction of the spectral flux due to stars during the kinematic extraction. The shape of the stellar distribution

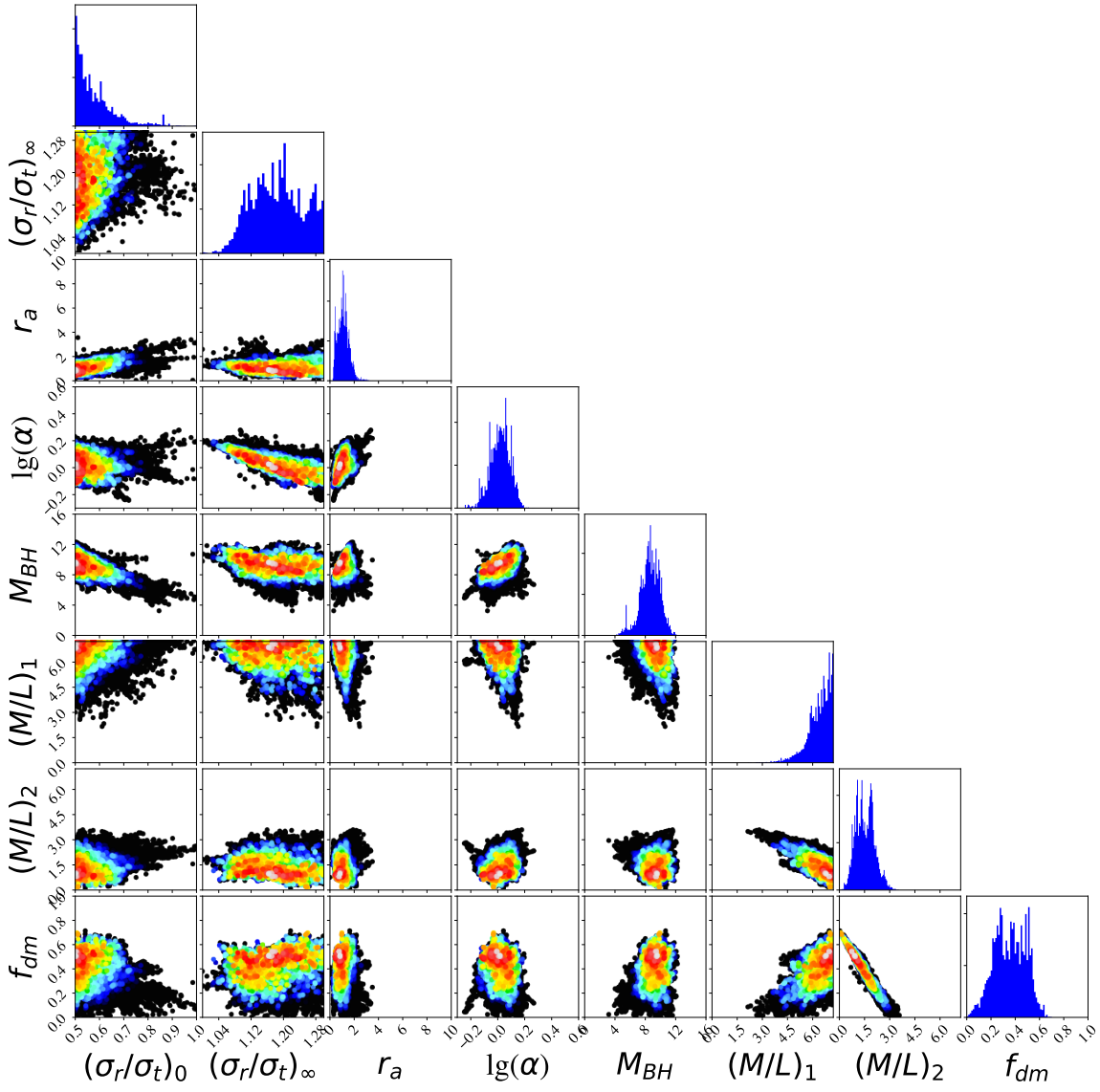

Figure 2.21: Corner plot for the joint MUSE RNI + OASIS RNI + SAURON data for the model with varying M/L and a DM halo.

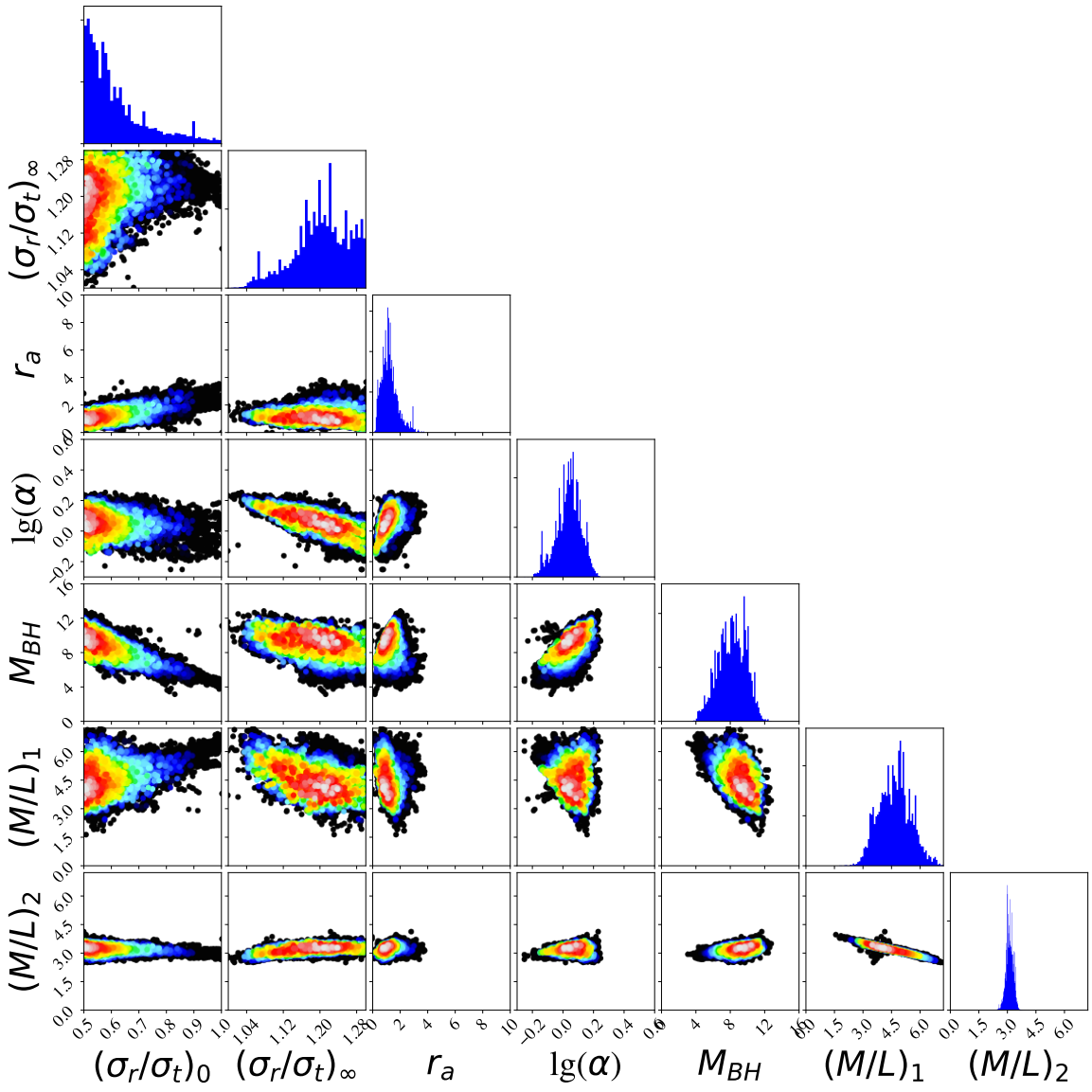

Figure 2.22: Corner plot for the joint MUSE RNI + OASIS RNI + SAURON data for the model with no dark halo but varying M/L.

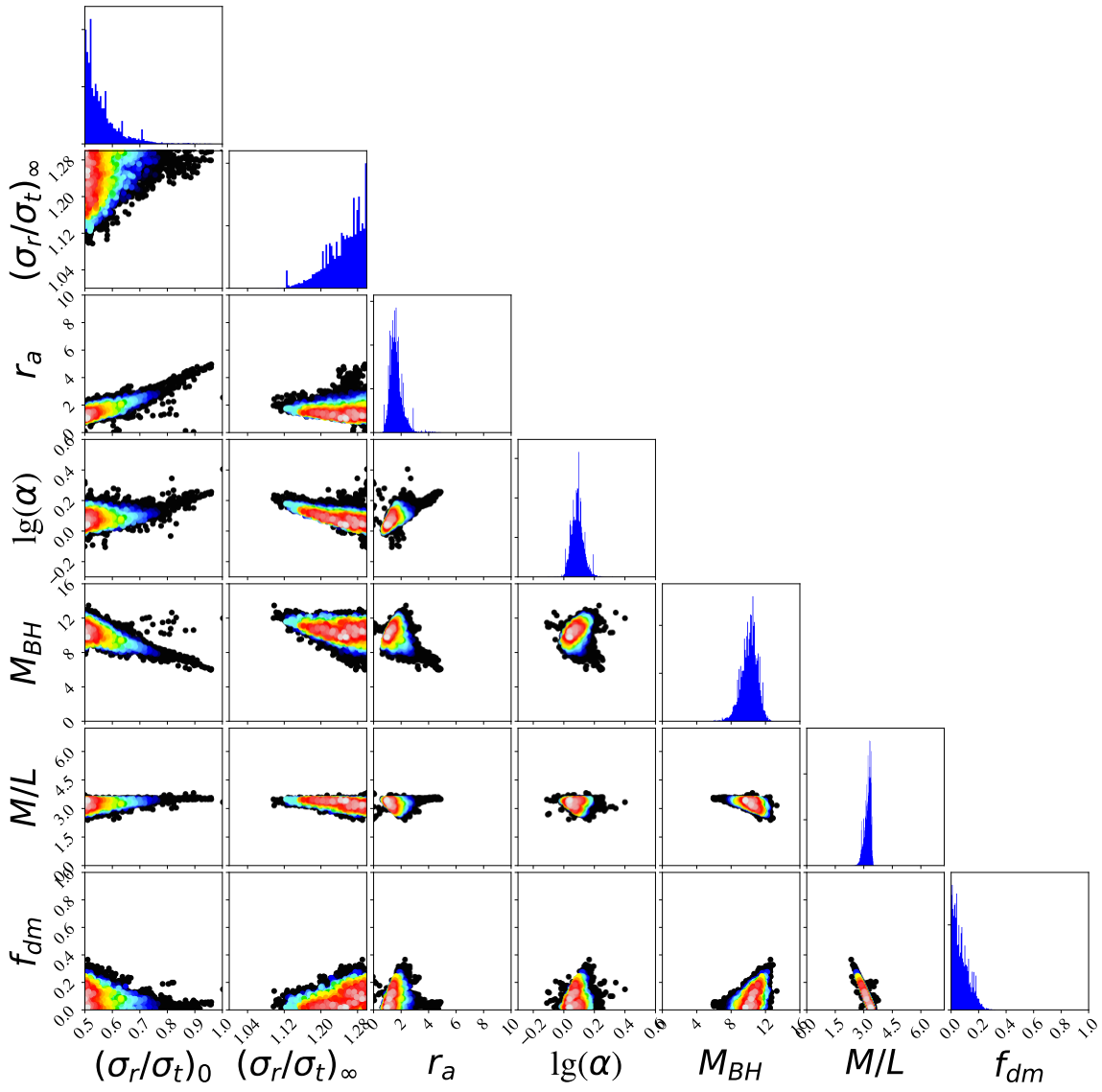

Figure 2.23: Corner plot for the joint MUSE RNI + OASIS RNI + SAURON data for the model with constant M/L and a DM halo.

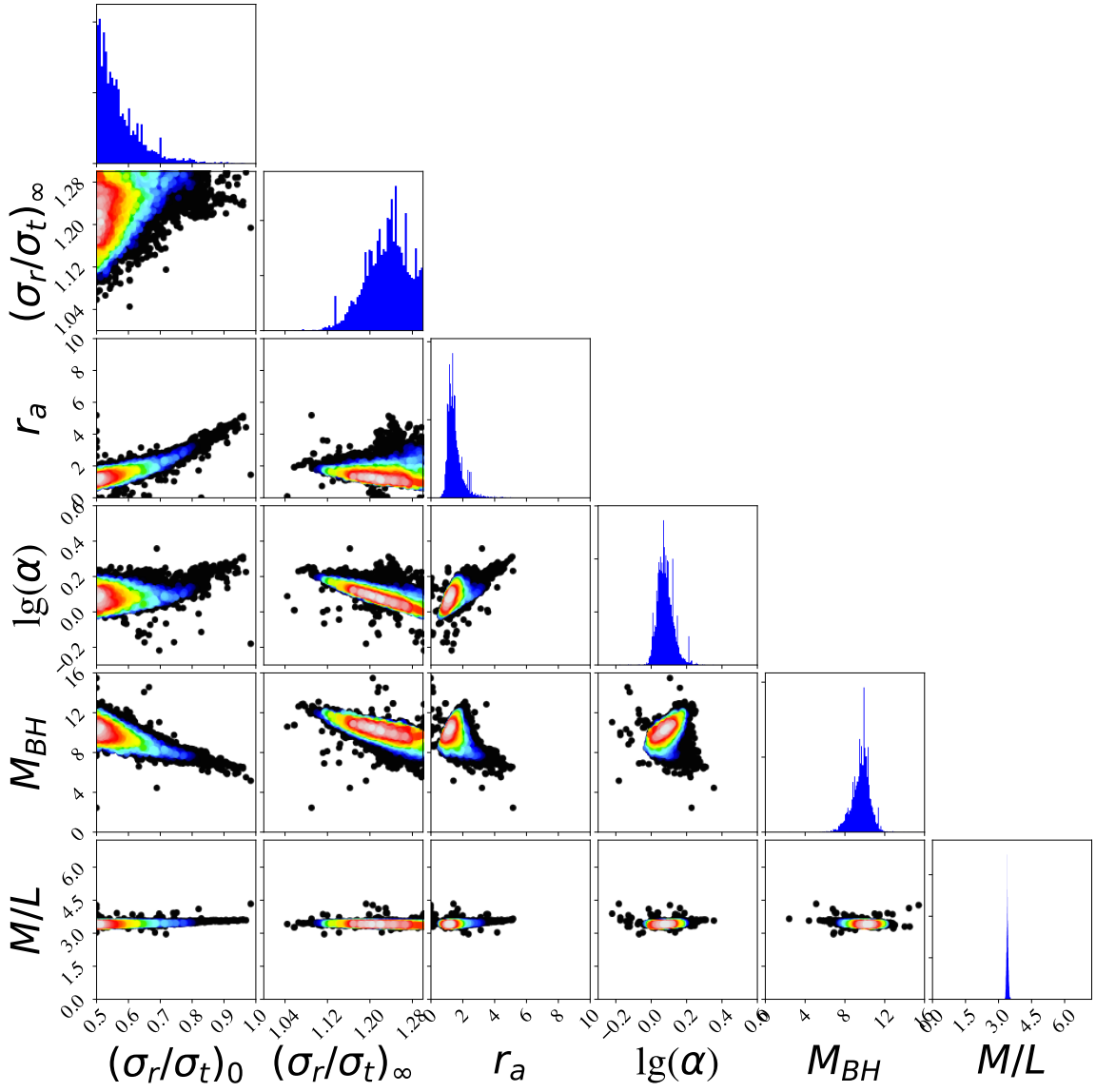

Figure 2.24: Corner plot for the joint MUSE RNI + OASIS RNI + SAURON data for the model with a constant M/L and no dark matter.

profile used in previous studies (Kormendy et al., 2009) overestimates the the stellar density in the central region of M87 by a factor of 2 (see Figure 2.10).

- For galaxies with an AGN, the PSF can be accurately measured in integral field spectroscopy by measuring the continuum flux during the kinematic extraction (see Figure 2.9 and Table 2.1). This is due to the fact that the AGN spectral contribution is thought to be smooth, and hence is well approximated by additive polynomials. We measure the FWHM of the OASIS PSF to be $0\farcs561$ and the FWHM of the MUSE PSF to be $0\farcs049$. These are consistent with what we expect from the seeing for OASIS and the AO capabilities of MUSE.

- We use JAM dynamical models of the kinematics in a Bayesian fashion. We find a preferred black hole mass of $M_{\rm BH} = (8.7 \pm 1.2 \pm 1.3) \times 10^9\ M_\odot$ with the second error showing the modelling and kinematic systematic uncertainty. This range is consistent with the EHT measurement and previous stellar dynamical models of M87, though with a distinct preference for a larger black hole mass. Our analysis also highlights the fact that, even with excellent data, the derived black hole mass is sensitive to a variety of assumptions on both the kinematic extraction and the M/L variation. The resulting uncertainties, when accounting for these systematics, are significantly larger than usually adopted.

- We find a strong preference for a radially decreasing M^*/L ratio at the lower end of what is found in Sarzi et al. (2018). This has the effect of expanding the range of allowed black hole masses to lower values.

- We measure the anisotropy profile of M87 assuming a new flexible analytic parametrization for the anisotropy which is a logistic function of logarithmic radius and find a strong preference for a radial increase. This also clearly shows the mass-anisotropy degeneracy which strongly contributes to the uncertainty in the black hole mass. We conclude that, contrary to what is sometimes assumed, one can obtain stringent constraints on both black hole masses and on the anisotropy profile, using the Jeans equations, by combining priors on the anisotropy, which is now well-understood in galaxy centres, with realistic parametrizations for the total density.

There remain many important questions about M87 that are well suited to be studied in the near future. Recent work has suggested a number of different ways that improved modelling of the gas disk is able to bring the supermassive black hole

measurements from gas dynamics into agreement with those from stellar kinematics and the EHT (Jeter et al., 2019; Jeter & Broderick, 2021; Osorno et al., 2023). This, however, relies on resolving details of the gas kinematics within the innermost arcsecond of the galaxy. New and improved datasets could be used to differentiate between these scenarios.

Furthermore, as this work shows, future studies of the supermassive black hole mass of M87 using stellar kinematics will also require detailed studying of systematic effects in order to produce reliable black hole mass results. High quality data for these tasks may not be far off. There is a cycle 1 JWST proposal (2228, PI: Jonelle Walsh) to measure the central supermassive black hole of M87 using NIRSpec. This will cover a wavelength range including the CO bandhead that is similar to the wavelength range covered in Gebhardt et al. (2011) though it will be unaffected by skylines and will have much higher spatial resolution and signal to noise. This will provide the most detailed view of M87's inner kinematics to date.

Chapter 3

Recovery of Kinematics without Systematic Effects

3.1 Summary

Kinematics play a fundamental role in understanding galaxies. I investigate the recovery of the kinematics from an ideal N-body simulation and find that, even in the ideal scenario, the second velocity moment $V_{\rm rms}$ depends in a non-trivial way on higher-order Gauss-Hermite moments h_i. We find that the typical systematic error in $V_{\rm rms}$ when not fitting higher-order moments is 5 per-cent. This drops to about 1 per-cent when including higher-order moments, and doesn't appreciably improve until a significantly higher number of moments are included.

3.2 Introduction

Understanding the kinematics of galaxies has played a fundamental role in progressing our knowledge of the Universe. The first spectrum of an external galaxy revealed that galaxies are assembled out of stars (Scheiner, 1899). The first redshift showed that galaxies moved with an order of magnitude more velocity than the stars in the local Universe (Slipher, 1913). And the first rotation curves (only within the innermost parts of the galaxy) revealed that galaxies rotate with immense speed (Slipher, 1914; Pease, 1916). More recent measurements of the rotation curves of spiral galaxies have shown that the Universe is not dominated by visible matter but by dark matter (Rubin & Ford, 1970; Rubin et al., 1978).

The early measurements of rotations in external galaxies used gas. This has the advantage of typically being quite bright. This poses a problem for early-type galaxies, however, which often have little to no gas. This problem was first properly

addressed in Bertola & Capaccioli (1975), which used an image-tube spectrograph to unveil the first rotation curve of an elliptical galaxy. Importantly, the kinematics were measured not from gas emissions, but from stellar absorption features (Ca II H and K). Since that time, integral field spectroscopy has revolutionised the field by allowing simultaneous two-dimensional spectra to be measured at a wide variety of observing facilities (see (Bacon, 2024) for a review).

The primary information extracted from spectroscopic data for stellar dynamical modelling is the line-of-sight velocity distribution (LOSVD). This was first done under the assumption of a Gaussian LOSVD (Simkin, 1974; Sargent et al., 1977; Tonry & Davis, 1979; Franx & Illingworth, 1988; Bottema, 1988). Observations of early-type galaxies revealed structures such as kinematically decoupled cores, showing that a Gaussian LOSVD is not sufficiently general, leading to updated versions of these methods that could handle a non-Gaussian LOSVD (Bender, 1990; Rix & White, 1992; Statler, 1995). A key breakthrough came with the use of the Gauss-Hermite (GH) series (van der Marel & Franx, 1993; Gerhard, 1993). This treats the LOSVD as approximately Gaussian, but includes higher-order terms to quantify deviations from pure Gaussianity. The GH series is attractive because it expresses the LOSVD as a sum of orthogonal basis functions. As with any complete orthogonal expansion, this guarantees that an arbitrary, sufficiently regular LOSVD can be reproduced to arbitrary accuracy when a sufficiently large number of terms is included.

Other approaches to parameterising the LOSVD exist. So-called "non-parametric" (e.g. Falcón-Barroso & Martig (2021),Mehrgan et al. (2019, 2023)) approaches parameterise the LOSVD using a large number of box functions or B-splines to create a histogram that approximates the LOSVD. It is important to note that both methods can approximate a LOSVD to arbitrary accuracy. For the GH series, this is achieved by including higher-order moments and for the so-called "non-parametric" methods this means including more basis box functions. We anticipate them to perform identically well on ideal data given sufficiently many terms are included.

Recently, there has been an increased interest in the role that using partial sums of the GH series has on measured stellar kinematics and dynamical modelling. Liepold et al. (2020) argued that moments up to h_{12} should be included in dynamical modelling, arguing that only including up to h_4 leads to weak constraints on the supermassive black hole mass and, in some codes, can result in a significant systematic offset in the black hole mass. Observationally, Mehrgan et al. (2023) measured the LOSVD of early-type galaxies with MUSE observations and found, using the non-parametric spectral fitting code WINGFIT (Mehrgan et al., 2019, 2023), broad wings that extend

out beyond what is usually captured by the lowest-order GH moments. Reiter et al. (2025) performed a reanalysis of SAURON galaxies using both the GH series and the non-parametric code BayesLOSVD (Falcón-Barroso & Martig, 2021) and found significant differences in the total stellar mass for a galaxy with counter-rotation, but small differences for a regular rotator.

Our goal is to better clarify the situation by systematically studying the impact of using finite truncations of the GH series for stellar kinematics and dynamical modelling of a high-resolution N-body galaxy that resembles a slow-rotator. We do this with both JAM modelling (Cappellari, 2008, 2020) and Schwarzschild modelling with the axisymmetric code of (Cappellari et al., 2002, 2006) which uses some code of van der Marel et al. (1998) as described in Section 3.4 of Cappellari et al. (2006), and the triaxial code DYNAMITE[1] (van den Bosch et al., 2008; Thater et al., 2022b). This code is primarily a Python wrapper for the code of van den Bosch et al. (2008) with a number of bug fixes and improvements for user-friendliness.

This chapter is structured as follows: in section 3.3 we describe the N-body simulation and initial conditions used to generate our slow rotator galaxy. In section 3.4 we provide some background on the Gauss-Hermite series and discuss some of its features and drawbacks. In section 3.5 we discuss in detail how we extract the GH parameters from our simulation and some of the limitations associated with this. We summarise our results in section 3.6.

3.3 N-Body Simulation

We use the simulation of Rantala et al. (2018) (which has the benefit of having previously been studied in Neureiter et al. (2023); de Nicola et al. (2022)). This uses the code GADGET-3 (Springel, 2005) with the black hole dynamics module KETJU (Rantala et al., 2017). Ordinarily, the stellar dynamics within the vicinity of the supermassive black hole in N-body simulations is unreliable due to the gravitational softening length being of the order or larger than the black hole sphere of influence. KETJU solves this problem by creating a region around each supermassive black hole with radius r_{KETJU} within which the orbits of the stars are accurately integrated without any softening. This stands in contrast to the rest of the galaxy where the orbits are integrated using a softened gravitational force calculated with the tree algorithm.

[1]The original code by van den Bosch can be found at `https://github.com/remco-space/TriaxSchwarzschild`. The updated code by Thater et al. is available at `https://github.com/dynamics-of-stellar-systems/dynamite`.

Table 3.1: Table of initial conditions for the progenitor galaxies.

Parameter	Symbol	Value
Stellar Mass	M_*	$4.15 \times 10^{11} M_\odot$
Effective Radius	R_e	7 kpc
DM Halo Mass	M_DM	$7.5 \times 10^{13} M_\odot$
DM fraction	f_DM	0.25
Black Hole Mass	M_BH	$8.5 \times 10^{9} M_\odot$
Number of Stellar Particles	N_*	4.15×10^{6}
Number of Dark Matter Particles	N_DM	1.0×10^{7}
Stellar Particle Mass	m_*	1.0×10^{5}
Dark Matter Particle Mass	m_dm	7.5×10^{6}

The simulation is initialised in the following way: two identical progenitor galaxies are initialised consisting of a stellar bulge, a dark matter halo, and a supermassive black hole. Both the stellar bulge and the dark matter halo are distributed according to a Hernquist profile (Hernquist, 1990)

$$\rho(r) = \frac{M}{2\pi} \frac{a}{r(r+a)^3} \tag{3.1}$$

with $R_e \approx 1.8153 \times a$. They have parameters as shown in Table 3.1. These galaxies are then set on a nearly parabolic merger orbit with an initial separation given by 30 kpc. From these initial conditions, the merger is complete before $t \sim 300$ Myr.

The resulting galaxy has a stellar mass of $8.3 \times 10^{11} M_\odot$, a supermassive black hole mass of $1.7 \times 10^{10} M_\odot$, a black hole sphere of influence, defined as the radius that encloses a spherical stellar mass equal to the black hole mass, of 1 kpc, and a half-mass radius $R_{1/2}$ defined as the radius that encloses a spherical stellar mass equal to half of the total stellar mass of the galaxy, of 14.3 kpc. By design, this simulation produces a black hole that is overmassive with respect to the $M - \sigma$ relation. This is because slow rotator galaxies tend to be overmassive compared to the predictions from the $M - \sigma$ relation (de Nicola et al., 2025, Fig. 1). We calculate the shape of the resulting galaxy using the routine BIWEIGHT_MIDCOVARIANCE of ASTROPY and find relative axes lengths of (1, 0.65, 0.6). This is a triaxial geometry, though close to prolate axisymmetry. We calculate the inertia tensor and align the galaxy such that the long axis is along the x direction, the medium axis is along the y direction, and the short axis is along the z direction.

We plot the enclosed mass for the dark matter and stars in Figure 3.1. The shape of the dark matter component is best described using a double power law fit. We use

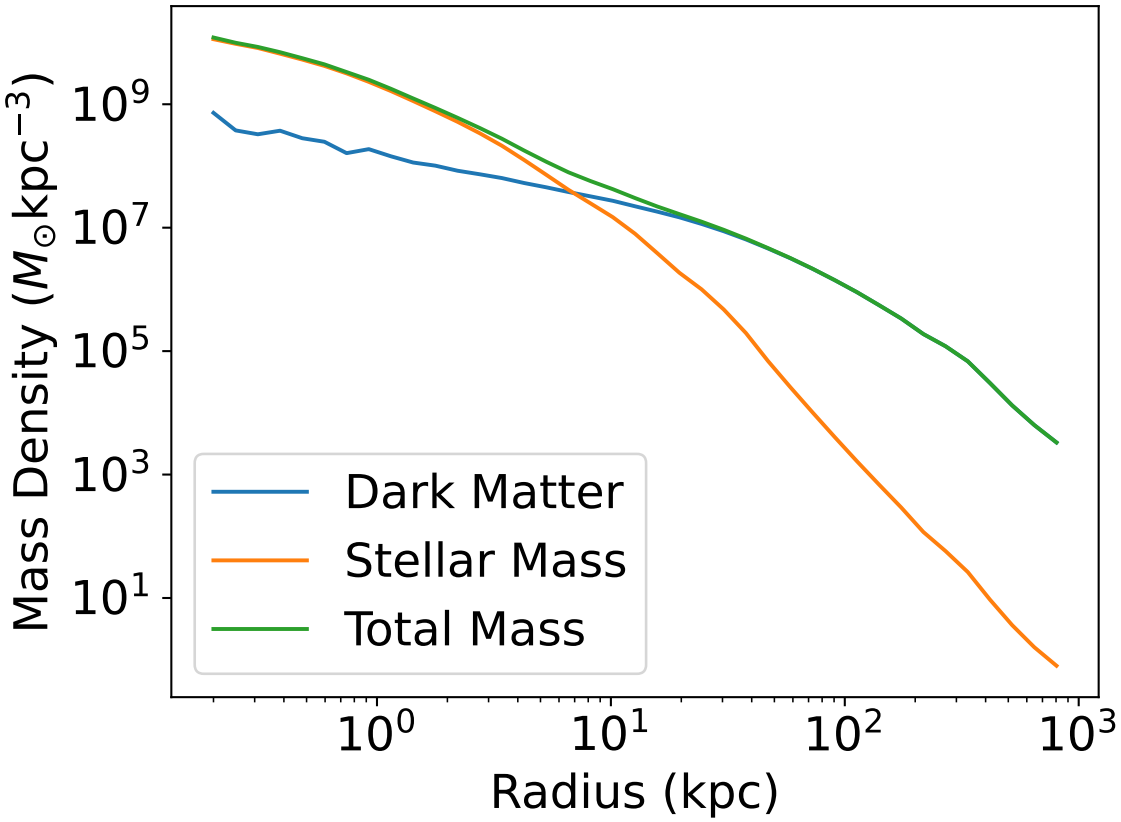

Figure 3.1: Plot of the 3D spherical density of the stars, dark matter, and total mass distribution in the simulation. The dark matter distribution is slightly more shallow than a NFW halo in the centre and drops off faster at large radii. This is due to the fact that the halo in the simulation is initialised as a Hernquist profile.

the Nuker profile (Lauer et al., 1995b) of the form

$$\rho(r) = \rho_0 \left(\frac{r}{r_b}\right)^{-\gamma} \left[1 + \left(\frac{r}{r_b}\right)^{\alpha}\right]^{\frac{\gamma-\beta}{\alpha}} \tag{3.2}$$

The best-fit recovery has an inner slope γ of 0.72 and outer slope β of 3.9, a break radius of 150 kpc, and α of 1.1. This halo is thus qualitatively different from a NFW (Navarro et al., 1997) or gNFW halo (Zhao, 1996; Wyithe et al., 2001). This shape is a result of the initial dark matter distribution being a Hernquist profile.

3.4 The Gauss-Hermite Series

3.4.1 Background

The LOSVD for most galaxies is approximately, but not perfectly, normally distributed. Deviations away from the Gaussian shape usually take the form of wings

of the distribution or asymmetries due to rotation. These higher-order moments of the LOSVD are often parameterised using a Gauss-Hermite series (van der Marel & Franx, 1993; Gerhard, 1993). In its most general form, a GH series can be written as

$$\mathcal{L}(v) = \frac{e^{-y^2/2}}{\sqrt{2\pi}\sigma} \sum_{m=0}^{\infty} h_m H_m(y) \tag{3.3}$$

where $y = (v - V)/\sigma$ and H_m are Hermite polynomials satisfying the relation

$$\int_{-\infty}^{\infty} H_m(y) H_n(y) \frac{e^{-y^2}}{2\pi} = \frac{\delta_{mn}}{2\sqrt{\pi}} \tag{3.4}$$

with δ_{mn} the Kronecker delta.

The convention we use is related to the more popular "physicist's convention" (the convention used for solutions to the quantum harmonic oscillator) via

$$H_m(y) = \frac{H_m^{\text{phys}}(y)}{\sqrt{m! 2^m}} \tag{3.5}$$

The first five terms under this convention are

$$\begin{aligned} H_0(y) &= 1 \\ H_1(y) &= \sqrt{2}y \\ H_2(y) &= \frac{1}{\sqrt{2}}(2y^2 - 1) \\ H_3(y) &= \frac{1}{\sqrt{3}}(2y^3 - 3y) \\ H_4(y) &= \frac{1}{\sqrt{24}}(4y^4 - 12y^2 + 3) \end{aligned} \tag{3.6}$$

One important mathematical feature of Hermite polynomials is that they form an orthogonal basis (under the above weighted inner-product) for the space of square-integrable functions. As we expect all LOSVDs to be square integrable, we are guaranteed that, in the limit that infinitely many terms are included in the series, Equation 3.3 will converge to any observed LOSVD. This is guaranteed due to the fact that Hermite polynomials are the solution of a Sturm-Liouville problem (Zettl, 2005), namely, the Schrödinger equation for a quantum harmonic oscillator.

Suppose $\mathcal{L}_0(v)$ is the true line-of-sight velocity distribution one wants to recover. In this case, the best-fit coefficients h_m in Equation 3.3 are given by fixing V and σ, multiplying by $H_n(y)\frac{e^{-y^2/2}}{\sqrt{2\pi}}$ and integrating to get

$$h_m = 2\sqrt{\pi} \int_{-\infty}^{\infty} \mathcal{L}_0(v) H_m(y) \frac{e^{-y^2/2}}{\sqrt{2\pi}} dv \tag{3.7}$$

It is worth emphasizing that V and σ can, in principle, be given any value and, provided enough terms in the Hermite series are included, Equation 3.3 will converge to $\mathcal{L}_0(v)$. However, it is practically much more convenient to choose V and σ to ensure that the series quickly converges to $\mathcal{L}_0(v)$. Practically, this means choosing V and σ such that the lowest-order non-trivial terms in the series (h_1 and h_2) vanish. This is satisfied when V and σ correspond to the best-fit Gaussian to $\mathcal{L}_0(v)$. To see this, note that the definition of the best-fit Gaussian minimizes the following χ^2

$$\chi^2(V,\sigma) = \int_{-\infty}^{\infty} \left(\mathcal{L}_0(v) - \frac{e^{-y^2/2}}{\sqrt{2\pi}} \right)^2 dv \tag{3.8}$$

This is minimized as a function of V and σ when

$$\frac{\partial \chi^2}{\partial V} = \frac{\partial \chi^2}{\partial \sigma} = 0 \tag{3.9}$$

and the corresponding second derivatives are positive. Evaluating the derivative with respect to V gives

$$\int_{-\infty}^{\infty} 2 \left(y\mathcal{L}_0(v) - y\frac{e^{-y^2/2}}{\sqrt{2\pi}} \right) e^{-y^2/2} dy = 0 \tag{3.10}$$

The second term equals zero by symmetry, leaving

$$\int_{-\infty}^{\infty} 2y\mathcal{L}_0(v) e^{-y^2/2} dy = 0 \tag{3.11}$$

As $H_1(y) \propto y$, this implies from Equation 3.7 that $h_1 = 0$. For the derivative with respect to σ we have

$$\int_{-\infty}^{\infty} \left(\mathcal{L}_0(v) - \frac{e^{-y^2/2}}{\sqrt{2\pi}} \right) \frac{e^{-y^2/2}}{\sqrt{2\pi}} \frac{y^2/2}{\sigma} = 0 \tag{3.12}$$

We can write $y^2 = H_2(y)/\sqrt{2} + 1/2$, and integrate to get

$$\frac{h_2}{2} = 0 \tag{3.13}$$

From which we see that $h_2 = 0$. This demonstrates an important fact often assumed when using the GH series. We can always find a V and σ such that $h_1 = h_2 = 0$. Practically, this means that we can set $h_1 = h_2 = 0$ and treat V and σ as free parameters and we still have the guaranteed mathematical property that the series will converge to the true solution given enough terms.

3.4.2 Normalization

As h_1 and h_2 can be set to zero, the Gauss-Hermite series is usually presented as

$$\mathcal{L}(v) = \frac{e^{-y^2/2}}{\sigma\sqrt{2\pi}} \left[1 + \sum_{m=3}^{M} h_m H_m(y)\right] \tag{3.14}$$

It is important to note that Equation 3.14 is not normalised if $h_m \neq 0$. In a typical case where $h_4 = 0.1$ and all other higher-order moments are zero there is an extra 6 per-cent of flux in Equation 3.14. In many cases this is of no practical importance as the determination of stellar kinematics from spectra in codes such as pPXF do not require the normalisation in order to recover the parameters of the LOSVD (van der Marel & Franx (1993) presents this method as a way to fit emission lines, and thus ignores the normalisation, preferring to use a factor γ that scales to the overall non-normalised flux). However, in the case where one is interested in the absolute weight of the stellar component in a spectral fit that contains higher-order GH moments, or one wants to calculate the higher-order velocity moments of the LOSVD, knowledge of the normalisation is of paramount importance.

We can express the normalisation as

$$C = \int_{-\infty}^{\infty} \mathcal{L}(v) dv \tag{3.15}$$

$$= \int_{-\infty}^{\infty} \frac{e^{-y^2/2}}{\sigma\sqrt{2\pi}} \left[1 + \sum_{m=3}^{M} h_m H_m(y)\right] dv \tag{3.16}$$

A complete calculation for this is given in section A.1. The result is

$$C = \left(1 + \sum_{m\in\{4,6,\ldots\}}^{M} h_m \frac{(m-1)!!}{\sqrt{m!}}\right) \tag{3.17}$$

with ! and !! denoting the single and double factorial, respectively. We thus write the normalised LOSVD as

$$\mathcal{L}_{\mathrm{norm}}(v) = \frac{e^{-y^2/2}}{C\sigma\sqrt{2\pi}} \left[1 + \sum_{m=3}^{M} h_m H_m(y)\right] \tag{3.18}$$

It is important to emphasise that the interpretation of this as a normalised probability distribution is challenged by the fact that the GH series, by nature of spanning the space of square-integrable functions, can include negative values. We discuss this issue in more detail in subsection 3.5.4.

3.4.3 Moments of the Gauss-Hermite Series

Jeans modelling describes the second velocity moment of the LOSVD. This is given by

$$V_{\rm rms} = \sqrt{\langle v^2 \rangle} = \sqrt{\int_{-\infty}^{\infty} v^2 \mathcal{L}_{\rm norm}(v) dv} \tag{3.19}$$

We call this $V_{\rm rms}$ because $V_{\rm rms} = \sqrt{V^2 + \sigma^2}$ with V and σ the mean and standard deviation of the LOSVD. In the pure Gaussian case these are just the parameters V and σ of the best-fit Gaussian. However, for a non-Gaussian LOSVD the mean and standard deviation must be calculated.

One way of calculating the integral in Equation 3.19 for a non-Gaussian LOSVD is numerical integration. However, it is possible to analytically perform the integral. To see this, we consider the more general problem of calculating the n-th moment of the distribution.

$$\langle v^n \rangle = \int_{-\infty}^{\infty} (\sigma y + V)^n \frac{e^{-y^2/2}}{C\sqrt{2\pi}} \left[1 + \sum_{m=3}^{M} h_m H_m(y) \right] dy \tag{3.20}$$

We show the full calculation for this in section A.2. The result is

$$\langle v^n \rangle = \frac{1}{C} \sum_{\ell=0}^{n} \sum_{m=0}^{M} \sum_{k=0}^{\ell} h_m \sigma^\ell V^{n-\ell} A_{k\ell mn} \tag{3.21}$$

with

$$A_{k\ell mn} = \frac{n! \sqrt{2^k m!}}{k!(n-\ell)!(m-k)!!(\ell-k)!!} \tag{3.22}$$

where $m - \ell, \ell - k \geq 0$ and $m - k$,$\ell - k$ are even, else $A_{k\ell mn}$ returns zero.

It is worth noting that the moments disproportionately weigh contributions to the LOSVD from the wings of the distribution. This is important when considering real observations as it is the wings of the LOSVD that are the most under-constrained by the data. To better clarify the dependence of the second moment on the wings, we plot the second (and fourth) moments as a function of the integration bounds for a pure Gaussian LOSVD and an LOSVD with $h_4 = 0.1$ in Figure 3.2. We find that, while it is true that measurements of the fourth moment are more sensitive to the wings than the second moment (they need to be integrated out further for an accurate recovery), the main factor driving the sensitivity of the wings is h_4. We see that increasing h_4 from 0 to 0.1 increases the range in which we must know the wings by around 1 σ. Going between the second and fourth moment, however, only increases this by a fraction of a σ. The reason for this is due to the fact that the

quantity we are looking at is not $\langle v^n \rangle$ but rather $\langle v^n \rangle^{1/n}$. If we produce the same plot for $\langle v^n \rangle$ (the lower panel of Figure 3.2) we recover the result that agrees with the standard intuition, and find that the requirements on the bounds are the same between a Gaussian LOSVD fitting the fourth moment or an LOSVD with $h_4 = 0.1$ fitting the second moment.

3.5 Measuring the Underlying Kinematics

3.5.1 Ideal Galaxy LOSVD

The output from the GADGET3-KETJU code is a file containing the position and velocity of every particle in the simulation. We need a way to take this information and determine the true underlying shape of the LOSVD across the field of view of a mock observation. To do this, we construct a "velocity cube" where the first two dimensions correspond to the spaxel location and the third dimension is a list of all of the velocities along the line-of-sight within that spaxel. We do this along an aperture that goes out to 14.3 kpc (the half-mass radius, $R_{1/2}$) and use a pixel size of 0.2 kpc (one fifth of the sphere of influence). We then bin the velocity cube to a target particle number of 5000 using the Voronoi tessellation algorithm of Cappellari & Copin (2003). For each binned spaxel we fit a GH series to a histogram of the velocities. We do this by normalising the velocity histogram and integrating a model LOSVD over the same velocity bins and then finding the best-fit parameters with the CAPFIT[2] routine of Cappellari (2023). We assume constant errors on each histogram. We show an example fit to a LOSVD with known parameters in Figure 3.3. We limit the bins to span a range of $\pm 7\sigma$ with σ the biweight standard deviation of the stellar LOSVDs within each Voronoi bin. The use of the biweight standard deviation over the normal standard deviation is designed to protect against any outliers in the LOSVD. It is important to include such a large velocity range in the histogram since this effectively enforces that the LOSVD goes to zero at large velocity. We found that choosing a smaller velocity range often led to highly over-fitted solutions with large amounts of non-Gaussianity outside of the velocity range of the histogram. We show the GH moments measured from our simulation along the y and z axes in Figure 3.6. We see that there is a lot of correlated structure in the higher-order moments beyond h_4.

We should note that the fitting procedure we describe is only necessary for measuring the higher-order GH moments. In principle, we can calculate the velocity first

[2]Available at `https://pypi.org/project/capfit/`.

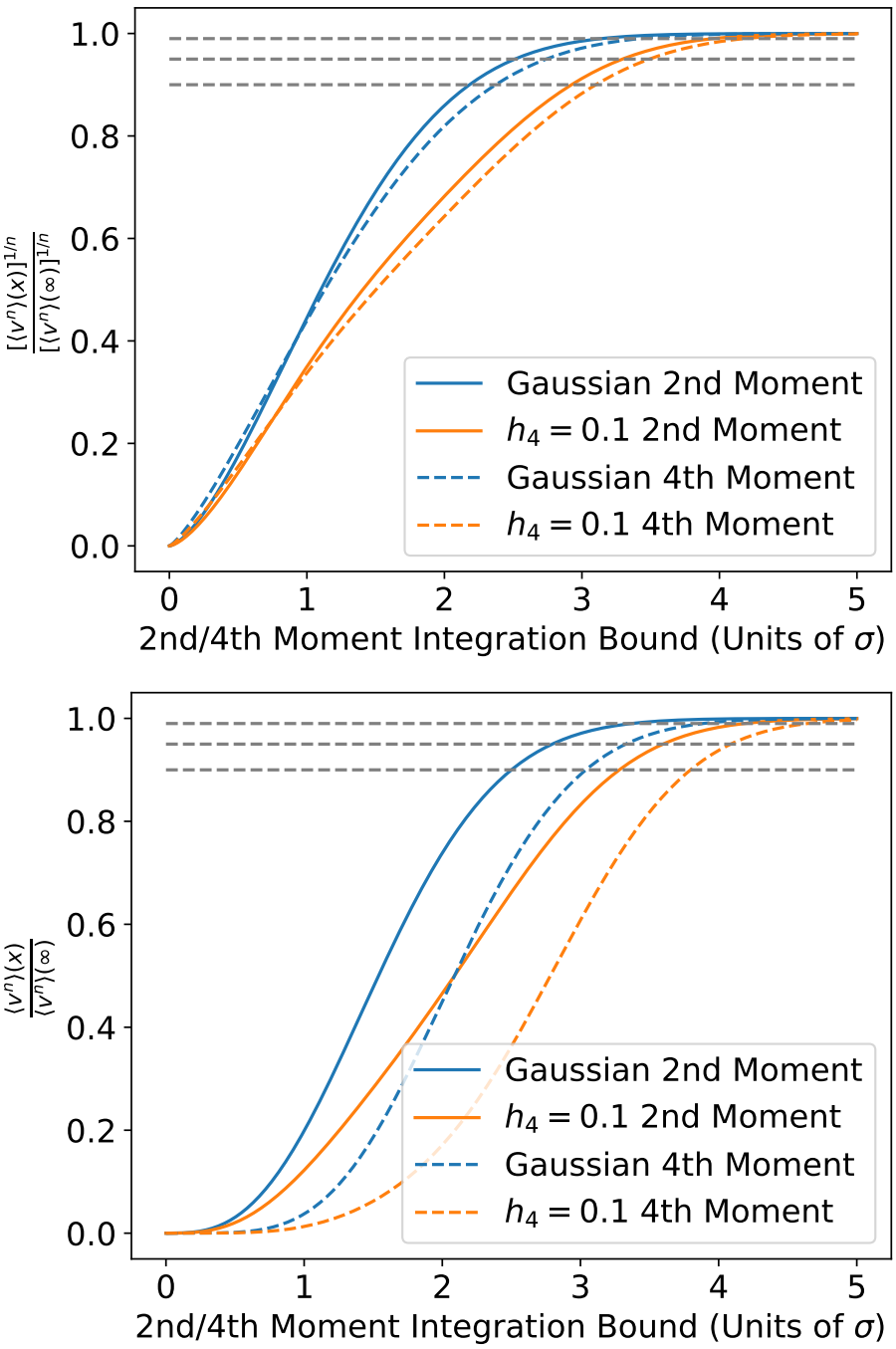

Figure 3.2: The top and bottom panels show $\langle v^n \rangle^{1/n}$ and $\langle v^n \rangle$ as a function of the integration bounds. From these we see that in order to accurately determine the second moment, we need to have a good knowledge of the wings of the LOSVD out to 3-4σ. Whether it is 3 or 4 σ depends on the underlying shape of the LOSVD (h_4 pushes this out) and the degree of the moments being fit (higher-order moments require more knowledge).

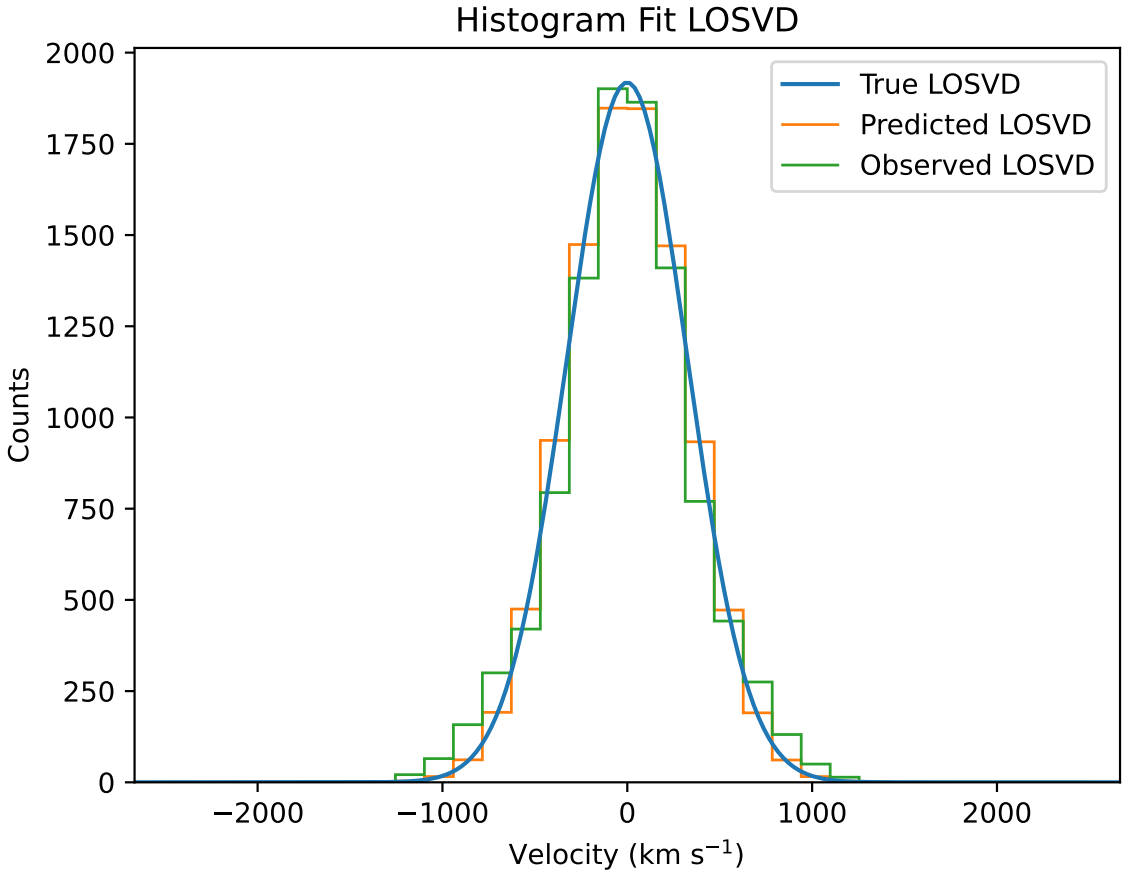

Figure 3.3: An example of a fit of a GH series to the histogram of a LOSVD. In this case the true underlying LOSVD has $v = 0, \sigma = 300, h_4 = 0.1$, and $h_3 = h_5 = h_6 = h_7 = h_8 = 0$ with five thousand particles sampled from it while we fit with just v and σ. The blue curve is the true underlying LOSVD exactly parameterised using the GH series, the green curve is the binned histogram version of the blue curve, calculated by integrating the blue curve over each bin, and the orange curve is the best fit match determined by minimising a χ^2 defined by the difference between the green and orange histograms.

moment and second moment directly as a summation of the particles in each Voronoi bin. Namely, for each Voronoi bin we have

$$\overline{V} = \frac{1}{N}\sum_{i=1}^{N} v_i \tag{3.23}$$

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N} (v_i - \overline{V})^2 \tag{3.24}$$

$$V_{\rm rms} = \sqrt{\overline{V}^2 + \sigma^2} \tag{3.25}$$

with N the number of particles in a given spaxel. For all of our tests, unless otherwise stated, we used the moments calculated directly from the particles. It is important to note that $\overline{V}$ as it appears above is not the same V as in the GH series. $\overline{V}$ is the mean of the line-of-sight velocity distribution, whereas V is an arbitrary number in the GH expansion that is only equal to the mean in general if all odd h_i are zero. Likewise for σ.

3.5.2 Limits to the LOSVD Resolution

One important detail to note is that, while the GH series is orthogonal, it is orthogonal with respect to a fixed choice of V and σ. In practice, we treat V and σ as free parameters. This means that the measured values of the higher-order GH moments depend on the total number of moments that we fit (a fit out to h_4 will have different values of h_3 and h_4 than a fit out to h_6 since V and σ can be different). This leads to complex relationships between the even and odd parameters of the GH series. To clarify this, we run an MCMC for a mock observed LOSVD consisting of 5000 particles and parameters $\sigma = 100$ and $h_4 = 0.1$, and the remaining parameters equal to zero assuming Poissonian errors. We generate this mock data using inverse transform sampling[3]. We show the results of this in Figure 3.5. Here, one clearly sees the degeneracy between different parameters in the GH series. In particular, we see that, as we expect, there is almost no correlation between the even and odd moments of the LOSVD. However, there is some correlation and anti-correlation between the even moments and between the odd moments. This mixture of positive and negative correlations is due to the fact that there are now four different parameters controlling the even/odd features of the GH LOSVD and these can be held roughly equal by increasing several parameters together while simultaneously decreasing others.

[3]It is important to note that this method will not work for any choice of parameters such that the LOSVD becomes negative. For this reason we choose as an example LOSVD one which is never negative.

A close look at Figure 3.5 reveals that not all of the posteriors perfectly recover the original input values. This is due to the fact that we sample a finite number of particles for our LOSVD. Knowing the intrinsic uncertainty in each parameter due to particle number is important for our analysis as it sets the lower limit to the uncertainty of the extracted LOSVD. This is especially important as we would ideally like to consider scenarios where our mock observations probe deep into the black hole sphere of influence. However, there is a trade-off between the field of view and particle number. To investigate these questions, we consider an idealised scenario where we assume a LOSVD, randomly sample N particles from it using the inverse transform method, fit a GH LOSVD to it, and measure the deviation from the underlying parameters in the GH series. In order to understand the statistics of this, we perform this 100 times. We show the results of this in Figure 3.4. We see that, for our choice of LOSVD, including 5000 particles leads to an uncertainty in the second moment ($V_{\rm rms}$) of approximately 1.5 per-cent. This is what we expect from the standard error, given by $1/\sqrt{n}$. For the Hermite moments, we see that the uncertainty for 5000 particles is about 0.01, approximately the same scale as the offsets we see in Figure 3.5.

It is important to note that we are making an assumption that the observed LOSVD in our simulated galaxy is a random sampling of the true underlying distribution which we might find by considering progressively higher-resolution simulations. This is not necessarily the case. This question was previously investigated in Harborne et al. (2024). In that work, the authors study how the LOSVD of a mock IFS observation varies in a simulation at different stellar mass resolutions. They find that errors in the velocity, h_3, and h_4 (as compared to higher-resolution versions of the same simulation) tend to be randomly distributed around zero, showing that lower-resolution simulations do tend to have LOSVDs centered around the same value as higher-resolution simulation. However, the dispersion σ tends to systematically be smaller in lower-resolution simulations. They quantify this effect and find that the systematic offset in the dispersion asymptotically flattens off with progressively higher-resolution, becoming close to flat at 500 particles per bin. As we use an order of magnitude more particles, we expect our results not to be sensitive to this systematic offset in the dispersion. This implies that the LOSVDs we recover share essentially the same underlying parameters as a simulation with realistic stellar particle masses. We thus don't anticipate there being any significant impact from the dispersion intrinsic to individual stellar particles.

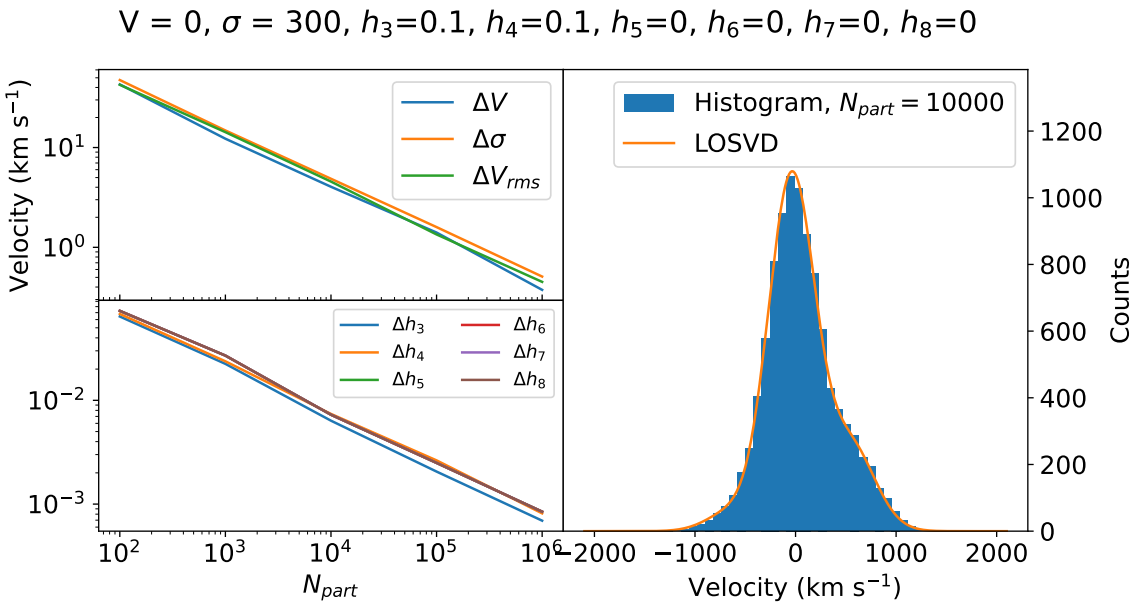

Figure 3.4: The right panel shows a histogram made from 10,000 particles randomly sampled from the LOSVD with the parameters $V = 0$, $\sigma = 300$, $h_3 = 0.1$, and $h_4 = 0.1$. The orange curve shows a smooth plot of the LOSVD as a function of velocity. The upper and lower left hand panels show the RMS per-cent error of the recovered parameters of the LOSVD determined by fitting a GH series to a histogram with between 100 and 1,000,000 particles. For each choice of the number of particles we consider 100 different fits to derive our RMS per-cent error.

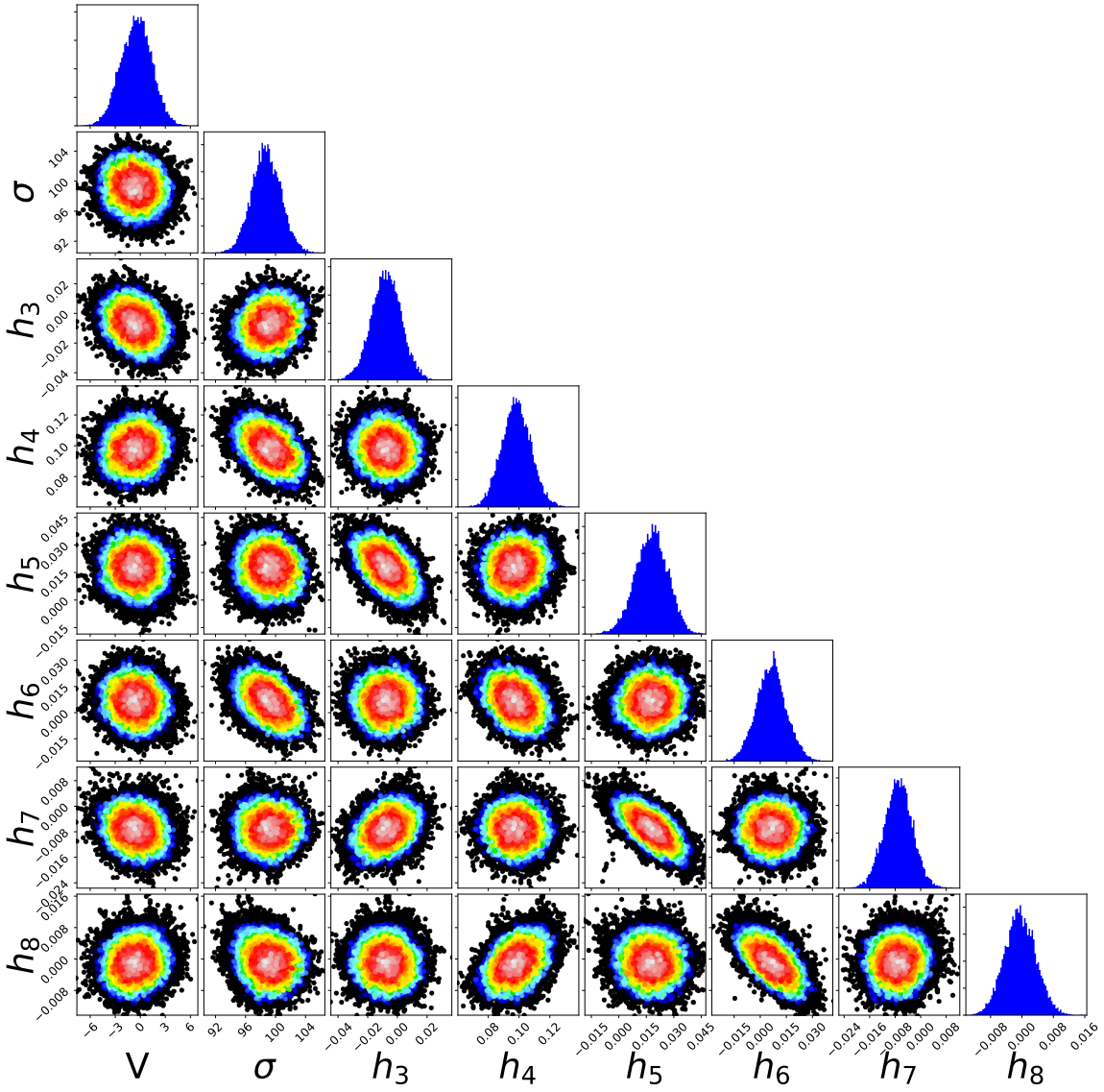

Figure 3.5: An example of a fit of a GH series to the histogram of a LOSVD. In this case the true underlying LOSVD has $v = 0, \sigma = 100,\ h_3 = 0,\ h_4 = 0.1,\ h_5 = 0,\ h_6 = 0$ with a thousand particles sampled from it while we fit with just v and σ.

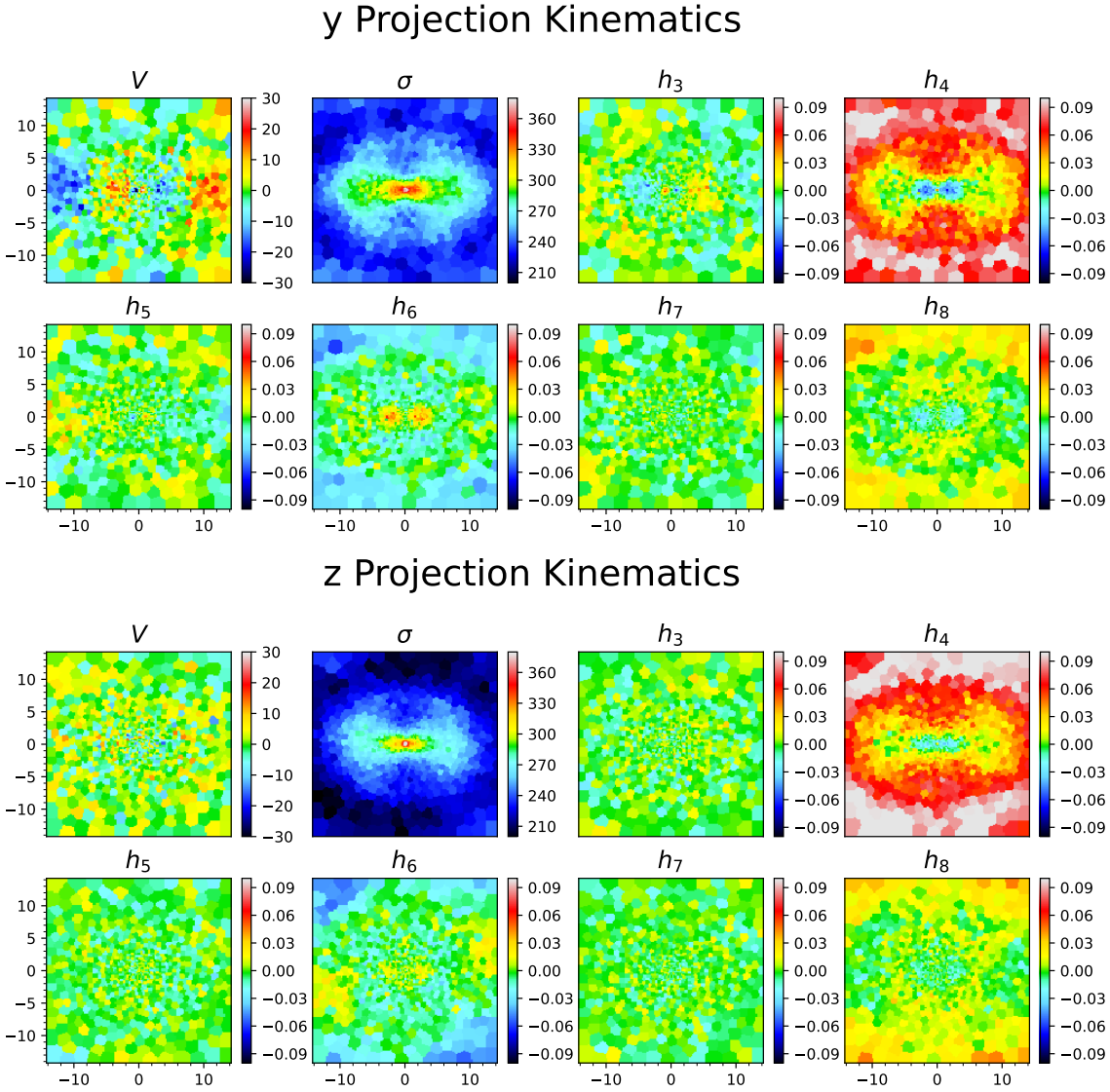

Figure 3.6: The top and bottom panels show y and z projection kinematics when the Voronoi binned data are fitted with a GH series out to moments = 8. In both cases but especially the y projection we see clear structure in the higher-order moments.

3.5.3 2nd Velocity Moment Recovery

For Schwarzschild modelling, we care about recreating the overall observed shape of the LOSVD from a superposition of orbits. This usually means calculating the GH moments from individual orbits and fitting them to the data. For Jeans modelling, we care about the recovery of the second velocity moment, a quantity that is derived from information coming from all GH moments in the LOSVD. In Figure 3.7 we calculate the second moment after fitting the Voronoi binned mock data with histograms including just a pure Gaussian, moments up to 4, 6, and so on, which we then compare to the true second velocity moment given by counting individual particles. We do this twice, once for the full field of view and once for a cropped field of view out to 3 kpc. Rather than each figure monotonically converging to the true value, the shapes of the maps oscillate, with moments = 6 and 10 appearing to do better at fitting the inner shape of the map while moments 4 appears better than moments 6 at fitting the outer regions of the map.

This is primarily due to the fact that, while including higher-order moments in the GH series ensures a better χ^2 for the LOSVD, it does not guarantee that the second velocity moment universally improves with higher-order moments. To see this more clearly, we plot the average bias and deviation of the second moment for the x, y and z projections in the simulation in Figure 3.8. There we see that there is a clear jump in the second moment going from the Gaussian case to moments = 4. After this, the bias slowly approaches zero, but not monotonically.

To better understand what is going on with the LOSVD that is causing this behavior, we plot the average residuals of each LOSVD fit for the moments = 2 and 4 case of the y projection in Figure 3.9. We see that in the moments = 2 case that the single Gaussian attempts to account for the wings of the distribution by shifting some flux away from the central peak and towards to outskirts of the distribution, but ultimately this falls short, leading to a 5 per-cent systematic offset in the second moment. The moments = 4 case significantly improves on this, though there is still some systematic offset.

We see from Figure 3.8 that the recovery of $V_{\rm rms}$ is sensitive to the number of moments included in the fit. This itself, however, depends on the underlying LOSVD. Maps that tend to have lower values of h_4 will have a $V_{\rm rms}$ that is less sensitive to h_4. Fu et al. (2025) fit h_3 and h_4 to a large sample of MaNGA galaxies. They found that, across all galaxies and pixels, the average value of h_4 is around 0.03, with a high end that peaks around 0.15 (though the majority of points are less than 0.1). To estimate the average $V_{\rm rms}$ offset between GH moments = 2,4 for real galaxies we fit a mock

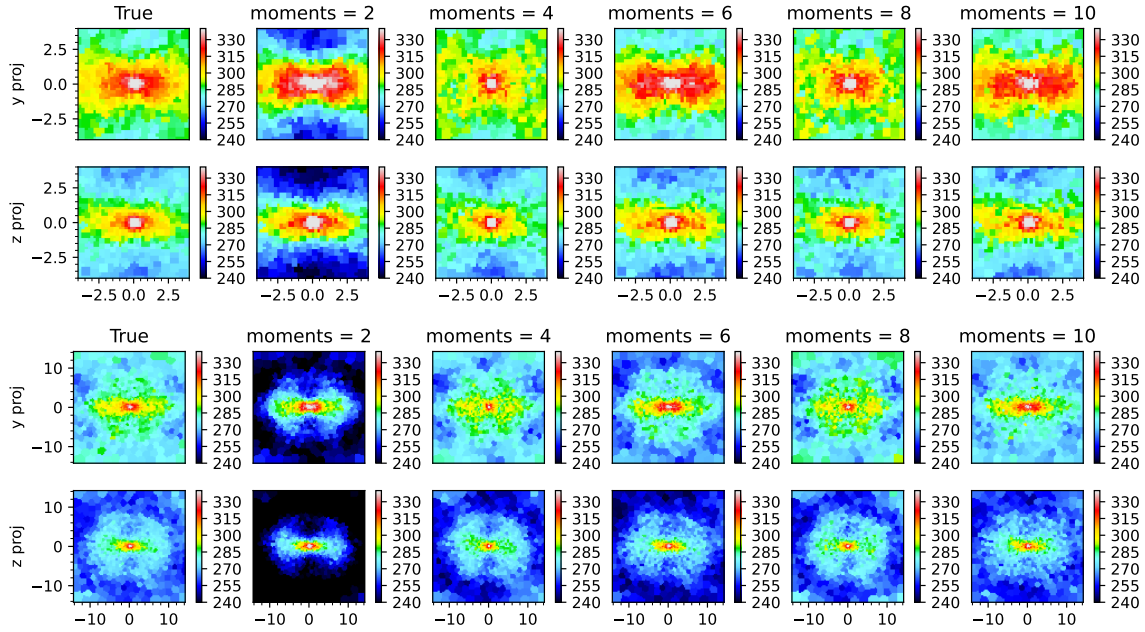

Figure 3.7: The top and bottom panels show, respectively, a 3 and 14.3 kpc field of view of the 2nd moment. The left-most column shows the true second moment that comes from Equation 3.25. One key feature is that the second moment doesn't monotonically approach the truth as the number of moments is increased, but rather oscillates depending on the number of moments.

LOSVD histogram with $V = 0, \sigma = 100, h_3 = 0, h_4 = 0.03$ with moments = 2 and 4 and calculate the second moment. We find that the two second moments have a 5 per-cent offset. At the large end where $h_4 = 0.1$ this increases to a 13 per-cent offset. As $V^2_{\rm rms} \propto M$ with M the mass of the system (Equation 5.4), this corresponds to a shift of the mass by ten per-cent at the low end and 27 per-cent at the high end. In practice, however, there are very few galaxies that have an $h_4 = 0.1$ across their full field of view. For this reason, if the kinematics one extracted from the spectra followed the same behaviour as the idealised fits to the velocity histograms, we estimate, purely from the fact that JAM models are almost exclusively done without higher-order moments, that there is a typical minimum modelling systematic uncertainty of 10 per-cent for JAM. If one were able to accurately measure higher-order moments, as we see from Figure 3.8, this would reduce to a 2 per-cent systematic uncertainty.

Clearly it is better to include higher-order GH moments when calculating $V_{\rm rms}$ in this idealised scenario. However, as we will investigate in chapter 4, this is not necessarily true when including other realistic observing effects such as template mismatch.

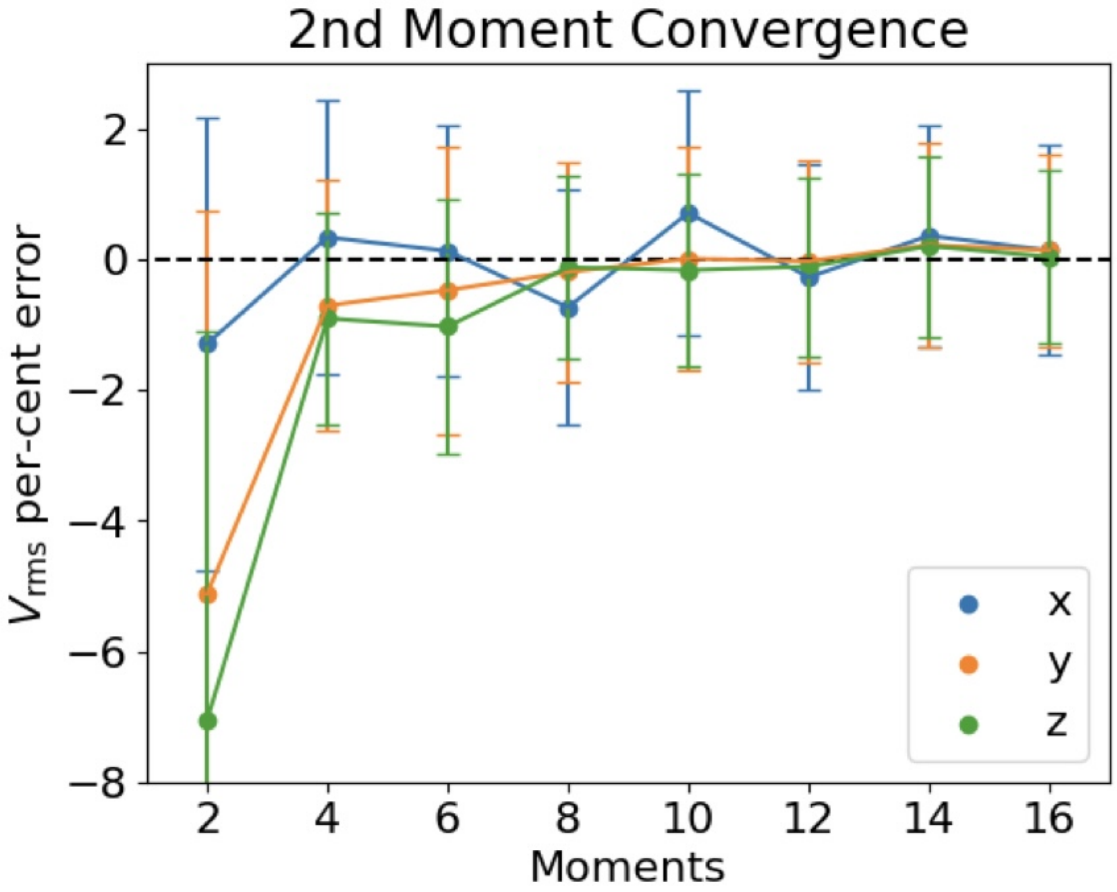

Figure 3.8: This figure shows, for each projection, the average bias and dispersion in the recovery of the second moment. We define the bias here as the mean of $(V_{\mathrm{rms,histogram}} - V_{\mathrm{rms,true}})/V_{\mathrm{rms,true}}$ across all spaxels in the projection, and the dispersion the standard deviation of this. The quantity $V_{\mathrm{rms,histogram}}$ is calculated using the analytic form in Equation 3.21. There is a clear jump going from moments = 2 to 4, followed by a somewhat dampened oscillation around 0.

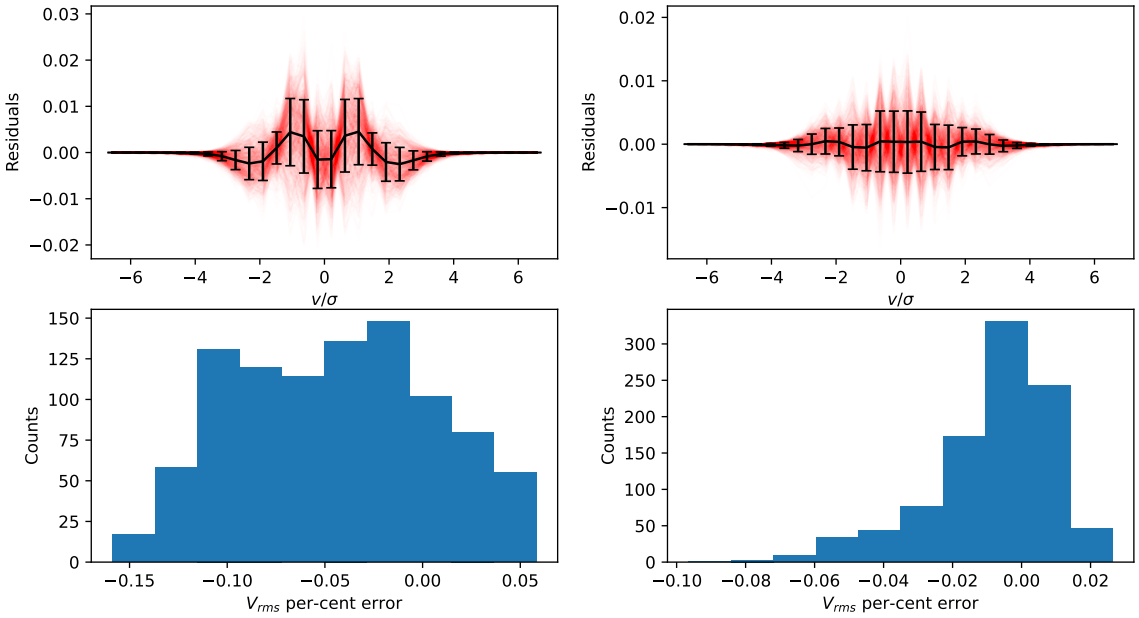

Figure 3.9: The top panel shows all of the residuals of the fits to each Voronoi bin in red and the average in black. The bottom panel shows the corresponding histogram of the offset in the second moment. The left panels are for moments = 2 and the right panels for moments = 4.

3.5.4 Challenges of the GH Series

The expression in Equation 3.18 is a probability distribution for the line-of-sight velocity. This interpretation, however, is challenged by the fact that the generality of the Gauss-Hermite series allows it to be negative. This, of course, is not physical, and no real galaxy LOSVD has a negative component. However, even realistic cases such as flat-topped LOSVDs require negative h_4 and thus always contain a negative component of the LOSVD at large radii. Additionally, for real observations, subject to noise and template mismatch (and the fact that we often use only a finite number of terms in the GH series), it is sometimes the case that the LOSVD measured in real galaxies does contain some negative component.

In our simulated galaxy we find that fitting the velocity cube with moments = 4 leads to approximately 20 per-cent of the Voronoi bins having a fit LOSVD with some negative component. Fitting with moments = 8 leads to nearly 50 per-cent of the bins having some negative component. Typically, the LOSVD is only negative by a very small amount, but because this occurs near the wings of the LOSVD, it is possible for it to have an impact on the second moment.

To investigate this, we compare the second moments that we measure from the

binned velocity cube where we directly apply the expression in Equation 3.21 (allowing for negative LOSVDs) to a modified version of Equation 3.19 where we numerically integrate the LOSVD and set all values of the LOSVD < 0 to 0. We show the results for this in Figure 3.10. We see that the deviation is almost always zero for values of $V_{\rm rms}$ less than 280 km/s. This corresponds to points that are outside of the nuclear region (see Figure 3.7). Within the nuclear region, there is much more variation in the shape of the LOSVD in the direction of negative wings (Figure 3.6), with both projections shown having their minimum in h_4 occur in the nuclear region. The reason why this scatter does not extend all the way up to the highest velocities partially has to do with the fact that there is a much larger volume of points on the one-to-one relation that there are outliers, and partially has to do with the fact that the exact spaxels closest to the black hole sometimes have sharp increases in h_4 and thus will not be outliers on the one-to-one relation.

Additionally, it is worth noting that, while a negative probability density is aesthetically unpleasing, forcing the LOSVD to be non-negative would systematically bias the second moments in the positive direction. For this reason, even some of the so-called "non-parametric" codes do not enforce positivity (see Mehrgan et al. (2023) for an example). For this reason we choose to proceed by calculating all of our second moments for the remainder of this thesis by applying the standard definition in Equation 3.19.

3.6 Conclusion

The GH series is a natural and convenient way to parameterise the LOSVD of galaxies. The use of higher-order moments around a central Gaussian allows, using only a few parameters, to clearly separate the wings and central peak of the distribution. However, as we have seen, the GH series has somewhat unintuitive convergence properties when it comes to measuring the second velocity moment $V_{\rm rms}$ (e.g. Figure 3.7). As we show in Figure 3.8, this results in there being a bias in the second moment that mostly disappears after including h_4, but which oscillates before converging. In this idealised scenario, we expect this to set a theoretical upper bound on the accuracy of dynamical models that do not take higher-order moments into account, with the typical uncertainty being around 10 per-cent for parameters contributing to the gravitational potential (and thus a 10 per-cent uncertainty in the black hole mass).

In the next chapter we turn our attention to the more realistic scenario where kinematics are not determined from histograms but from spectra.

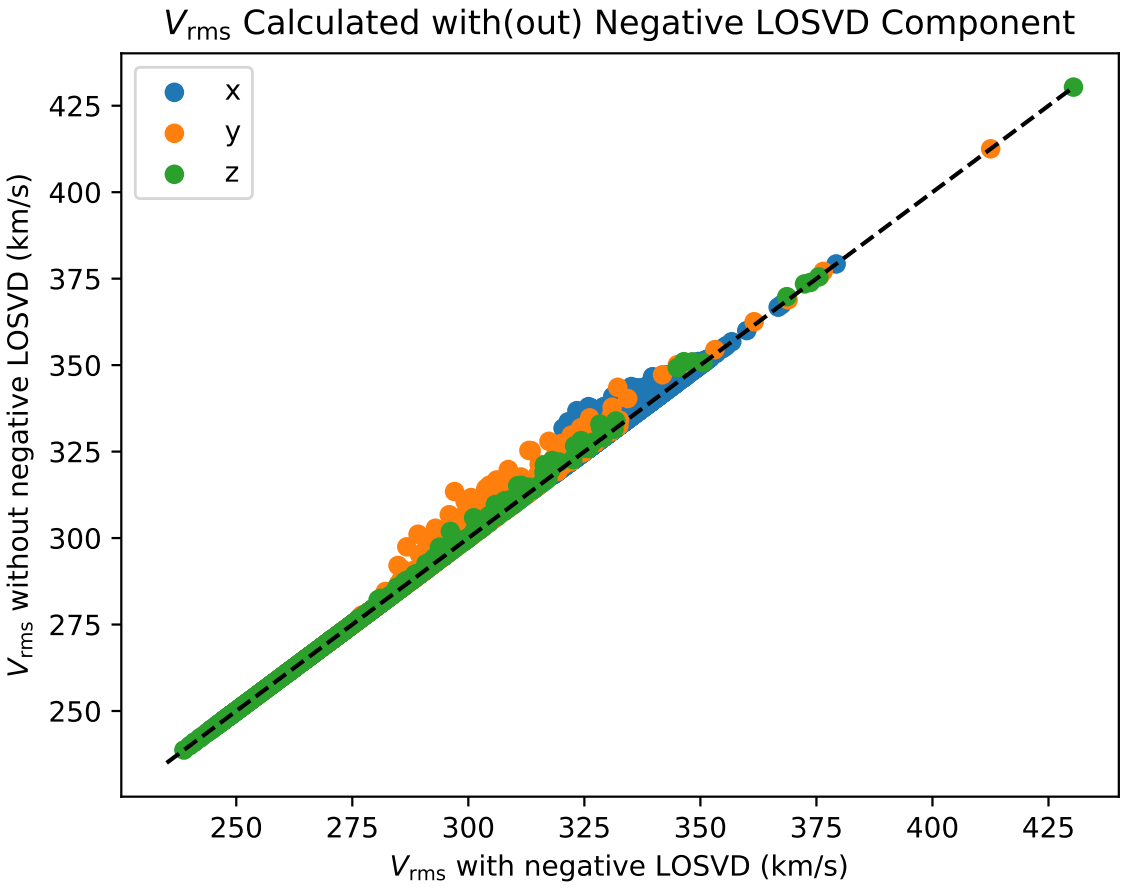

Figure 3.10: The top panel shows the residuals of fits of the LOSVD using only moments = 2 and moment = 4, respectively. The bottom panel shows a histogram of the per-cent error in the second moment for each of the Voronoi bins in the mock observation. We see that using moments = 2 leads to a systematic 5-10 per-cent under-estimation of the second moment, while using moments = 4 gives no bias but an approximately two per-cent scatter.

Chapter 4

Recovery of Kinematics Subject to Systematic Uncertainties

4.1 Summary

Template mismatch can cause systematic errors in the extracted line-of-sight velocity distribution. This is especially important for studying early-type galaxies which are more alpha-enhanced than the local Milky Way stars that are observed for empirical stellar libraries. We quantify the error in derived kinematics due to template mismatch in early-type galaxies by creating a mock datacube from a high-resolution N-body simulation. We then fit this datacube using realistic template mismatch and study the best choice of pPXF parameters to minimize the error in the second moment.

4.2 Introduction

Many techniques have been developed to extract information about the line-of-sight velocity distribution (LOSVD) from galaxy spectra. The first redshifts were measured using microscopes to compare reference spectra to galaxy spectra on photographic plates (e.g. Slipher (1913, 1914, 1917)). At this time, it was difficult to measure galaxy dispersions compared to redshifts, with some being determined purely by visually comparing with convolved stellar spectra (e.g. Faber & Jackson (1976)).

The first quantitative approaches used the Fourier quotient technique (Simkin, 1974; Sargent et al., 1977), and later the correlation method (Tonry & Davis, 1979; Franx & Illingworth, 1988; Bottema, 1988). With the increased speed of computation, a shift away from Fourier space towards pixel space was made (Rix & White, 1992; Kuijken & Merrifield, 1993; van der Marel, 1994; Saha & Williams, 1994; Merritt,

1997; Gebhardt et al., 2000a; Kelson et al., 2000). This method works by taking stellar spectra, convolving them with a line-of-sight velocity distribution, usually parameterised using a set of orthogonal functions such as a Gauss-Hermite (GH) series (van der Marel & Franx, 1993; Gerhard, 1993), with histograms (Rix & White, 1992; Saha & Williams, 1994; Merritt, 1997) or B-Splines (Falcón-Barroso & Martig, 2021), and fitting it to the observed galaxy spectra. This offered many important conveniences, such as being able to readily mask gas lines, study stellar populations, and include continuum contributions to the modeled spectra. There are now a number of different codes available that use pixel fitting to fit integral field data (Cappellari & Emsellem, 2004; Cappellari, 2017, 2023; Cid Fernandes et al., 2005; Wilkinson et al., 2017; Sánchez et al., 2016; Lacerda et al., 2022).

Among the methods fitting in pixel space, pPXF has been shown to be the fastest and most accurate for determining stellar populations (Woo et al., 2024). In the case where the stellar templates span the galaxy spectra and the measured features are resolvable within the spectral resolution, we expect the pPXF method to perfectly recover the underlying kinematics. However, it is almost always the case that there is some residual template mismatch between the galaxy spectra and the stellar template libraries. This problem is exacerbated for early-type galaxies. Stellar libraries are constructed by observing individual stars in our galaxy, limiting the population of stars used in empirical stellar libraries to the stellar population of the Milky Way. Early-type galaxies, however, tend to have stellar populations that are older, redder, and more alpha-enhanced than local stars.

This raises several key questions. How accurate is the pPXF method when fitting spectra with template mismatch? What are the best practices for alleviating template mismatch when this is the case? Is there an optimal number of GH moments one should fit? What about additive and multiplicative Legendre polynomials? In this work, we seek to answer these questions by creating a fully realistic mock observation of a simulated galaxy and studying how the recovery of the kinematics depends on these choices.

This chapter is structured as follows: in section 4.3 we introduce the pPXF method. In section 4.4 we discuss the standing of the current empirical data and some of the motivations for this study. In section 4.5 we describe how we create realistic mock integral field datacubes from our N-body simulation. In section 4.6 we describe our choice of stellar libraries and in section 4.7 we present our fits and tests of different parameters. Finally, in section 4.8 we list a number of best practices that aim to guarantee the most reliable extraction of the LOSVD from realistic data.

4.3 The pPXF Method

pPXF (Cappellari & Emsellem, 2004; Cappellari, 2017, 2023) is a full-spectrum fitting code designed primarily to measure the stellar kinematics of external galaxies. It was developed during the SAURON project to extract kinematics from SAURON galaxies. It has since been used in a wide number of applications including the ATLAS^{3D} project, as a part of the MaNGA data reduction pipeline (Westfall et al., 2019), and others.

The method works by fitting stellar templates, typically empirically observed stars, to the observed galaxy spectrum. The pPXF fitting function is given by (Cappellari, 2023, eq. 11)

$$G_{\rm mod}(x) = \sum_{n=1}^{N} w_n \left[(T_n(x) * \mathcal{L}_n(cx)) \sum_{k=1}^{K} a_k \mathcal{P}_k(x) \right] + \sum_{l=0}^{L} b_l \mathcal{P}_l(x) \qquad (4.1)$$

where T is the template spectrum, $\mathcal{L}(cx)$ is the corresponding unnormalised LOSVD for each template given by Equation 3.14, $\mathcal{P}_k(x)$ are multiplicative Legendre polynomials and $\mathcal{P}_l(x)$ are additive Legendre polynomials with degree K and L, respectively. Multiplicative polynomials and additive polynomials have been used to improve the fit of stellar templates to observed galaxy spectra by reducing template mismatch and accounting for non-stellar contributions to the continuum. Our goal is to optimise this set of parameters with the goal of accurately recovering the kinematics.

4.4 Empirical Data

The question of how to optimally measure the velocity second moment from a GH LOSVD goes back to the SAURON (de Zeeuw et al., 2002) and ATLAS^{3D} (Cappellari et al., 2011a) projects. There are two possible approaches. The first approach is to fit the kinematics with pPXF using a pure Gaussian LOSVD. This has the advantage that the velocity and dispersion are less sensitive to template mismatch than the higher-order moments, and one expects that the dispersion will adjust to account for large wings of the distribution in a way which improves the recovery of $V_{\rm rms}$. Alternatively, in the presence of ideal data, we expect using the full GH series to provide the most accurate result given its generality.

In the absence of realistic simulated data where the truth is known, the best approach is to consider both approaches and see which ones return the most reasonable results. In Figure 4.1 we show kinematic fits for twelve different ATLAS^{3D} galaxies[1]

[1]Available at `https://purl.org/atlas3d`.

with maps of the second moment determined using just a pure Gaussian LOSVD and using the Gauss-Hermite series up to h_4 along with mass-follows-light JAM fits to each of these (more on JAM in subsection 5.3.2). We see that the maps have systematic differences between them depending on the shape of the h_4 map. It is clear that the second velocity moment including higher-order GH moments is noisier than the pure Gaussian case, often to the extent of wiping out some of the structure evident in the second moment map using a pure Gaussian LOSVD. The lack of clear structure in many of these maps results in the JAM fits visually appearing quite poor as the model does not know which features to reproduce. On the other hand, JAM models to the data using only a Gaussian LOSVD provide visually much better fits to the data, often reproducing the overall shape and amplitude of the $V_{\rm rms}$ maps. Given that JAM is a predictive model (it produces a unique second moment map, unlike Schwarzschild models which in principle can produce fits to a wide variety of unrealistic data), this suggests (though it is not a proof) that the second moment without higher-order moments is more reliable than the one with higher-order moments.

Given these results showing the noisiness of the second moment including higher-order GH moments and the poor fit of JAM models to this data, the most common choice in the literature so far has been to take the first approach and assume a pure Gaussian LOSVD for JAM modelling. This choice, however, has never been tested in an idealised setting with mock observations where the ground truth is known.

4.5 Mock IFS Observations

4.5.1 Stellar Population Assignment

Each star must be assigned a spectrum for the mock datacube. As our goal is to create a realistic datacube we must carefully choose our stellar population assignment in order to best reproduce the limitations met when working with real data. Key constraints we would like to reproduce are that the stellar populations observed have some alpha enhancement while the spectra used for the fitting are purely empirical. With this in mind, we choose to use the semi-empirical sMILES SSPs of Knowles et al. (2023) as they span a wide range parameters at high spectral resolution.

For the distribution of stellar population parameters, we consult Figure 12 of Kuntschner et al. (2010), which plots age, [Z/H], and [α/Fe], as a function of radius for the SAURON population of early-type galaxies. The key result is that both age and [α/Fe] are relatively flat with radius while [Z/H] varies linearly with the log of the

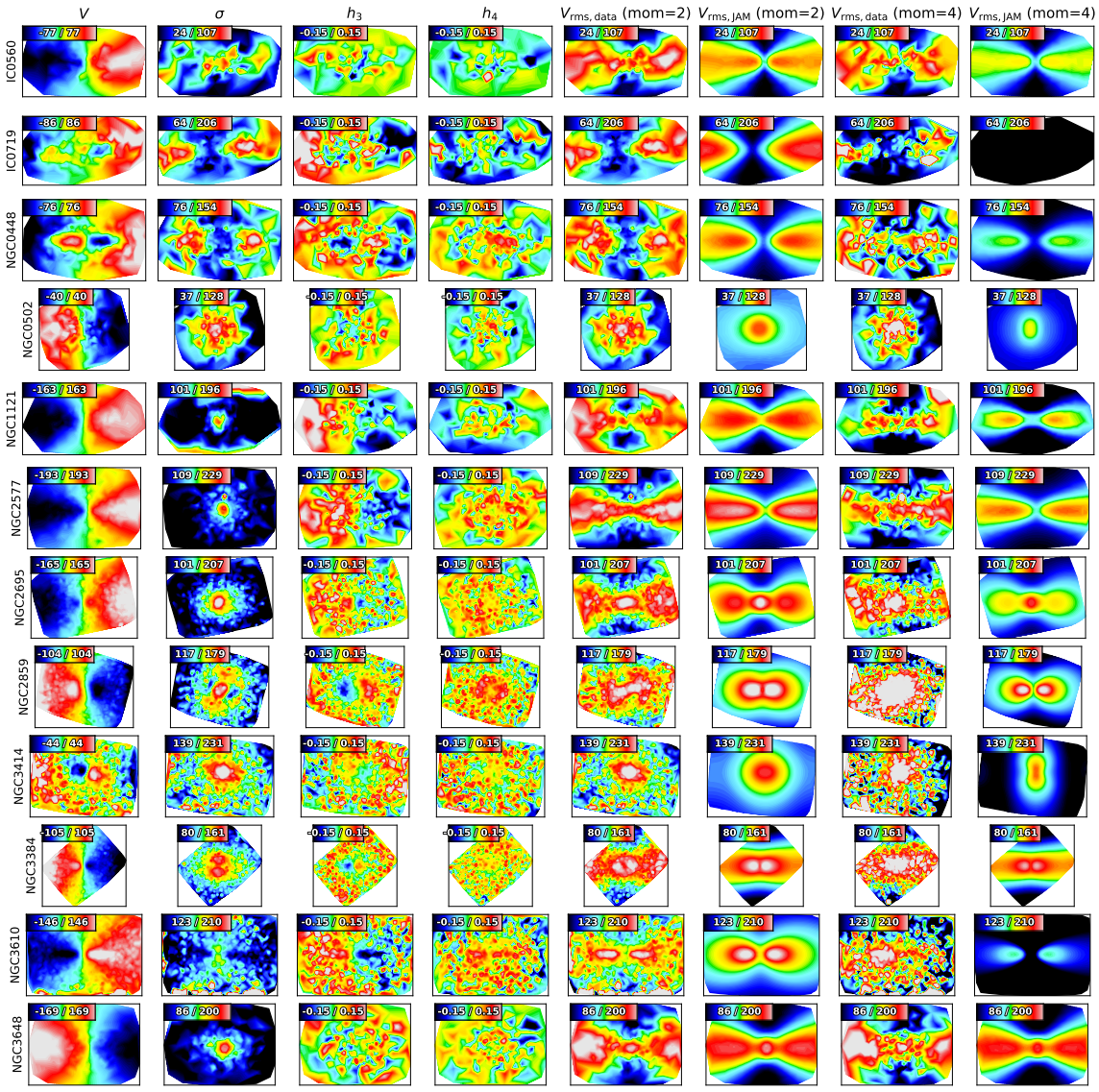

Figure 4.1: Atlas3D galaxies with their measured kinematics and mass-follows-light constant anisotropy $\mathrm{JAM_{cyl}}$ fits. From left to right we plot the velocity, σ, h_3, h_4 (all determined from a pPXF fit with moments = 4), V_{rms} (Assuming a pPXF fit with moments = 2), a $\mathrm{JAM_{cyl}}$ fit to that V_{rms}, and the same for V_{rms} calculated with moments = 4. The velocity is plotted with symmetric bounds, h_4, h_4 are plotted from -0.15 to 0.15, and σ and all second moments are plotted on the same scale for the same galaxy. We see that the two methods of measuring the second moment produce qualitatively different maps, with a pure Gaussian LOSVD being noticeably less noisy than those including higher-order GH moments. We also observe that the JAM models tend to be better at describing the pure Gaussian second moments compared to those with higher-order GH moments.

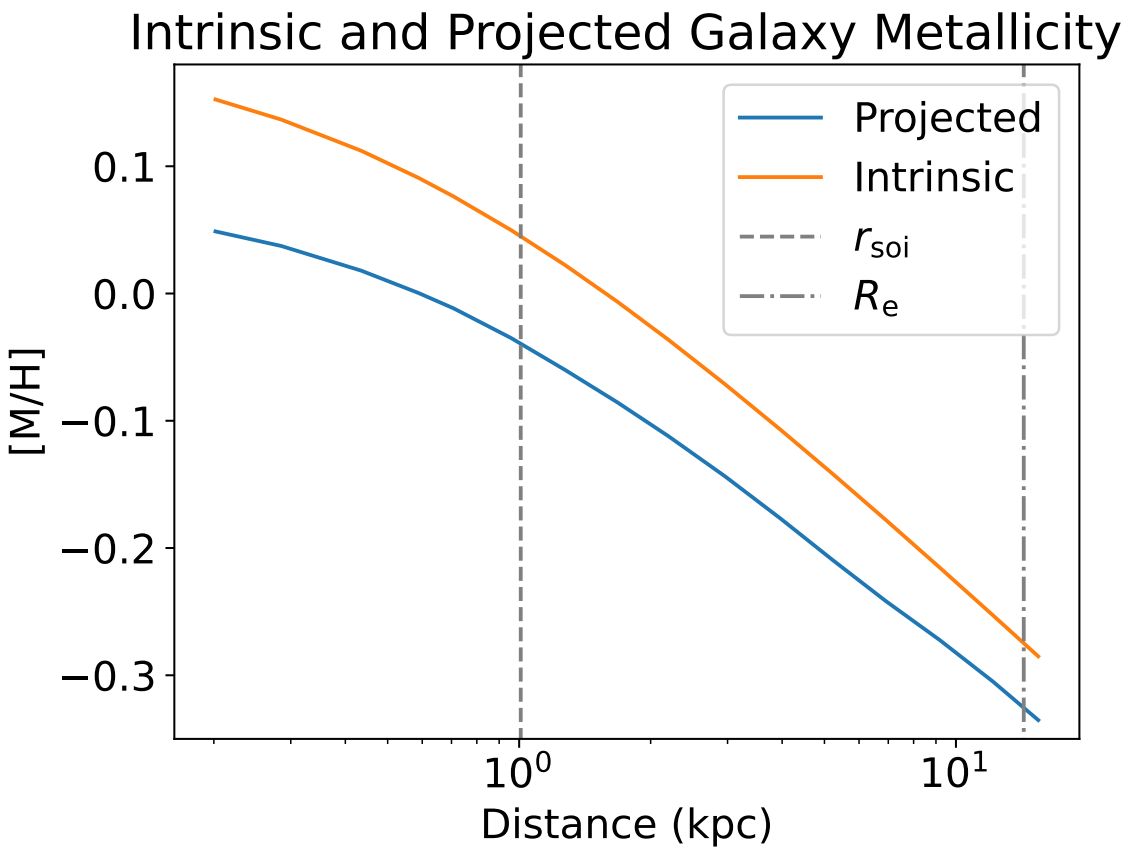

Figure 4.2: Intrinsic and projected metallicity for the simulated galaxy. This curve is meant to emulated (Kuntschner et al., 2010, Figure 12).

radius. With this in mind, we assign the stellar population according to the following phenomenological parameterisation

$$[Z/H] = a \log(r/R_e + b) + c \tag{4.2}$$

with $a = -0.14$, $b = 0.0348$, and $c = -0.27$. It is important to note that this parameterisation describes the observed metallicity which is distinct from the intrinsic metallicity which we need to assign each star a metallicity. To better emphasize this, we plot this intrinsic metallicity and the corresponding projected metallicity in Figure 4.2. As expected, the projected metallicity is lower than the intrinsic metallicity, due to the fact that each line-of-sight has a maximum metallicity given by the intrinsic value, but also has lower metallicity contributions.

The sMILES stellar library provides spectra in units of erg s^{-1} $Å^{-1}$ $M_\odot^{-1}$. We account for the mass dependence by multiplying each spectrum by the particle mass of each stellar particle.

Next, we apply the kinematic information to the spectra. We begin by logarithmically rebinning our spectra to a target velocity scale of 36.4 km s^{-1}. We then Doppler-shift our spectra by shifting the wavelength of each spectrum by a factor $1 + z$ with $z = v/c$ and v the component of the velocity along the line-of-sight. We also divide the total flux by $1 + z$ in order to conserve flux. Next, we reinterpolate this spectrum to the target wavelength range. In order to ensure that we are not affected by the uncertainty in the interpolation, we downsample the spectral resolution by summing adjacent spectral pixels. This has the benefit of maintaining the logarithmic binning, while increasing the velocity scale to 72.8 km s^{-1}. The stellar spectra we use are based on empirical spectra, and thus automatically contain instrumental broadening from the line spread function. We do not investigate the impact of uncertainties in the line spread function nor do we allow for any spatial variations in the line spread function, even though they can in principle have significant variations across a real detector (Hau, 2017, Fig.22-25). The impact of including this may be significant for low dispersion systems where the instrumental dispersion is close to the true dispersion of the galaxy. Another case where this is potentially quite important is when it comes to recovering the higher order GH moments, as a LSF with broad wings has the potential to mimic higher order GH moments like h_4. While interesting, this is beyond the scope of our analysis.

4.5.2 Spectra Template Mismatch and Field of View

Real IFS spectra contain a mixture of errors based on the exposure time of the galaxy (Poisson noise), as well as from systematic sources such as flux calibration uncertainties. Woo et al. (2024) studied the recovery of stellar populations using different spectral fitting codes and assigned a systematic uncertainty to their spectra by taking the residuals from previously fit spectra to real data and co-adding them to their mock spectra. One significant limitation of this, however, is that these systematic uncertainties are not necessarily independent over the sky but can be spatially correlated. Randomly mixing up residuals from real galaxies and uniformly distributing them to mock spectra removes this realistic spatially correlated information.

We consider a new approach. In order to add systematic errors to our spectra while simultaneously maintaining the impact of spatially correlated uncertainties, we adopt both the spatial binning and residuals from real observations in a single mock datacube. We show an example of this in Figure 4.3.

The detailed procedure for this is as follows: we first bin our target galaxy to the target signal-to-noise ratio using the Voronoi tessellation algorithm of Cappellari

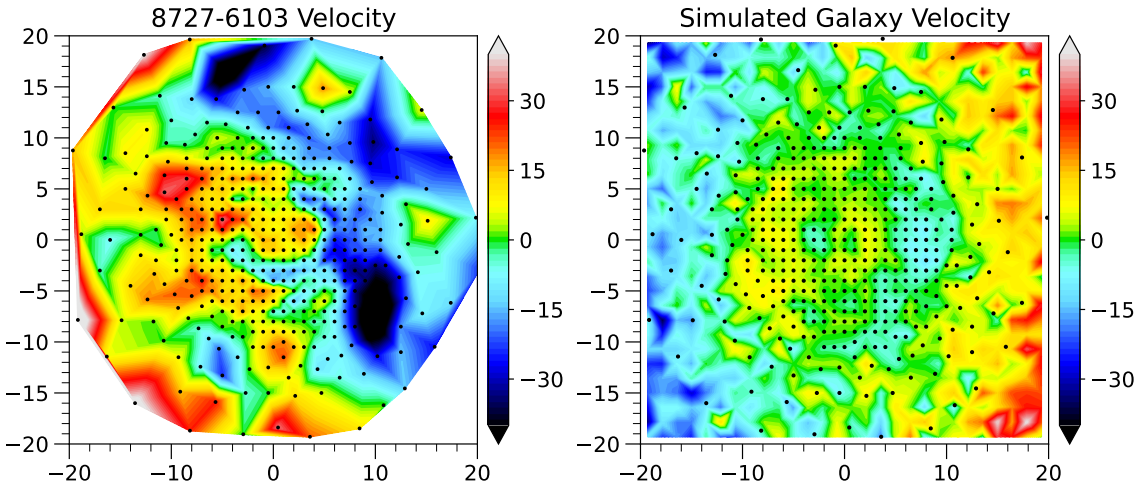

Figure 4.3: The left panel shows the velocity of MaNGA galaxy 8727-6103 binned to a signal-to-noise ratio of 40. The right panel shows the velocity of our simulated galaxy with the same binning scheme imposed on it.

& Copin (2003). For our target galaxy we use galaxy 8727-6103 from MaNGA. We measure the kinematic axis using the routine PAFIT[2] of Krajnović et al. (2006) and rotate the galaxy so that the principal kinematic axis is aligned with the x-axis. We run a preliminary PPXF fit over the sum of all spaxels in the cube to determine the redshift of the galaxy, which we then divide the wavelength by to place the galaxy in its rest frame. Next, we run a proper PPXF run over each binned spaxel of the real galaxy IFS cube and record the wavelength and residuals of each spectrum.

We then superimpose the binning scheme of the observation on top of a projection of our simulation (Figure 4.3). For this work we project the simulated galaxy along the intermediate axis. We scale the binning scheme so that the outermost spaxels have a distance from the origin of $\sqrt{2}R_{\rm e}$. From there, we construct a datacube by summing each of the spectra from each stellar particle within each spaxel. The sMILES stellar spectra only contains data from 3540.5Å to 7409.6Å, so we truncate the spectra to be in the overlapping region between these bounds and the wavelength range of the real IFS cube.

The last step is to interpolate the residual from the true datacube to the wavelength points of the simulated datacube and add the residuals. We show an example of a spectrum before and after adding the residuals in Figure 4.4. It is important to note that adding the residuals also imposes a signal-to-noise ratio equal to that of the original datacube. In practice we bin to a signal-to-noise ratio of 40. It is possible

[2]Available at `https://pypi.org/project/pafit/`.

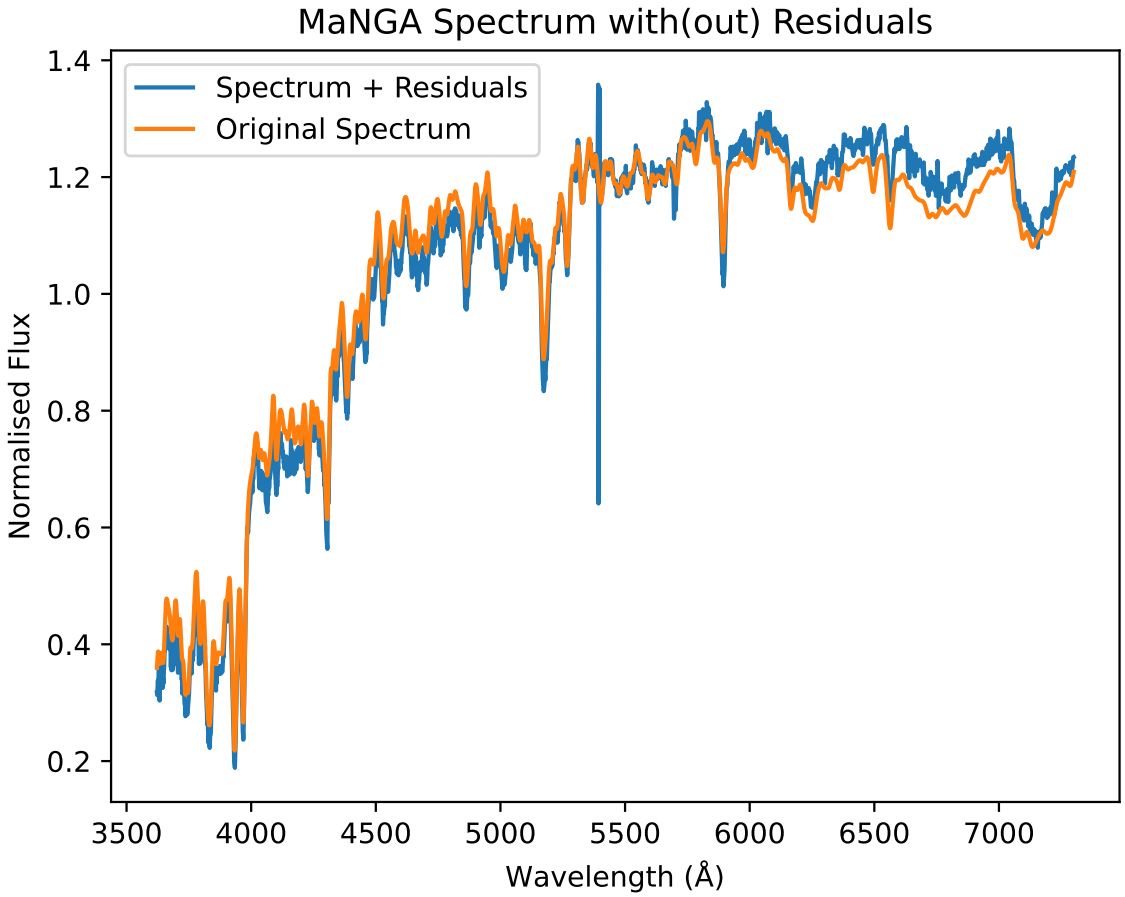

Figure 4.4: Mock datacube spectrum before and after adding residuals from the MaNGA datacube. The MaNGA data have a signal-to-noise ratio of 40.

to further decrease the signal-to-noise ratio by adding white noise. However, our primary interest is systematic offsets due to the uncertainty from template mismatch. Thus we do not do a detailed study of the impact of the signal to noise here.

4.6 Spectral Fitting Approach

4.6.1 pPXF Stellar Templates

Now that we know the true LOSVD of the galaxy, what remains is to compare this with the kinematics measured from the spectra. This requires making several choices. The first choice is which stellar library to use for the extraction. It is important to ensure that the results we derive are generally true for an arbitrary choice of stellar library. As such, we choose four different stellar libraries to fit the datacubes. These are the eMILEs library (Vazdekis et al., 2016), the X-Shooter stellar library (Verro

et al., 2022), the galaxev library (Bruzual & Charlot, 2003), and the fsps stellar library (Conroy et al., 2009; Conroy & Gunn, 2010). Each of these libraries span a grid of ages and metallicities, with each of the above libraries having dimensions of (6,25), (8,26), (5,43), and (9,43), respectively. The typical age range is between 6 Gyr and 10.2 Gyr, and the typical metallicity range is between -1.75 and 0.25. Note that none of these stellar libraries feature alpha-enhanced spectra. This choice is intentional. The goal of this study is to determine how well current methods are able to extract true kinematic information in the presence of realistic template mismatch. As an additional test, we also include the library of individual MILES stars (Sánchez-Blázquez et al., 2006; Falcón-Barroso et al., 2011).

With this choice made, we now have to decide which templates to use to fit the kinematics. The standard choice for fitting stellar kinematics is to first sum up all of the spectra in the mock datacube and perform a pPXF fit to this single spectrum. We then take the linear combination of templates that best describes this average galaxy spectrum and use that single template to fit each of the Voronoi bins individually. We do this for each spectral library and show the fits to the single averaged spectrum for each spectral library in Figure 4.5. This allows us to visually gauge the amount of template mismatch present.

4.7 Measuring Mock Observation Kinematics

4.7.1 Testing the Degree of Additive Legendre Polynomials

In order to determine the optimal number of additive Legendre polynomials to include, we measure the impact on the second moment while holding the total number of GH moments we fit constant and varying the number of additive Legendre polynomials. We fix the number of moments we fit to 2 and set the parameter mdegree for the multiplicative Legendre polynomials to -1 (we do not include them in the fit). We show the result for this in Figure 4.6. We find that across all stellar libraries, the shape of each curve clearly shows that there is an optimal degree of additive Legendre polynomials. The value varies per stellar library but tends to be in the range of 5 to 15.

In a few cases we see that the optimal degree is -1 (that is, additive polynomials are not included). While this is a plausible outcome in our mock observations, it is in principle never recommended to not include additive polynomials for real galaxies as real galaxies often contain spectral light not just from stars, but from gas and other continuous sources such that, fitting with even a perfect stellar library does not

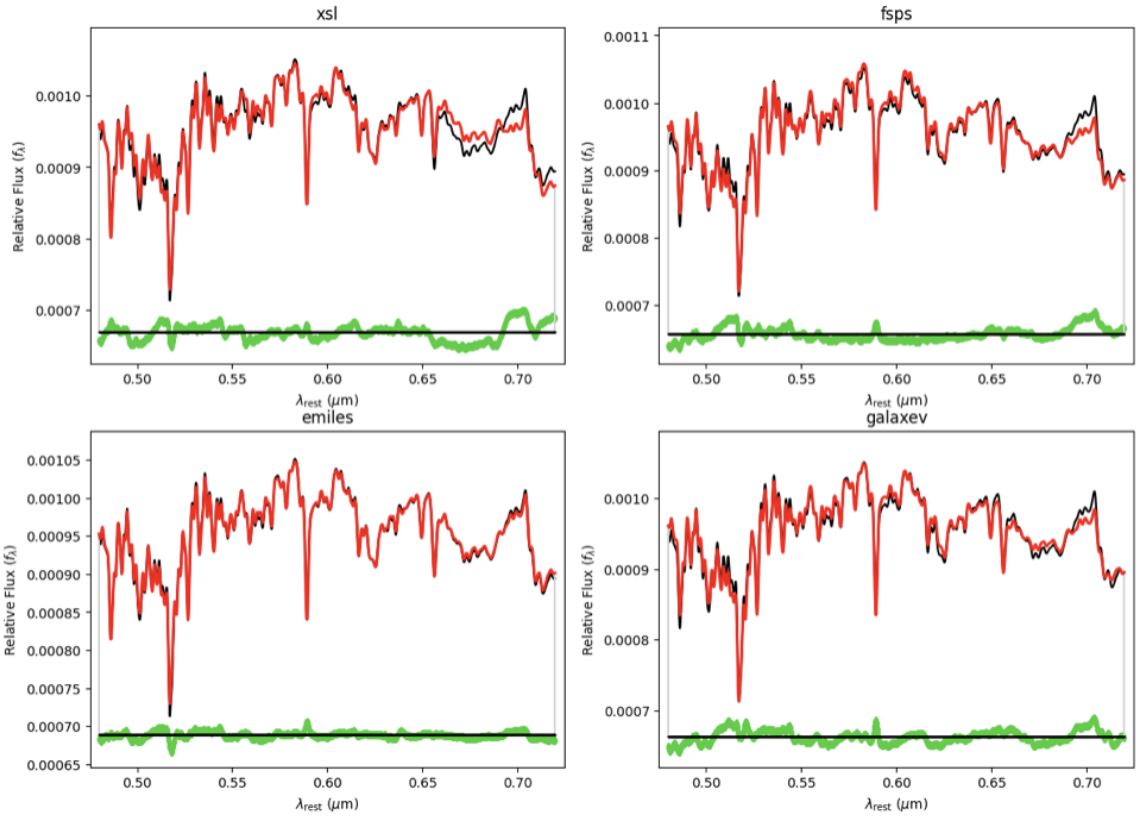

Figure 4.5: pPXF fits to the same stellar spectrum with four different stellar libraries. Each library contains a different amount of mismatch, though all seem to show the largest mismatch at the boundaries of the wavelength region.

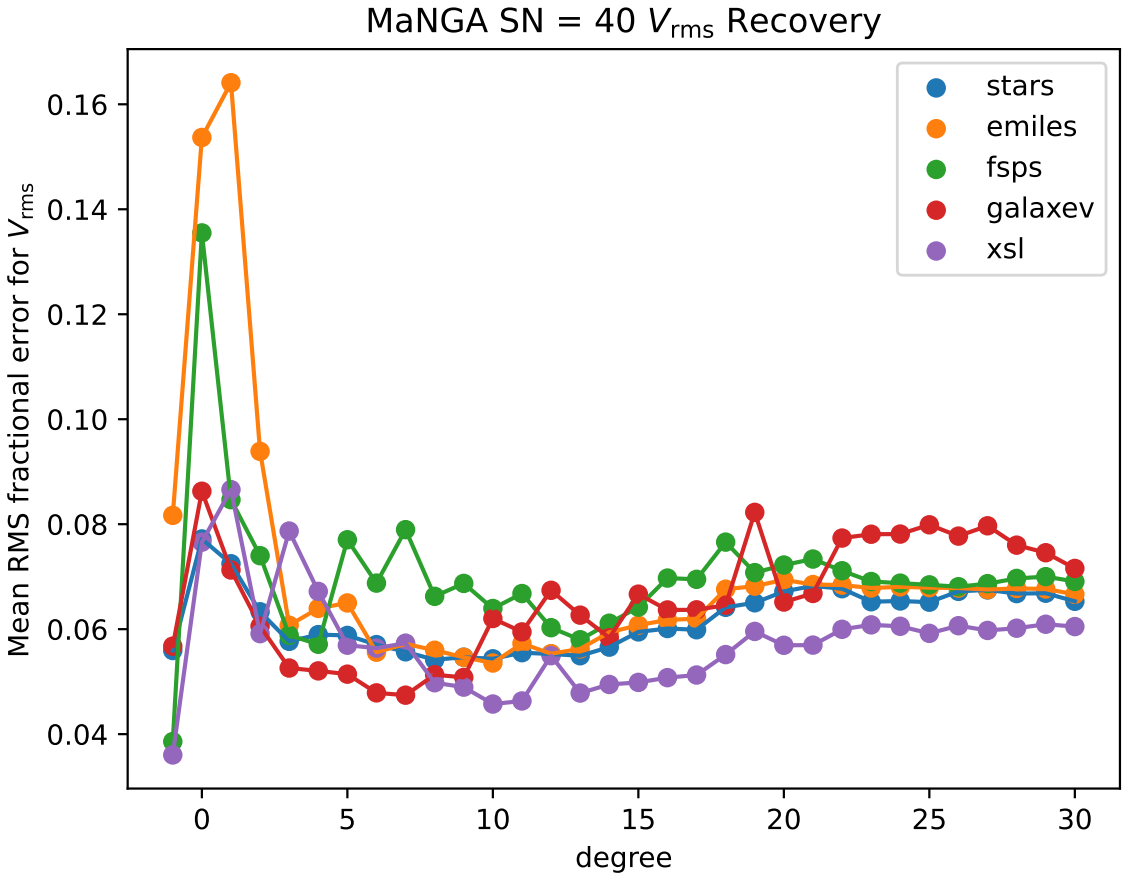

Figure 4.6: Mean root mean square fractional error in $V_{\rm rms}$ as function of the additive Legendre polynomial degree. Each stellar library has its own behavior but they, on average, tend to optimally return the second moment in the range of 5 and 15. Degree equal to -1 corresponds to polynomials not being included in the fit.

provide a complete basis for the possible spectra encountered. A useful example of this comes from chapter 2 where we use additive polynomials explicitly as a tool to measure the light from the AGN of M87.

It is interesting to investigate if we can find metrics that don't make use of the true underlying values of the kinematics that also reveal the optimal number of Legendre polynomials to fit. One attempt at this is in figure 14 of Westfall et al. (2019). There, the authors studied the impact of the number of included additive Legendre polynomials on the recovered σ of the LOSVD (they fit with moments = 2) for MaNGA galaxies. They look at the quantity $\zeta = \sigma_{\rm obs}/\langle\sigma_{\rm obs}\rangle_p$ with the denominator being the mean of all sigma across different choices of Legendre polynomial. They find an initial sharp decrease in the spread of ζ which later flattens out. We perform our own test of this and show the results in Figure 4.7. We find a similar behavior,

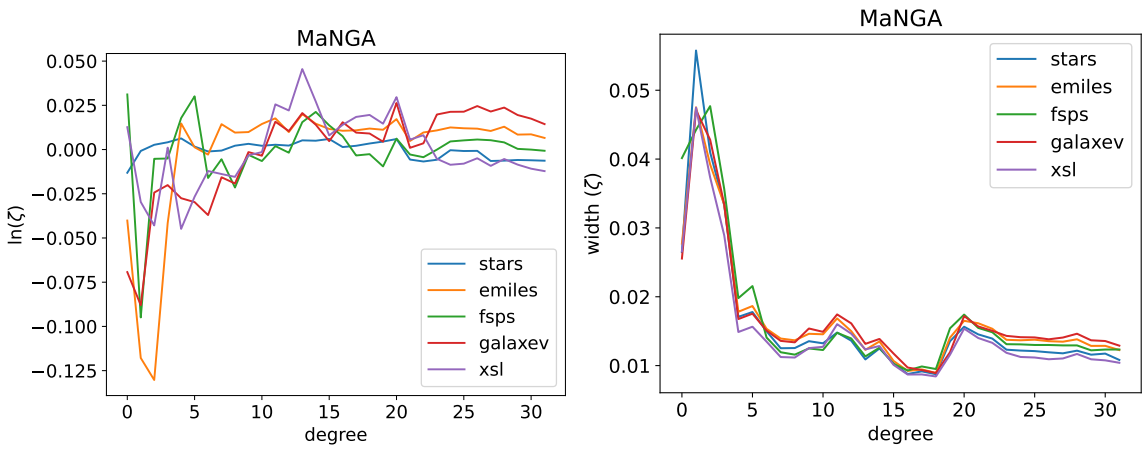

Figure 4.7: The left and right panels show the mean and standard deviation of the parameter ζ for each stellar library. We see that both plots behave very abruptly near the origin but quickly settle down after degree = 5. This is the start of the range of degree values where we find accurate recoveries of the second moment. Degree equal to -1 corresponds to polynomials not being included in the fit.

with a large spike for small values which then plateaus around degree = 5. This is around the minimum value of degree that we find leads to the most accurate second moment recovery. This suggests that making the ζ plot may be a reliable method for determining the lower bound of the optimal range of degree choices for real data.

4.7.2 Testing the Degree of Multiplicative Legendre Polynomials

We also investigate the optimal choice of multiplicative Legendre polynomial degree in Figure 4.8. In the previous case with additive polynomials we clearly saw that for degree > -1 all choices of stellar library show a clear global minimum in the RMS fractional error in $V_{\rm rms}$ that occurs in the range of degree 5 to 15. In this case, however, we find no clear optimal choice of multiplicative Legendre polynomial across the different choices of stellar library. The dependence of the RMS fractional error in $V_{\rm rms}$ varies significantly on the choice of stellar library, with fsps and galaxev having no strong preference for including multiplicative polynomials, while emiles and xsl are able to improve the accuracy by 1-2 per-cent by including multiplicative polynomials. Given these conflicting results, we conclude that it is best not to include multiplicative polynomials when the goal is the optimal recovery of the kinematics.

Previously, the use of multiplicative Legendre polynomials has been typically reserved for measurements of stellar populations. In this case, individual line depths

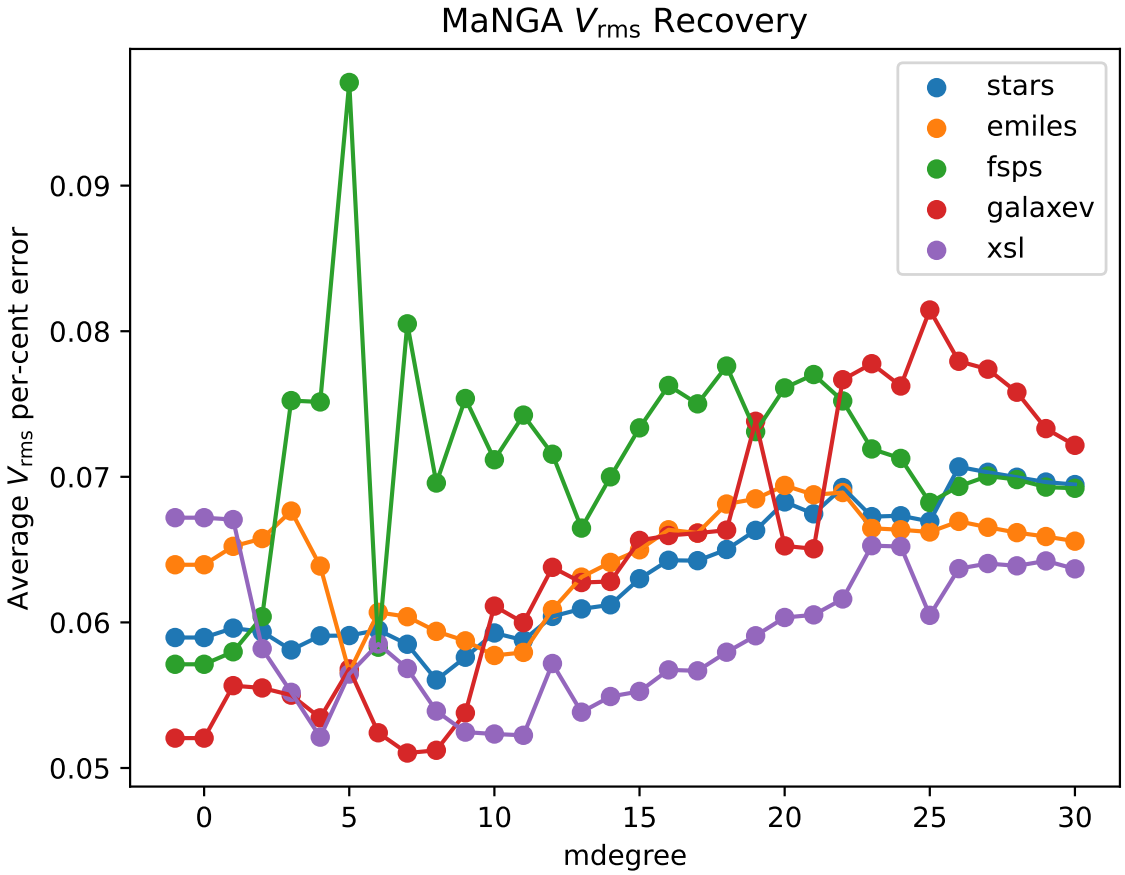

Figure 4.8: Mean RMS fractional error in $V_{\rm rms}$ as function of the multiplicative Legendre polynomial degree. In this case, we see that it is almost always optimal not to include multiplicative polynomials if the objective is to accurately recover the kinematics. mdegree equal to -1 corresponds to polynomials not being included in the fit.

contain important information about the population of stars present in a spaxel. Additive polynomials have the potential to artificially change line depths, whereas multiplicative polynomials ideally keep individual line depths unchanged but scaled. We do not investigate the recovery of stellar populations in this chapter, though we expect that multiplicative polynomials may play a role there, even if they do not meaningfully contribute to the recovery of the kinematics.

4.7.3 Testing the Number of Moments

A wide variety of choices have been made about the maximum number of moments to include in the GH expansion of the LOSVD for pPXF fits. These have ranged from moments = 2 (a pure Gaussian), to moments = 4, all the way up to moments = 8 in some cases. We directly measure the optimal number of moments by running pPXF

fits on our mock datacube using our full selection of stellar libraries, and varying the number of moments we fit. We fix the degree of the additive Legendre polynomials to 7 and do not include multiplicative Legendre polynomials. We show the results for this in Figure 4.9. The results here stand in stark contrast to those of the previous chapter. In the previous chapter we saw that the accuracy in the second moment improved by up to 7 per-cent (though not uniformly) as higher-order moments are included (Figure 3.8). However, in this case, where we have strived to reproduce the most realistic scenario to date involving fitting mock datacubes with realistic template mismatch, we find that the most accurate approach to resolving the second moment comes from moments = 2, the pure Gaussian case. Including up to moments = 4 decreases the accuracy by a few per-cent, and moments = 6 further decrease the accuracy by up to 20 per-cent.

The accuracy of the recovery of the second moment, however, has well-understood downstream effects on the kinematics as equation Equation 5.4 implies that $V_{\rm rms} \propto \sqrt{M}$ with M the mass of the system. Just looking at the net bias in $V_{\rm rms}$, however, is insufficient to draw clear conclusions as it is possible there could be a radially dependent bias which may impact the dynamical models in a non-trivial way. In order to have a more detailed understanding of the biases in dynamical models, we plot the recovery of the second moment as a function of radius for a few choices of the number of moments in Figure 4.10. We average over all of the stellar library choices in this plot. There we clearly see that fitting with higher-order moments results in an overestimated second moment. This over-estimation is approximately constant across the field of view, though the slope does change towards the largest radii. The fact that it is mostly constant across the field of view suggests that the downstream impact on dynamical models is probably close to a constant scaling of the measured quantities.

The most significant contributions to the second moment come from σ and h_4. It is natural to ask which of these are not accurately recovered, leading to the systematically large second moment we recover with increasing numbers of moments. We show a histogram for σ and h_4 in Figure 4.11 and see that both are overestimated when we fit with moments = 4. We see that h_4 is overestimated by 0.02 on average while σ is overestimated by close to 10 km/s. Systematic offsets of this scale and larger have been observed in real galaxy observations. A particularly illustrative example is that of M87, for which different sets of kinematics were compared in Al-Amri et al. (2025). There, they compare the kinematics of chapter 2 with kinematics extracted from Gemini NIFS, as well as their own reduction of JWST NIRSpec data. They

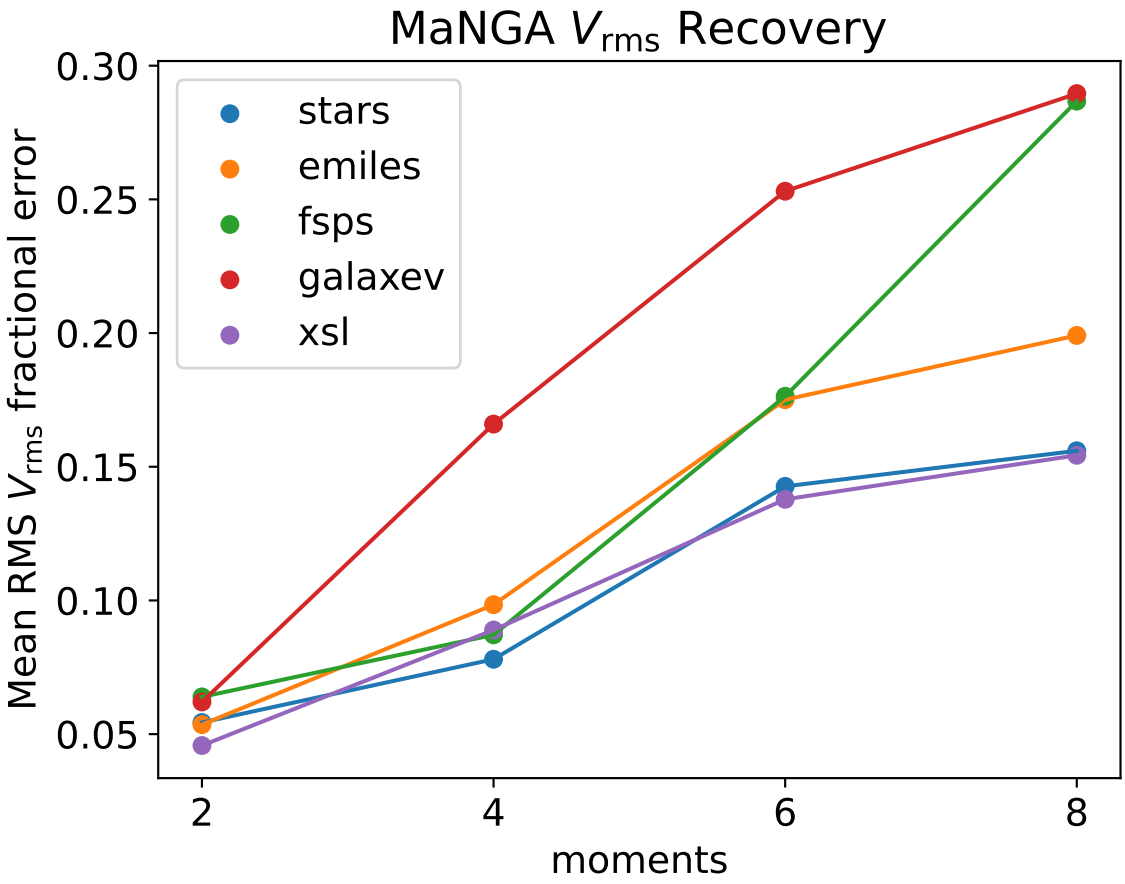

Figure 4.9: The RMS $V_{\rm rms}$ fractional error as a function of the number of GH moments fit. Unlike the idealised case in the previous chapter, here we find that the optimal number of moments to fit is 2, not 4. This is due to the fact that adding the realistic template mismatch decreases the reliability of the fit.

find systematic offsets in σ of up to 50 km/s between NIRSpec and the NIFS data, with the MUSE data having closer to a 30 km/s offset with the NIRSpec observations. M87 is a particularly well resolved galaxy, emphasising the significance of these offsets being observed.

Another illustrative example is NGC 1277, for which initial fits of the GH series out to h_4 found values up to 0.2 (van den Bosch et al., 2012), while a later analysis found that h_4 is almost always less than 0.1 (Walsh et al., 2016). It appears a similar phenomenon may be present in the ATLAS$^{\rm 3D}$ data in Figure 4.1, as many galaxies have large values of h_4, and subsequently appear to have unreliable second moments.

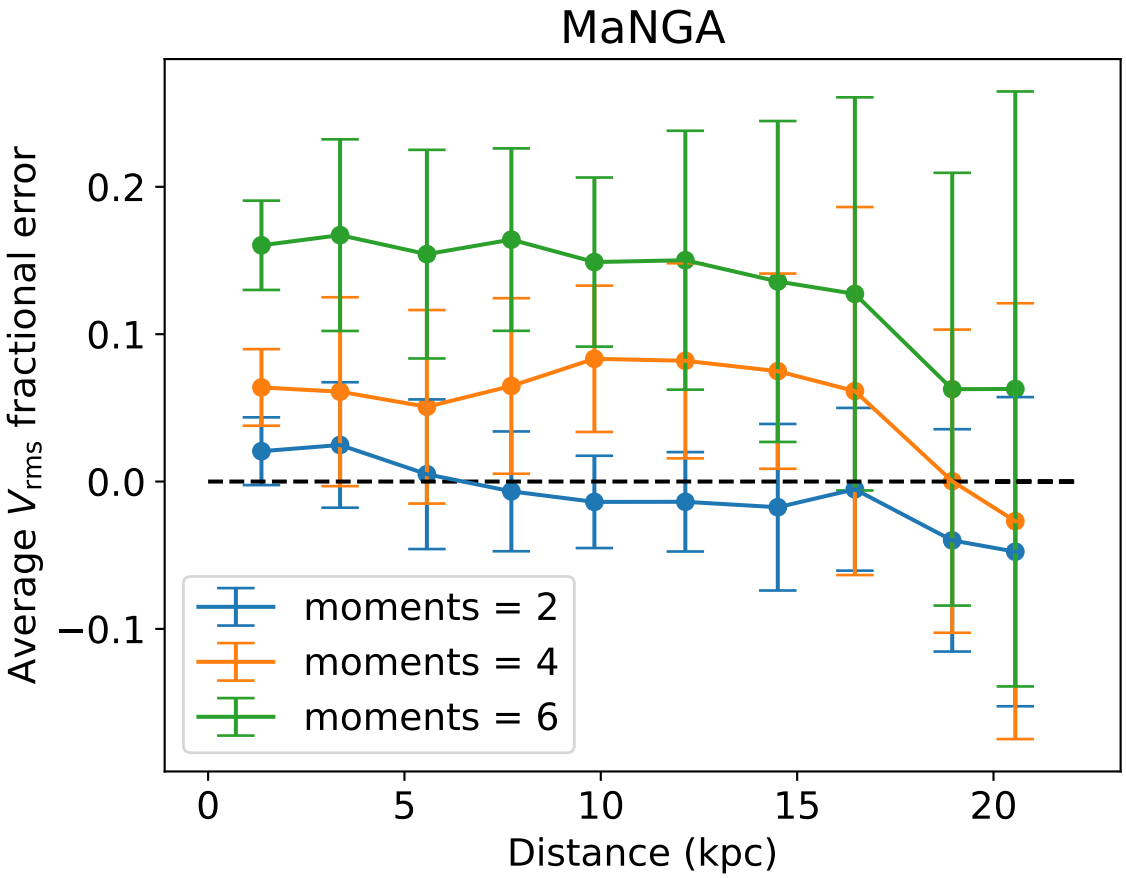

Figure 4.10: The $V_{\rm rms}$ fractional error for different numbers of moments as a function of radius. We see that adding higher-order moments systematically over-estimates the second moment.

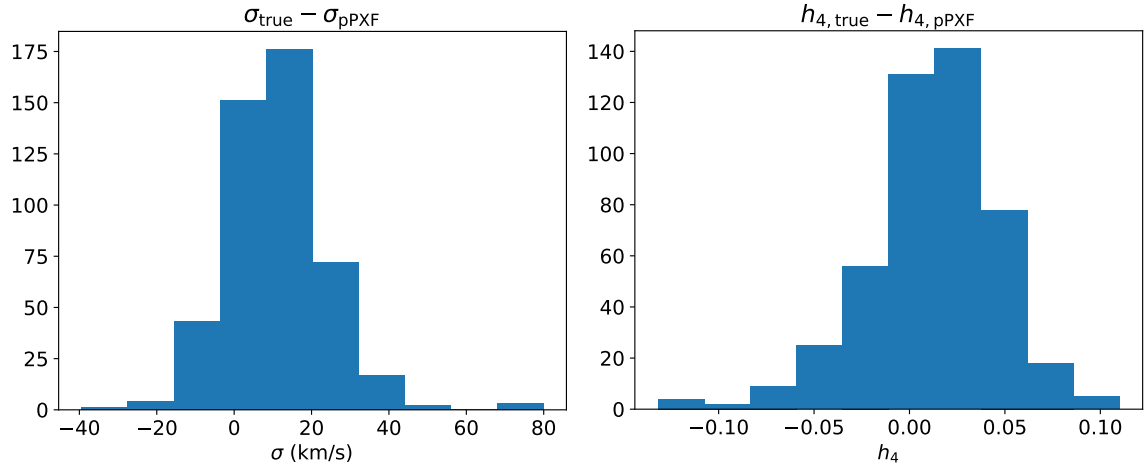

Figure 4.11: The left and right panels show a histogram of the residuals for σ and h_4 when fit with moments = 4. The ground truth we compare to is a histogram fit of the true kinematics with moments = 4. We see that both σ and h_4 are overestimated.

4.7.4 Caveats

It is important to acknowledge the limitations of our study. The first main limitation is that we are only studying a single galaxy. The LOSVD depends on the details of the progenitor galaxies and their merger history. Notably, our galaxy is a slow rotator. Results pertaining to the velocity thus appear most susceptible to change if the galaxy type were to be modified.

Another consideration is the fidelity of our mock datacubes. While we strive to produce datacubes that closely resemble real galaxies, we do not include certain spectral features that can appear in real data such as AGN, ionized gas, or nebular continua. We anticipate this to be well approximated by low-order additive polynomials, but this has never been shown in a convincing way.

In this study we only use the full-spectrum fitting code pPXF for the fitting. One interesting question that we do not address is whether or not it would be possible to improve the accuracy of the recovered kinematics using a larger number of moments but with some regularisation to better control spurious behavior. This is effectively the approach of the non-parametric fitting codes, which we expect are roughly equivalent to a GH expansion with moments = 20 or 30 but with a strong regularisation to prevent overfitting. How to find a reliable regularisation that works well for real data, however, is a challenging problem and one we do not attempt here. This question was partially addressed by Reiter et al. (2025), which found that using a non-parametric LOSVD compared to a GH series out to h_6 had relatively small differences in the recovered M/L ratio and shape of the galaxy, though the orbit composition changed by up to 10 per-cent.

4.8 Conclusions

We summarise our key results as follows:

- We find that the optimal number of Gauss-Hermite moments to fit for accurately recovering $V_{\rm rms}$ is 2 (fitting a pure Gaussian).
- We find that additive Legendre polynomials do improve the fit to the kinematics in the presence of template mismatch. An estimate of what degree to use for real data can be given by producing a ζ diagram such as in Figure 4.7.
- We find that multiplicative Legendre polynomials tend not to significantly improve the recovery of the kinematics.

Chapter 5

Robustness of Black Hole Mass Measurements

5.1 Summary

In the previous chapters we have studied how kinematics measured from mock IFS observations depend on a variety of factors, from the intrinsic shape of the GH series to template mismatch. In this chapter we shift focus to study how well different dynamical modelling methods are able to recover the gravitational potential given a set of kinematics. Here, we focus only on the ideal kinematics determined by directly measuring the LOSVD from histograms in chapter 3. For JAM models this means directly computing the second moment from particles, and for Schwarzschild modelling this means using a fit to the LOSVD histograms in each spaxel up to h_{12}. We find that Axisymmetric Schwarzschild modelling recovers black hole masses with an average accuracy of $M_{\rm bh}/M_{\rm bh,true} = 1.09^{+0.08}_{-0.13}$ and $M_{\rm bh}/M_{\rm bh,true} = 1.09^{+0.06}_{-0.06}$ for DYNAMITE. For Jeans modelling we find recovered black hole masses for $\rm JAM_{cyl}$ of $M_{\rm bh,JAM}/M_{\rm bh,true} = 1.04^{+0.10}_{-0.32}$ and for $\rm JAM_{sph}$ of $M_{\rm bh,JAM}/M_{\rm bh,true} = 0.80^{+0.21}_{-0.06}$. Both dynamical modelling approaches, however, fail to accurately recover the black hole mass when the line of sight is within 30 degrees of the intrinsic long axis of the triaxial simulated galaxy.

5.2 Data

5.2.1 Ideal Kinematics

Previously we studied how the line of sight velocity distribution depends on both the order of truncation of the GH series as well as on various systematic effects present in real measurements from IFS data such as template mismatch. In this chapter we shift

our attention to dynamical modelling and investigate how well Schwarzschild models and JAM models can recover supermassive black hole masses when provided ideal data. For Schwarzschild modelling, we use the histogram fits to the simulated galaxy described in chapter 3. We extract GH moments up to h_{12}. For JAM modelling we use second moments determined by using individual particles in each spaxel.

5.2.2 Uniformly Sampling Projections

In order to improve the robustness of our results, we uniformly sample projections of our galaxy. As our galaxy is triaxial, it is only necessary to sample one octant of projections to describe the full galaxy.

One simple approach to randomly sample this space is to uniformly sample a three-dimensional vector according to a normal distribution and normalise it. This uniformly samples the surface of the sphere but has the downside of requiring a large number of points to obtain statistically certain results. A better approach is to uniformly sample the surface of the sphere such that each sampled point corresponds to an approximately equal surface area.

In general, the problem of generating such a sampling for an arbitrary number of points is only known when the number of points corresponds to that of the Platonic solids, leading to a maximum number of points of 20, otherwise the best configuration is problem-specific.

One algorithm which provides an approximate solution to this for a general choice of the number of points is the spherical Fibonacci lattice (González, 2010), or Fibonacci sphere. The Fibonacci sphere has the property of providing optimal packing, that is, each point occupies approximately equal area. For this reason, the Fibonacci lattice appears regularly in nature. We show two examples of this in an octant of a sphere in Figure 5.1.

Our approach for uniformly sampling projections of the simulated galaxy is as follows. For JAM models we generate a Fibonacci sphere with 1600 points and then restrict the points only to the octant with $x, y, z \geq 0$. This results in 202 points. For Schwarzschild models we repeat this but with 75 points (resulting in 11 points). We use fewer points for the Schwarzschild models as it takes much longer to run a Schwarzschild model. We then align the galaxy so that the origin is at zero and the intrinsic major axis is aligned in the direction of each point on the Fibonacci sphere.

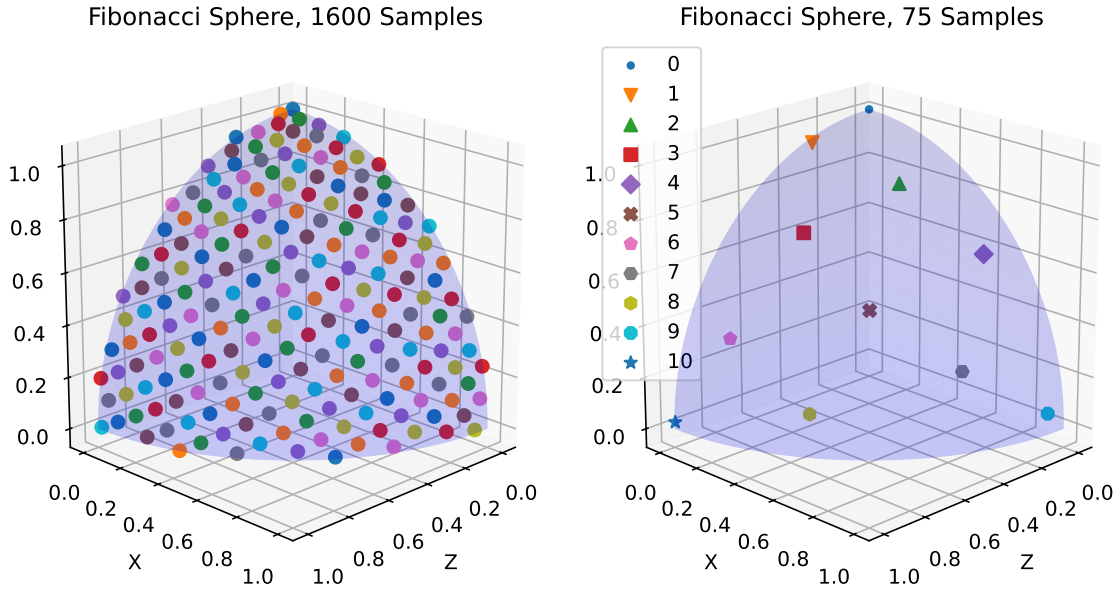

Figure 5.1: Demonstrations of the Fibonacci lattice projected on an octant of a sphere where the total number of points sampled across the whole sphere is 1600 and 75, respectively.

5.3 Dynamical Modelling Methods

5.3.1 Approach

The use of simulated data gives us the special opportunity to do dynamical modelling where we know what the true result is. For this reason, we test several different dynamical modelling methods, namely, Jeans Anisotropic Modelling (Cappellari, 2008, 2020), Axisymmetric Schwarzschild modelling van der Marel et al. (1998); Cappellari et al. (2002, 2006), and triaxial Schwarzschild modelling with the code DYNAMITE[1] (van den Bosch et al., 2008; Thater et al., 2022b). Our goal is to test each of these methods against each other and to see under which conditions each code is able to best model the simulated galaxy.

5.3.2 Jeans Axisymmetric Modelling

For our Jeans analysis we use the code JAMPY[2] of Cappellari (2008, 2020). We start by briefly introducing the method. The evolution of a collisionless stellar system can

[1]Available at `https://github.com/dynamics-of-stellar-systems/dynamite`, and based on the code van den Bosch et al. (2008), available at `https://github.com/remco-space/TriaxSchwarzschild`.

[2]Available at `https://pypi.org/project/jampy/`.

be modelled using the collisionless Boltzmann equation (Binney & Tremaine, 2011, Eq. 4.11)

$$\frac{\partial f}{\partial t} + \sum_{i=1}^{3} \left(v_i \frac{\partial f}{\partial x_i} - \frac{\partial \Phi}{\partial x_i} \frac{\partial f}{\partial v_i} \right) = 0 \tag{5.1}$$

Where t is time, x_i is position and v_i is velocity, Φ is the gravitational potential, and f is the 6-dimensional distribution function describing the probability of finding a star at a given position x_i and velocity v_i. Assuming that the system is stationary gives $\partial f / \partial t = 0$. Practically, it is often difficult to deal explicitly with the distribution function (DF) f for external galaxies since the DF is 6-dimensional but we are only able to observe 3-dimensional data, namely the 2D surface brightness and line-of-sight velocity distribution via integral field data. We thus reduce this problem to one better suited for the data in three steps. First, we convert Equation 5.1 to either cylindrical coordinates or spherical coordinates as these better reflect the symmetries of real galaxies. Next, we assume axisymmetry and drop terms that involve derivatives with respect to ϕ. This gives the two equations

$$v_R \frac{\partial f}{\partial R} + v_z \frac{\partial f}{\partial z} + \left(\frac{v_\phi^2}{R} - \frac{\partial \Phi}{\partial R} \right) \frac{\partial f}{\partial v_R} - \frac{\partial \Phi}{\partial z} \frac{\partial f}{\partial v_z} - \frac{v_R v_\phi}{R} \frac{\partial f}{\partial v_\phi} = 0 \tag{5.2}$$

for cylindrical coordinates, and

$$\begin{aligned} & v_r \frac{\partial f}{\partial r} + \frac{v_\theta}{r} \frac{\partial f}{\partial \theta} + \left(\frac{v_\theta^2 + v_\phi^2}{r} - \frac{\partial \Phi}{\partial r} \right) \frac{\partial f}{\partial v_r} \\ & + \frac{1}{r} \left(\frac{v_\phi^2}{\tan\theta} - v_r v_\theta - \frac{\partial \Phi}{\partial \theta} \right) \frac{\partial f}{\partial v_\theta} - \frac{v_\phi}{r} \left(v_r + \frac{v_\theta}{\tan\theta} \right) \frac{\partial f}{\partial v_\phi} = 0 \end{aligned} \tag{5.3}$$

for spherical coordinates. Lastly, in order to write this in terms of observable quantities (i.e. the velocity moments) we multiply each of these equations by v_R and v_z, and v_r and v_θ, respectively, and integrate over all velocities to get (Binney & Tremaine, 2011, Eqs. 4.222a-b)

$$\frac{\nu \overline{v_R^2} - \nu \overline{v_\phi^2}}{R} + \frac{\partial (\nu \overline{v_R^2})}{\partial z} + \frac{\partial (\nu \overline{v_R v_z})}{\partial z} = -\nu \frac{\partial \Phi}{\partial R} \tag{5.4}$$

$$\frac{\nu \overline{v_R v_z}}{R} + \frac{\partial (\nu \overline{v_z^2})}{\partial z} + \frac{\partial (\nu \overline{v_R v_z})}{\partial R} = -\nu \frac{\partial \Phi}{\partial z} \tag{5.5}$$

for cylindrical coordinates and (de Zeeuw et al., 1996, eq. 2.4)

$$\frac{\partial (\nu \overline{v_r^2})}{\partial r} + \frac{1}{r} \left[\frac{\partial (\nu \overline{v_r v_\theta})}{\partial \theta} + 2\nu \overline{v_r^2} - \nu \overline{v_\theta}^2 - \nu \overline{v_\phi^2} + \frac{\nu \overline{v_r v_\theta}}{\tan\theta} \right] = -\nu \frac{\partial \Phi}{\partial r} \tag{5.6}$$

$$r \frac{\partial (\nu \overline{v_r v_\theta})}{\partial r} + \frac{\partial (\nu \overline{v_\theta^2})}{\partial \theta} + 3\nu \overline{v_r v_\theta} + \frac{\nu \overline{v_\theta^2} - \nu \overline{v_\phi^2}}{\tan\theta} = -\nu \frac{\partial \Phi}{\partial \theta} \tag{5.7}$$

for spherical coordinates. Here we use the notation

$$\nu\overline{v_i v_j} = \int v_i v_j f d^3\mathbf{v} \tag{5.8}$$

to denote the integral over all of the velocity components.

Each of these sets of two equations has 4 unknown parameters. We thus require two more assumptions in order for there to be a unique solution. The first assumption we make is that the velocity ellipsoid is aligned with the same coordinate system as the equations. Practically, this means that the cross terms $\nu\overline{v_R v_z}$ and $\nu\overline{v_r v_\theta}$ are equal to zero.

The last assumption we make is on the shape of the velocity anisotropy. For this, we have two choices. First, for cylindrical and spherical coordinates, respectively, we take

$$\overline{v}_R^2(1-\beta_z) = \overline{v_z^2} \tag{5.9}$$

$$\overline{v_r^2}(1-\beta) = \overline{v_\theta^2} \tag{5.10}$$

with β_z and β constant[3].

An alternative is to assume a special form of β which does not increase the number of quadratures in solving the Jeans equations. Simon et al. (2024) found that using $\mathrm{JAM_{sph}}$ with an anisotropy of the form

$$\beta(r) = \beta_0 + \frac{\beta_\infty - \beta_0}{1 + (r_a/r)^\alpha} \tag{5.11}$$

allowed good fits to the slow rotator galaxy M87. This profile was previously used in (Baes & van Hese, 2007, eq. 30). Note that for $\alpha = 2$ this reduces to the Osipkov-Merrit profile (Osipkov, 1979; Merritt, 1985), and for $\alpha = 1$ this reduces to the homographic anisotropy of Bacon (1985). We refer to this as logistic as this function is simply the logistic function for $\log(r/r_a)$. Note that this option only makes sense for $\mathrm{JAM_{sph}}$, as that is the case where we believe a radially varying anisotropy profile is physically justified.

In summary, there are three different possible model choices for our Jeans analysis. There are both cylindrically aligned and spherically aligned models (which we call $\mathrm{JAM_{cyl}}$ and $\mathrm{JAM_{sph}}$) with constant anisotropy, and there is $\mathrm{JAM_{sph}}$ with varying anisotropy according to the logistic profile. Practically, however, only two of the

[3]It can be shown that it is only necessary to make the assumption of constant anisotropy on each individual component in an MGE expansion (see Cappellari (2008) for details). However, here we only consider the simple case of global constant anisotropy.

models are physically meaningful. We know that a varying anisotropy is necessary to describe slow rotators so for that reason we only run $\mathrm{JAM_{sph}}$ with the varying anisotropy[4].

In summary, there are two JAM models that we use. $\mathrm{JAM_{cyl}}$ with constant anisotropy and $\mathrm{JAM_{sph}}$ with varying anisotropy. We test each of these choices in section 5.4 and describe their respective advantages and disadvantages.

5.3.3 Limitations of JAM Modelling

JAM models are not guaranteed to provide a solution that has a non-negative distribution function. In general there is not a simple way to check if the distribution function has some negative component, but there are some approaches that suggest something has gone wrong. One such approach is checking for violations of the slope-anisotropy theorem (An & Evans, 2006). This theorem states that if the density of a gravitating system is cusped like $r^{-\gamma}$ then the limiting value of the anisotropy parameter $\beta = 1 - \frac{v_T^2}{2v_r^2}$ at the center cannot be more than $\gamma/2$. Thus for a Hernquist profile one would expect a maximum β in the centre of 0.5. For a perfectly flat cored galaxy this would give $\beta = 0$. Additionally, Wang et al. (2021) showed that JAM models with $\beta > 0.7\epsilon_{\mathrm{intr}}$ correspond to solutions with negative distribution function. Here it is straightforward to determine that the distribution function is negative as the model second moments themselves are negative.

However, even if positivity of the distribution function was guaranteed, there is a more fundamental issue with JAM modelling which is that real galaxies cannot be globally described by any simple coordinate system like spherical or cylindrical coordinates Evans et al. (2016).

Given these limitations, it is important to test how well JAM models recover the gravitational potential of simulated stellar systems, and to ensure that JAM models do not clearly have a negative distribution function, either through checking the slope-anisotropy theorem or checking for negative moments in the JAM models.

5.3.4 Axisymmetric Schwarzschild Modelling

One natural approach to dynamical modelling is to use individual stellar orbits as basis functions to reconstruct the entire galaxy. This idea was first proposed by

[4]We ran tests measuring the black hole mass assuming a constant anisotropy profile and always found that the recovered black hole is a factor of two or smaller than the true value.

Martin Schwarzschild in Schwarzschild (1979). When combined with requiring a self-consistent description of the galaxy (Richstone & Tremaine, 1984) and regularisation to improve the behavior of the models (Richstone & Tremaine, 1988), this proved a powerful tool for studying the dynamical properties of galaxies.

The first implementation of Schwarzschild's orbit superposition method that fit the full line-of-sight data came in Rix et al. (1997) which made spherical models with the LOSVD parameterised using a Gauss-Hermite series. This was then generalised to the axisymmetric case in van der Marel et al. (1998) and Cretton et al. (1999). Other independent implementations of axisymmetric Schwarzschild modelling were presented in Gebhardt et al. (2000a) (sometimes referred to as the NUKER code), as well as the code MASMOD of Valluri et al. (2004). These codes were originally designed to fit long slit spectroscopic observations, though they were later updated to fit integral field data. Another independent code was that of Cappellari et al. (2002, 2006) which shares some code with that of van der Marel et al. (1998), as described in Sec. 3.4 of Cappellari et al. (2006). This is sometimes referred to as the LEIDEN code (though it should not be confused with other independent codes that were developed in Leiden, such as that of Cretton et al. (1999)). This code has previously been used extensively, first as a part of the SAURON project (Cappellari et al., 2007), and later in a number of black hole studies (Shapiro et al., 2006; Cappellari et al., 2009; Krajnović et al., 2009, 2018; Ahn et al., 2018; Thater et al., 2019). This is the code that we use for our axisymmetric modelling.

The first step of any Schwarzschild code is to generate a library of orbits that will be used to fit some observed kinematics. There are two popular choices for this. The first is to generate the orbits by sampling the three integrals of motion. This is typically done by generating a grid logarithmically spaced in energy E with a maximum value equal to the energy of the circular orbit at the maximum radius considered, and linearly spaced in angular momentum L_z with maximum value given by the angular momentum of the circular orbit at a given energy. The third integral is more difficult to sample since it typically does not have a closed form. The standard approach is to note that each orbit with three isolating integrals touches the zero velocity curve a finite number of times. The third integral can thus be varied by sampling points on the zero velocity curve (van der Marel et al., 1998, Fig. 5). A second approach, and the approach taken in our code, is to uniformly sample the sky in polar coordinates and for each location uniformly sample the energy (Cappellari et al., 2006). We set the number of energy samples to be 21, the number of angular samples equal to 8, and the number of radial samples equal to 7. These values have

previously been used in a number of studies, principally in the study of ATLAS^{3D} galaxies Cappellari et al. (2006). We ran some tests increasing the number of orbits and found that it did not significantly change our results. The radii are logarithmically spaced between 0.3 kpc and 316 kpc. This allows us, at the bottom range, to probe within the sphere of influence and, at the top range, to include stars out to over 20 times the half-mass radius.

Both of the codes we use make use of orbit dithering to improve the orbit sampling, allowing for a smoother, representative dynamical model. This method was originally employed in a spherical system in Richstone & Tremaine (1988); Rix et al. (1997), and later expanded to all three integrals of motion in Cappellari et al. (2006). This method was adopted unchanged for triaxial geometry in van den Bosch et al. (2008). Our axisymmetric code allows for an even number of dithers so we set the number of dithers to the default value of 6. For this sampling method the orbital starting points are first calculated and then integrated using a Runge-Kutta method.

The contribution to the kinematics from each individual dithered orbit bundle can then be calculated in the same field of view as the observation aperture by projecting each dither onto the aperture. This generates a velocity histogram for each orbit and spaxel. Once projected, the higher order Gauss-Hermite moments can be measured using Equation 3.7 where the velocity and σ are fixed to the values given by the data in the corresponding spaxel.

It is worth emphasising that each of these stars is treated as test particles in the galaxy potential. To ensure self-consistency, it is necessary to require that any candidate superposition of orbits recreates the stellar tracer density, to a given accuracy. This is done at two levels. Within the range of the observing aperture, the orbits are directly projected onto the aperture and are required to recreate the observed MGE surface brightness and density. Outside of the range of the observed data, a polar grid is created to ensure that the surface brightness and density match the MGE outside of the observation aperture.

Lastly, to determine the optimal orbital weights, the code runs a non-negative least squares (NNLS) fit to the χ^2 (Lawson & Hanson, 1995). The χ^2 is defined as

$$\chi^2 = \chi^2_{\rm phot} + \chi^2_{\rm kin} + \chi^2_{\rm reg} \tag{5.12}$$

where the photometric $\chi^2_{\rm phot}$ is given by

$$\chi^2_{\rm phot} = \sum_{i=1}^{N_p} \left(\frac{M_i^{\rm obs} - \sum_k \gamma_k M_i^k}{\Delta M_i^{\rm obs}} \right)^2 \tag{5.13}$$

where N_p is the number of photometric constraints, γ_k is the weight of an individual orbit, $M_i^{\rm obs}$ is the observed mass constraint in aperture i, and M_i^k is the contribution of orbit k to the mass constraint in aperture i. The mass constraints typically include the surface brightness in each IFU aperture, surface brightnesses from a wide field polar grid, and 3D densities in the same wide field polar grid. The kinematic $\chi^2_{\rm kin}$ is defined as (Rix et al., 1997, Eq. 13)

$$\chi^2_{\rm kin} = \sum_{i=1}^{N_k} \sum_{m=1}^{M} \frac{M_i^{\rm obs} h_{m,{\rm obs}}^{(i)} - \sum_k \gamma_k M_i^k h_m^{(i,k)}}{\Delta(M_i^{\rm obs} h_m^{(i,k)})} \tag{5.14}$$

where $h_{m,{\rm obs}}^{(i)}$ is the observed GH moment h_m in aperture i, and $h_m^{(i,k)}$ is the GH moment h_m of the kth orbit in aperture i. Lastly, there is a regularisation term that is used to smooth the distribution function and guarantee a unique solution even when the number of fitted orbits is much greater than the number of constraints. The form of this regularisation is (Rix et al., 1997, Eq. 15)

$$\chi^2_{\rm reg} = \lambda \sum_{i=1}^{N_o} (\gamma_k - \frac{1}{P} \sum_{p=1}^{P} \gamma_{k_p}) \tag{5.15}$$

where the P orbits γ_{k_p} are the immediate neighbors of orbit k in phase space.

One key subtlety with Schwarzschild modelling is that, while the above χ^2 is minimized internally in the code, this is typically not the χ^2 that is used for generating contours or finding best fit parameters. Typically, the χ^2 chosen for this is

$$\chi^2 = \sum_{i=1}^{N} \left[\left(\frac{V_i^{\rm obs} - V_i^{\rm model}}{\Delta V_i^{\rm obs}} \right)^2 + \left(\frac{\sigma_i^{\rm obs} - \sigma_i^{\rm model}}{\Delta \sigma_i^{\rm obs}} \right)^2 + \sum_{m=3}^{M} \left(\frac{h_{m,i}^{\rm obs} - h_{m,i}^{\rm model}}{\Delta h_{m,i}^{\rm obs}} \right)^2 \right] \tag{5.16}$$

This is chosen as it directly relates to the key kinematic parameters measured. This is the χ^2 that we show in our subsequent results for our Schwarzschild models.

We have made several changes to the code to improve its usability. All of the major routines of the code are written in Fortran 90 with a wrapper that was originally written in the Interactive Data Language (IDL). We have rewritten this wrapper in Python and modified the input method to mimic the approach taken by DYNAMITE, allowing greater compatibility between the two methods. Parts of the conversion of the IDL code to Python were done using AI. To ensure that no errors were introduced during this process, we ran both the original IDL code as well as the python code on the same data and confirmed that the output was identical. We have replaced many of the original input files with a single master configuration file that allows the user, in one place, to specify the orbital library parameters, gravitational potential

parameters, etc. The implementation of an MGE dark halo to the code was done by Youquan Fu and Haitong Wang.

It was previously observed in Quenneville et al. (2021) that the triaxial code of van den Bosch et al. (2008) spends a significant amount of time applying the PSF to the orbit library. Given that the code of van den Bosch et al. (2008) shared some development with our axisymmetric code, we performed a test using our modified version of the axisymmetric code to see if a similar bug was present in the evaluation of the PSF in our axisymmetric code. We found that running a single Schwarzschild model with a PSF takes around 7 hours to run, whereas modifying the code to avoid running the PSF results in the code taking 5 hours to run. The issue stems from the RANDOM_GAUSS subroutine of the orblib.f90 file. This subroutine uses a rejection sampling implementation of the Box-Muller transform as it is presented in Numerical Recipes (Press, 2007). Quenneville et al. (2021) solves this issue by implementing the trigonometric form of the Box-Muller transformation, claiming that this vectorizes the code, resulting in their observed speedup. We are primarily interested in modelling massive well-resolved galaxies and determining the pure model systematic uncertainties. The impact the PSF has on this becomes increasingly significant as it approaches the size of the black hole sphere of influence. While studying the impact of the PSF resolution on the recovery of the gravitational potential of our mock galaxy is interesting, it is beyond the scope of our current study. Because we do not use a PSF in any of our mock observations, we have modified the code so that when the PSF sigma is equal to 0 the PSF module skips over RANDOM_GAUSS and returns the orbital points unchanged.

5.3.5 Triaxial Schwarzschild Modelling

The first triaxial code to fit both kinematics and the tracer distribution was van den Bosch et al. (2008). This code has been used extensively for black hole studies (van den Bosch et al., 2012; Walsh et al., 2015, 2016), as part of the analysis of the CALIFA survey (Zhu et al., 2018), and for studying the orbit distribution of MaNGA galaxies (Jin et al., 2020).

It was discovered in Quenneville et al. (2022) that this code contained a bug in the orbit mirroring. Thater et al. (2022b) found that the impact this had on dynamical analyses was small compared to the statistical and systematic errors. However, this resulted in the code being forked, forming the basis for the TriOS code of Quenneville et al. (2022), as well as the basis of the code DYNAMITE Jethwa et al. (2020); Thater et al. (2022b).

Several other triaxial codes have been developed such as SMILE Vasiliev (2013) and its successor FORSTAND (Vasiliev & Valluri, 2020), as well as the code SMART of Neureiter et al. (2021). For our study, we use the code DYNAMITE. DYNAMITE is primarily a Python wrapper for the original Fortran code of van den Bosch et al. (2008), though it has now expanded to include a number of new features, including a treatment of barred galaxies (Tahmasebzadeh et al., 2022, 2024), and the classification of orbit components (Zhu et al., 2018; Santucci et al., 2022).

DYNAMITE is qualitatively similar to the axisymmetric version with a few caveats. The orbit library initialises the set of orbits in the same way as in the axisymmetric code. One issue with this, however, is that it does not generate a significant number of box orbits. Box orbits are known to be very important for the support of triaxial galaxies (Schwarzschild, 1979; Hunter & de Zeeuw, 1992), so DYNAMITE increases their number by explicitly sampling additional box orbits. This is done by generating the equipotential for each energy E and linearly sampling points in the spherical angles θ and ϕ. (van den Bosch et al., 2008). This code also allows for dithering. We set the number of dithers equal to the default value of 5. We use the same range of radii as in the axisymmetric code.

5.3.6 Multi-Gaussian Expansion

The deprojection of an axisymmetric system is unique only in the edge-on case (Rybicki, 1987; Kochanek & Rybicki, 1996), and becomes very severe when the galaxy is close to face-on (Gerhard & Binney, 1996; Romanowsky & Kochanek, 1997; van den Bosch, 1997; Magorrian, 1999). In the triaxial case the situation is compounded and the deprojection is never unique (Gerhard, 1996). One natural choice of deprojection is a multi-Gaussian expansion (Monnet et al., 1992), which has the nice property of allowing the roundness of the model to be enforced (Cappellari, 2002).

For both our Jeans models and Schwarzschild models, the stellar tracer distribution is parameterised using a multi-Gaussian expansion (MGE) (Emsellem et al., 1994; Cappellari, 2002). Under this scheme, the surface brightness is written as a sum of Gaussians

$$\Sigma(R', \theta') = \sum_{k=1}^{N} I_k \exp\left[-\frac{1}{2\sigma_k'^2}\left(x'^2 + \frac{y'^2}{q_k'^2}\right)\right] \tag{5.17}$$

with

$$x' = R' \sin(\theta' - \psi_k') \tag{5.18}$$

$$y' = R' \cos(\theta' - \psi_k') \tag{5.19}$$

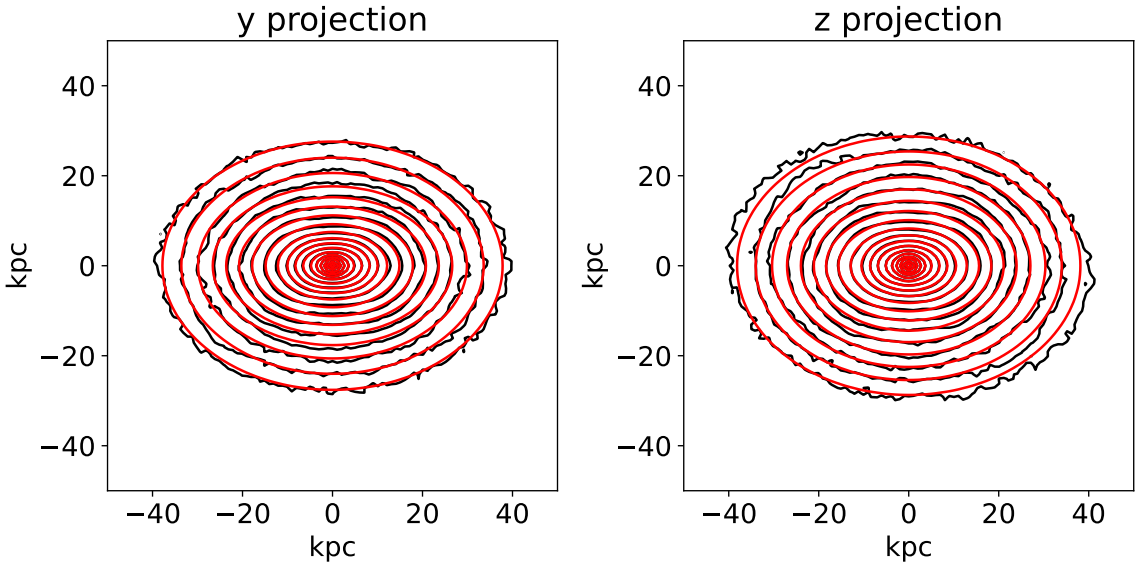

Figure 5.2: The left and right panels show the fit of the MGE contours to our mock photometry. The y projection has no observed twist so we fit it assuming that all MGEs are aligned with the same axis. For the z projection, we see that almost all MGE contours are aligned except for the outer two which experience a very small amount of twist.

where ψ'_k is the position angle of the kth Gaussian component, and where I_k is given by

$$I_k = \frac{L_k}{2\pi\sigma_k^2 q'_k} \tag{5.20}$$

with L_k the observed total luminosity of the kth Gaussian component and q'_k is the observed axial ratio satisfying $0 \leq q'_k < 1$. Here primed terms correspond to projected quantities whereas unprimed terms are the corresponding deprojected quantities.

We fit MGEs to our simulation by creating mock photometric observations. We then use the MGEFIT[5] package of Cappellari (2002). As we see in equation Equation 5.31, the deprojection is limited to values of the inclination satisfying $\min(q'^2_k) > \cos^2 i$. In order to allow for the largest range of possible deprojections, we use the MGE_FIT_SECTORS_REGULARIZED routine which increases the minimum value of q' if it does not significantly harm the χ^2 of the MGE fit. We show the results of our MGE fits for the y and z projections in Figure 5.2. We note that neither projection appears to require significant isophotal twist to match the photometry. For this reason, we proceed assuming that there is no twist.

[5]Available at `https://pypi.org/project/mgefit/`.

5.3.7 Triaxial Deprojection

If we assume an intrinsically triaxial geometry, then the intrinsic density in the MGE formalism is given by

$$\rho(x,y,z) = \sum_{k=1}^{N} \frac{L_k}{(\sigma_k\sqrt{2\pi})^3 q_k p_k} \exp\left[-\frac{1}{2\sigma_k^2}\left(x^2 + \frac{y^2}{p_k^2} + \frac{z^2}{q_k^2}\right)\right] \tag{5.21}$$

The deprojection of a triaxial ellipsoid is given by (Cappellari, 2002, Eqs. 7-9)

$$1 - q_k^2 = \frac{\delta_k'[2\cos 2\psi + \sin 2\psi(\sec\theta\cot\phi - \cos\theta\tan\phi)]}{2\sin^2\theta[\delta_k'\cos\psi(\cos\psi + \cot\phi\sec\theta\sin\psi) - 1]} \tag{5.22}$$

$$p_k^2 - q_k^2 = \frac{\delta_k'[2\cos 2\psi + \sin 2\psi(\cos\theta\cot\phi - \sec\theta\tan\phi)]}{2\sin^2\theta[\delta_k'\cos\psi(\cos\psi + \cot\phi\sec\theta\sin\psi) - 1]} \tag{5.23}$$

$$\sigma_k^2 = \frac{\sigma_k'^2 q_k'}{\sqrt{p_k^2\cos^2\theta + q_k^2\sin^2\theta(p_k^2\cos^2\phi + \sin^2\phi)}} \tag{5.24}$$

with $\delta_k' = 1 - q_k'^2$ and where the angles θ and ϕ are the polar coordinates that determine the orientation of the line-of-sight with respect to the principal axes of the object, and ψ denotes the rotation along the line-of-sight. The angles θ, ϕ, and ψ are related to the intrinsic shape parameters q, p, and u via (van den Bosch et al., 2008, eq. 10)

$$\cos^2\theta = \frac{(u^2 - q^2)(q'^2u^2 - q^2)}{(1 - q^2)(p^2 - q^2)} \tag{5.25}$$

$$\tan^2\phi = \frac{(u^2 - p^2)(p^2 - q'^2u^2)(1 - q^2)}{(1 - u^2)(1 - q'^2u^2)(p^2 - q^2)} \tag{5.26}$$

$$\tan^2\psi = \frac{(1 - q'^2u^2)(p^2 - q'^2u^2)(u^2 - q^2)}{(1 - u^2)(u^2 - p^2)(q'^2u^2 - q^2)} \tag{5.27}$$

where $u = \sigma'/\sigma$. To ensure that the right-hand side of these equations is never negative, the shape parameters must satisfy the constraints

$$\begin{aligned} &q \le q_k' \\ &q \le p \le 1 \\ &\max(q/q', p) \le u \le \min(p/q', 1) \end{aligned} \tag{5.28}$$

It is important to note that a single choice of q' must be made to generate the angles θ, ϕ, ψ that are used for the deprojection of all of the MGE components. A popular choice is to use the smallest axial ratio q', in which case the deprojection can be represented using the intrinsic q, p, and u of the MGE with the smallest q'.

One difficulty comes when sampling the shape parameters for a triaxial ellipsoid. Our optimisation code only allows for square bounds to be used on the parameters

sampled (see subsection 5.3.10). Because the shape parameters satisfy a series of inequality conditions, we internally use a reparameterisation in terms of the variables x, y, z given by

$$\begin{aligned} q &= q_{\min} + x(q_{\max} - q_{\min}) \\ p &= L_p + y(p_{\max} - L_p) \\ u &= L_u + z(U_u - L_u) \end{aligned} \tag{5.29}$$

where $L_p = \max[p_{\min}, q]$, $L_u = \max(q/q', p, u_{\min})$, and $U_u = \min(p/q', u_{\max})$. To ensure numerical stability, DYNAMITE does not allow $q, p, u > 0.999$ or < 0.01. For this reason, we set the minimum and maximum values for q, p, u to these. Additionally, we find that the orbit integration sometimes fails when the values of q, p, and u are too close to each other. This is a natural result of the conversion from q, p, u to the angles θ, ϕ, ψ, as this conversion depends on a number of ratios like one over $p^2 - q^2$ or $u^2 - p^2$ which are poorly behaved numerically as they approach zero. As a result, we implement a minimum distance between each parameter of 0.03.

5.3.8 Axisymmetric Deprojection

The oblate axisymmetric limit of the triaxial deprojection can be achieved by taking $p_k = 1$ and all ψ equal to $\pi/2$. In this limit the 3D density becomes

$$\rho(x, y, z) = \sum_{k=1}^{N} \frac{L_k}{(\sigma_k\sqrt{2\pi})^3 q_k} \exp\left[-\frac{1}{2\sigma_k^2}\left(x^2 + y^2 + \frac{z^2}{q_k^2}\right)\right] \tag{5.30}$$

From the previous section we see that in the oblate axisymmetric limit Equation 5.22 and Equation 5.23 reduce to the same result, that

$$q_k^2 = \frac{q_k'^2 - \cos^2\theta}{\sin^2\theta} \tag{5.31}$$

from which we identify θ with the inclination, i. Likewise, we see that $u = 1$ or $\sigma_k = \sigma_k'$. Note that the prolate case corresponds to replacing $q \to 1/q$ and $q' \to 1/q'$, so that we have

$$q_k = \frac{\sin^2\theta}{1/q_k'^2 - \cos^2\theta} \tag{5.32}$$

for prolate symmetry.

5.3.9 Parameter Assumptions

For JAM, we follow chapter 2 and use priors on the anisotropy parameters in order to guarantee physically reasonable shapes for the anisotropy. We summarise our choice

Table 5.1: Table of free parameters and their permitted upper and lower bounds for $\mathrm{JAM_{sph}}$. The limits on $(\sigma_r/\sigma_t)_0$ and $(\sigma_r/\sigma_t)_\infty$ come from the large literature of observations and simulations of core galaxies.

Parameter	Lower Bound	Upper Bound
$(\sigma_r/\sigma_t)_0$	0.5	1
$(\sigma_r/\sigma_t)_\infty$	1	1.3
M_{bh}	0	∞
M/L	0.5	1.5
f_{dm}	0	1

Table 5.2: Table of free parameters and their permitted upper and lower bounds for $\mathrm{JAM_{cyl}}$. The limit on (σ_r/σ_t) is motivated.

Parameter	Lower Bound	Upper Bound
(σ_r/σ_t)	1	1.3
M_{bh}	0	∞
M/L	0.5	1.5
f_{dm}	0	1

of bounds for $\mathrm{JAM_{cyl}}$ and $\mathrm{JAM_{sph}}$ in Table 5.2 and Table 5.1, respectively. Unlike in chapter 2, we fix the parameters α and r_a to equal 2 and the sphere of influence, respectively. This is to reduce the number of degeneracies and better reflect the amount of model assumptions one would make with data of typical quality. We run a test where we vary r_a to twice the sphere of influence and find it minimally impacts the recovered black hole masses. This is important as for real data knowledge of the sphere of influence is often imperfect. We choose the fit the parameters all on a linear scale. The reason for this is that we already have a good idea of the true values for each parameter. Additionally, we note that this range of parameters is guaranteed to satisfy the slope-anisotropy theorem as the central slope of this galaxy is slightly greater than zero but the maximum β in the center of the galaxy is 0.

For both JAM, the axisymmetric Schwarzschild code, and DYNAMITE, we have to carefully choose what to use as our dark matter distribution. As we saw in Figure 3.1, the true dark matter distribution in our simulation deviates from a NFW halo (Navarro et al., 1997) in both the inner and outer regions. The deviation in the outer region is due to the fact that the dark halo in the simulation is initialised as a Hernquist profile (Hernquist, 1990). In many cases with real galaxies, the standard choice is to assume a fixed NFW halo which is scaled according to some parameter like the dark matter fraction within one effective radius, f_{dm}. In our case, this assumption is not justified since we know that the dark matter in the simulation does not necessarily

reflect realistic galaxies. In order to ensure that our modelling is being constrained only by the systematics of the kinematics rather than the parameterisation of the dark halo, we choose as our halo the exact Nuker profile that we fit in section 3.3. We do, however, leave the overall amplitude of the halo as a free parameter that we fit. The parameter we use is the dark matter fraction within 1 effective radius $f_{\rm dm}$. The equation for this is

$$f_{\rm dm} = \frac{M_{\rm dm}(r < R_{\rm e})}{M_*(r < R_{\rm e}) + M_{\rm dm}(r < R_{\rm e})} \tag{5.33}$$

We calculate the effective radius for each projection using the MGE_HALF_LIGHT_RADIUS routine of the JamPy package. To solve for the radially enclosed mass, an analytic form has been derived in the axisymmetric case (Mitzkus et al., 2017) and implemented in the MGE_RADIAL_MASS routine of Jampy. For the triaxial case, we derived (for a single MGE component)

$$M_k(r) = \int_0^{2\pi}\int_0^{\pi}\int_0^{r} \frac{L_k r'^2}{(\sigma_k\sqrt{2\pi})^3 q_k p_k} \exp\left[-\frac{1}{2\sigma_k^2}\left(x^2 + \frac{y^2}{p_k^2} + \frac{z^2}{q_k^2}\right)\right] \sin\theta dr' d\theta d\phi \tag{5.34}$$

$$= \int_0^{2\pi}\int_0^{\pi}\int_0^{r} \frac{L_k r'^2}{(\sigma_k\sqrt{2\pi})^3 q_k p_k} \exp\left[-\frac{r'^2}{2\sigma_k^2}\left(\sin^2\theta\left(\cos^2\phi + \frac{\sin^2\phi}{p_k^2}\right) + \frac{\cos^2\theta}{q_k^2}\right)\right] \sin\theta dr' d\theta d\phi \tag{5.35}$$

Similar to the axisymmetric case (Cappellari et al., 2015), the θ integral can be evaluated to give

$$M_k(r) = \frac{L_k}{2\pi}\int_0^{2\pi}\int_0^{r} \frac{r'}{A} \exp\left[\frac{-r'^2 B}{2\sigma_k^2}\right] \mathrm{erf}\left[\frac{rA}{\sigma_k p_k q_k \sqrt{2}}\right] dr' d\phi \tag{5.36}$$

with

$$A = \sqrt{p_k^2(1 - q_k^2\cos^2\phi) - q_k^2\sin^2\phi} \tag{5.37}$$

$$B = \frac{p_k^2\cos^2\phi + \sin^2\phi}{p_k^2} \tag{5.38}$$

Note that for $p = 1$ (the axisymmetric limit) this recovers the result in (Cappellari et al., 2015, footnote 11). Similar to that case, we can analytically evaluate the r integral using integration by parts, giving

$$\begin{aligned} M_k(r) = \int_0^{2\pi} \frac{L_k^2}{2\pi AB}\Bigg[&\frac{A}{\sqrt{A^2 + Bp_k^2 q_k^2}}\mathrm{erf}\left(r\frac{\sqrt{A^2 + Bp_k^2 q_k^2}}{\sqrt{2}\sigma_k p_k q_k}\right) \\ &- \exp\left(-\frac{Br^2}{2\sigma_k^2}\right)\mathrm{erf}\left(\frac{Ar}{\sqrt{2}\sigma_k p_k q_k}\right)\Bigg] d\phi \end{aligned} \tag{5.39}$$

This leaves one final integral which we numerically evaluate.

An alternative approach, and one which allows for a more robust confirmation of the accuracy of our formula, is to compare to the circularised and axisymmetrised version of the triaxial MGE. For the circularised case this is determined by modifying σ_k and I_k by taking $q_k, p_k = 1$ such that the total luminosity and the peak luminosity for each MGE component are held constant. This is achieved with $\sigma_{l,\mathrm{circ}} = \sigma_k (p_k q_k)^{1/3}$ and $I_{k,\mathrm{circ}} = I_k \frac{\sigma_k^2 q_k'}{\sigma_{k,\mathrm{circ}}^2}$. For the axisymmetrised case this is achieved with $\sigma_{k,\mathrm{axi}} = \sigma_k p_k^{1/3}$ and $I_{k,\mathrm{axi}} = I_k \frac{\sigma_k^2}{\sigma_{k,\mathrm{axi}}^2}$. We plot the results of these different methods for a case of weakly triaxial geometry in Figure 5.3. There we clearly see that the axisymmetric approximation is better than the circular one, but that neither deviate significantly from the triaxial result.

It is worth noting that this treatment of the dark halo where we assume the shape is equal to the known one is impossible with real data. For real galaxies, it is more typical to assume a fixed NFW or logarithmic profile with the overall amplitude (typically parameterised with f_{dm}) free. Studies with a sufficiently wide field of view may also fit for the break radius r_{break}. In our case, we believe we are justified in using our form of the dark halo for two reasons. The first is that the dark halo in the simulation is initialised as a Hernquist profile (Hernquist, 1990). Notably, this has a steeper outer profile than NFW so we expect a naive implementation of the NFW halo to result in systematics that aren't present in real galaxies. Second, as we discuss more in subsection 5.6.1 and subsection 5.6.2, our simulated galaxy has an unrealistically large dark halo compared to real galaxies. For this reason, we do not want our results to rely too strongly on the dark halo, so for these reasons we proceed using the fit to the true halo (up to the scaling f_{dm} which we fit).

DYNAMITE natively supports several dark matter halo options, but none are suitable for our double power-law profile. To overcome this, we write our own DYNAMITE wrapper which includes our dark matter halo parameterisation by adding it to the MGE potential. We generate our MGE of the profile using the MGE_FIT_1D routine of the MGEFIT package (Cappellari, 2002).

5.3.10 Parallel Optimisation

Running a single Schwarzschild model takes on the order of several hours to one day to run. Determining the best-fit model parameters requires iteratively running many different models. Historically, the approach to this problem has often been to run parameter grids or parameter grid searches to find the best-fit model. This approach

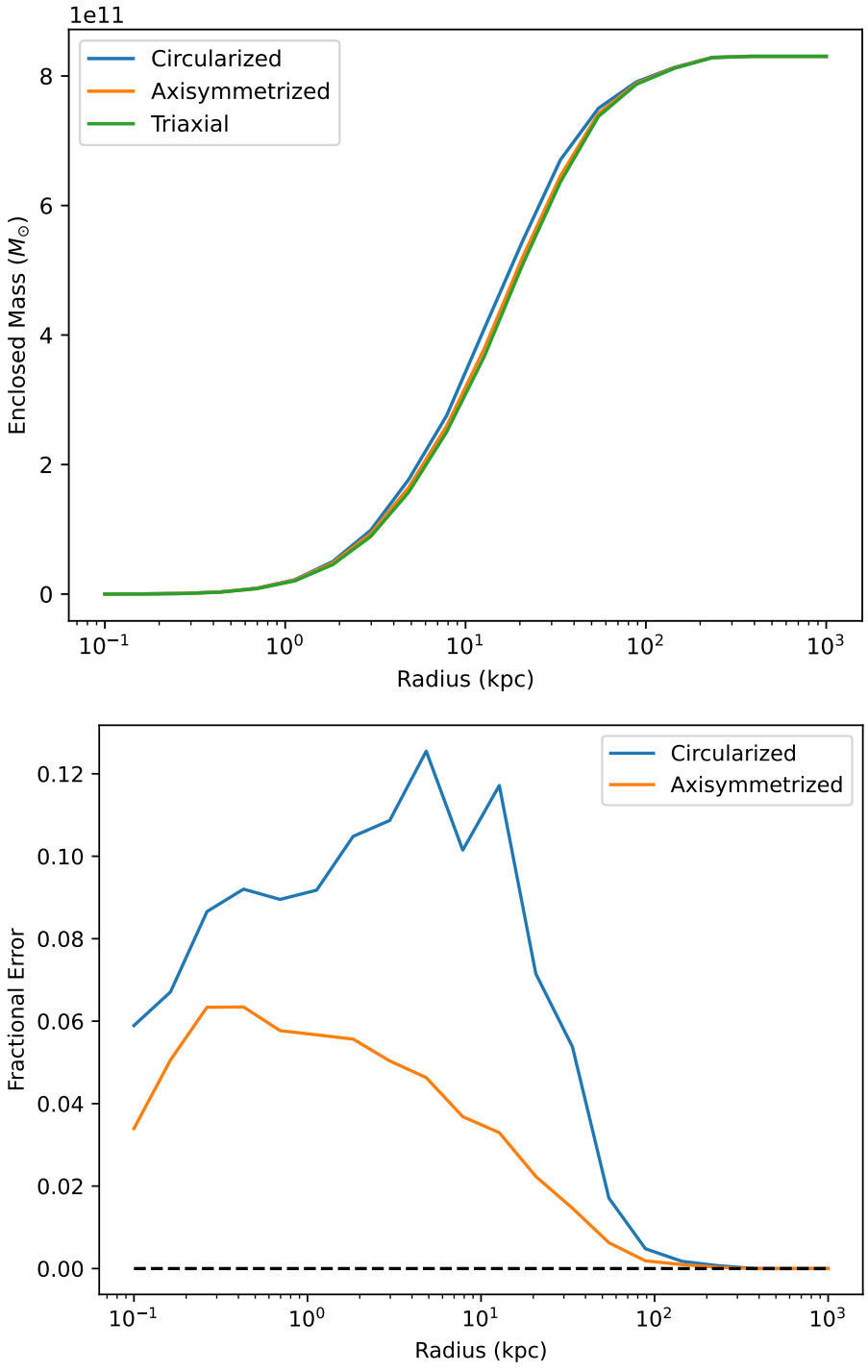

Figure 5.3: The top panel shows the enclosed mass using the circuliarsed, axisymmetrised, and triaxial MGE. The bottom panel shows the fractional error in each of the methods compared to the triaxial formula. We find a close agreement with each method, with the axisymmetric case understandably outperforming the circularised case.

usually requires thousands of models to be run, with the numbers of models required increasing exponentially with the number of parameters included.

We circumvent this issue by using Bayesian optimisation (Jones et al., 1998). Bayesian optimisation is designed for black box functions that are slow to compute, have no derivatives available, and have noisy outputs. This exactly describes the sort of problem faced in Schwarzschild modelling. The full description of this method will be given in Haitong Wang et al. (in prep). We provide a brief outline of the method here.

We use the Bayesian optimisation implementation in the Python package SCIKIT-OPTIMIZE (Head et al., 2021). This method works by iteratively fitting a Gaussian process to some points sampled in the χ^2, suggesting the next optimal point to sample, and repeating. In this form the optimisation proceeds in series. It can be parallelised by assuming a value of the χ^2 at the proposed point and then asking what the next best point to sample is. There are several options for the choice of assumed χ^2 values. Two common choices are assuming the χ^2 is equal to the max of the values seen so far or the mean value seen so far. These approaches are well suited for particularly pathological χ^2 functions with many multiple minima, as this encourages the method to sample a diverse range of points. Another approach, if the χ^2 is already well sampled and relatively well behaved, is to assume that the value of the χ^2 can be interpolated from the Gaussian process fit. We have compared these different approaches and found that the mean/max method tends to converge slowly, especially as the number of dimensions increases. As a result, we proceed with the Gaussian process fit approach.

For our Gaussian process (Jones et al., 1998), we use a Matérn kernel with $\nu = 2.5$. In our tests we found that it was occasionally challenging for the GP to accurately model the area around the minimum of the χ^2 as typical confidence intervals are the result of a $\Delta\chi^2$ of order 10, whereas our χ^2 spans three orders of magnitude. To improve the fits, and thus the convergence of the method, we fit the Gaussian process to the log of the χ^2. We practically find that this significantly improves the convergence of the method.

The parallel optimisation code was implemented in the axisymmetric code by Haitong Wang. I updated the DYNAMITE code to use OpenMPI to facilitate running multiple models in parallel. This code was provided by generative AI and tested to ensure that it behaved as expected.

5.4 JAM Modelling Results

5.4.1 Oblate and Prolate Symmetry

JAM allows for both oblate and prolate axisymmetric galaxies to be modelled. We first perform a systematic test of models with oblate/prolate symmetry across both choices of JAM modelling and assuming a mass-follows-light gravitational potential and a model with a dark halo. We show the results for $\mathrm{JAM_{cyl}}$ in Figure 5.4 and for $\mathrm{JAM_{sph}}$ in Figure 5.5. As our data we use the true second moments calculated for individual particles in each aperture. We only include data within 3 kpc. The way we plot the data is that we take each point in the octant of a sphere that we sample (where each point corresponds to the projection of the galaxy such that the line from the origin to the point corresponds to the intrinsic major axis of the galaxy).

Several features immediately jump out of these results. We see that for all cases, galaxies that are aligned such that the intrinsic major axis is aligned with the line-of-sight (z axis) have the worst recovery of the supermassive black hole mass. For $\mathrm{JAM_{cyl}}$ this results in a systematic underestimate of the black hole mass by around a factor of two, while for $\mathrm{JAM_{sph}}$ this can result in either an over-estimation or an under-estimation of the black hole mass. This makes sense as there is the largest uncertainty in the deprojection along the long axis of the galaxy.

One surprising result is that $\mathrm{JAM_{sph}}$ consistently tends to miss the true black hole mass by around 20 per-cent. The reason for this likely has to do with the mass-anisotropy degeneracy combined with the fact that $\mathrm{JAM_{sph}}$ makes assumptions that aren't perfectly matched by the data. Most of the best-fit $\mathrm{JAM_{sph}}$ models have anisotropy parameters that are up against the prior boundaries, typically with $(\sigma_r/\sigma_t)_0 = 1$ and $(\sigma_r/\sigma_t)_\infty = 1.3$. This higher radial anisotropy in the central region will result in a smaller black hole mass. This is especially clear when looking at Figure 5.13, where we clearly see that fits with the true black hole mass also provide a good fit to the data. The exception to this is the case with oblate symmetry and mass-follows-light potential, in which case we typically have $(\sigma_r/\sigma_t)_0 = 0.8$ and $(\sigma_r/\sigma_t)_\infty = 1$. This case emphasises the importance of the mass anisotropy degeneracy in dynamical modelling.

We observe several other features. We consistently see that, especially for $\mathrm{JAM_{cyl}}$, prolate symmetry tends to do a better job of recovering the black hole mass than oblate symmetry. In fact, the results for the recovery of the black hole mass tend to appear symmetric around the z axis (the long axis). Indeed, as we know from section 3.3, the galaxy, while triaxial, is close to prolate. If it is the case that most

slow rotator galaxies are elongated in this way, then it may be the case that they should be treated as approximately prolate when modelled assuming axisymmetry.

Lastly, we see that the addition of dark matter can significantly change the best-fit black hole mass. For $\mathrm{JAM_{cyl}}$ this improves the range of projections where a good fit is obtained, while for $\mathrm{JAM_{sph}}$ the impact is less clear.

It is important to note that the best-fit black hole mass on its own is insufficient to judge the quality of the recovery as model maps that fail to accurately reproduce the data are not believable even if they do recover the true black hole mass. For black hole studies this means the central region of the kinematics must provide an accurate visual fit to the data. We plot our best-fit maps for a selection of projections that uniformly sample the angle between the line-of-sight and the intrinsic long axis of the galaxy (θ_{long}). We plot the results for this in Figure 5.6. There we see very clearly that the quality of the fits along the long axis of the galaxy is very poor, improving significantly after $\theta_{\mathrm{long}} = 30$. We also clearly see, especially in the case of $\mathrm{JAM_{cyl}}$, that the model is best-fit under the assumption of prolate symmetry rather than oblate symmetry.

Taking into consideration only the best-performing models ($\mathrm{JAM_{cyl}}$ and $\mathrm{JAM_{sph}}$ each with prolate symmetry) and only including the maps with an acceptable fit ($\theta_{\mathrm{long}} \geq 30$), we find a mean black hole mass recovery across all reliable projections of $\mathrm{JAM_{cyl}}$ of $M_{\mathrm{bh,JAM}}/M_{\mathrm{bh,true}} = 1.04^{+0.10}_{-0.32}$ and for $\mathrm{JAM_{sph}}$ of $M_{\mathrm{bh,JAM}}/M_{\mathrm{bh,true}} = 0.80^{+0.21}_{-0.06}$.

5.4.2 Optimal Field of View

In the previous section we cropped our data to a field of view of 3 kpc (three times the sphere of influence). Barring limited data availability, the field of view is a free parameter which can be arbitrarily chosen. We expect that the accuracy with which we can recover the black hole mass will depend on the field of view of our mock observations as this can interpolate between radii where the assumptions underlying JAM modelling (shape of the velocity ellipsoid, axisymmetry, etc.) are more or less accurate.

To investigate this question, we vary the field of view of our mock data and measure the best-fit JAM models using the two different parameterisations described in subsection 5.3.2. We show the results for the recovery of the black hole mass across different model assumptions in Figure 5.7 and Figure 5.8. For our data we use the true second moment calculated using individual particles.

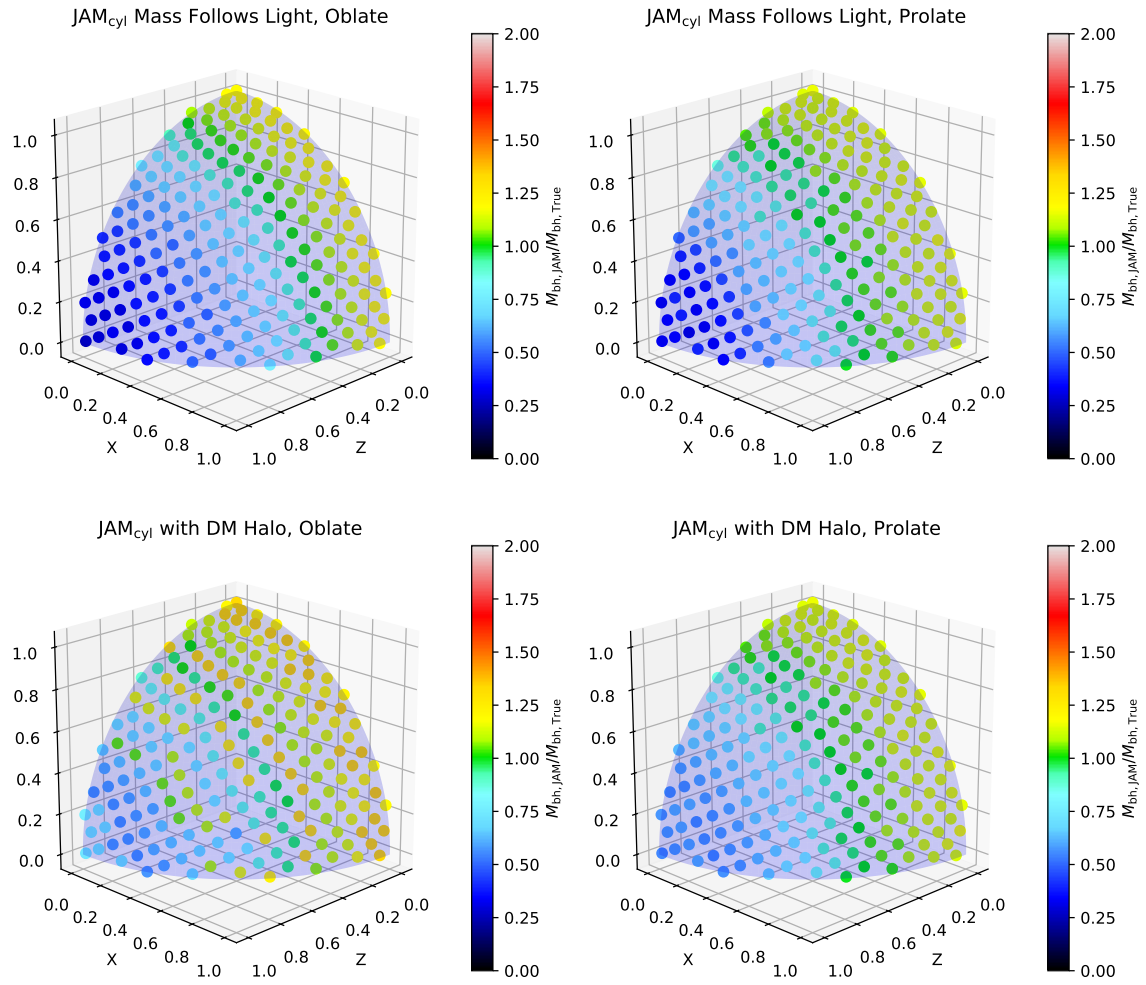

Figure 5.4: Three dimensional plot showing the recovered supermassive black hole mass for different JAM$_{cyl}$ model assumptions. The meaning of the three dimensional plot is that each point on the surface of the sphere corresponds to the projection of the galaxy such that the major axis is oriented in that direction where the z axis is the line-of-sight. We see that the recovery of the black hole mass is close to 1 for intermediate projections but is very low when the long axis is along the line-of-sight.

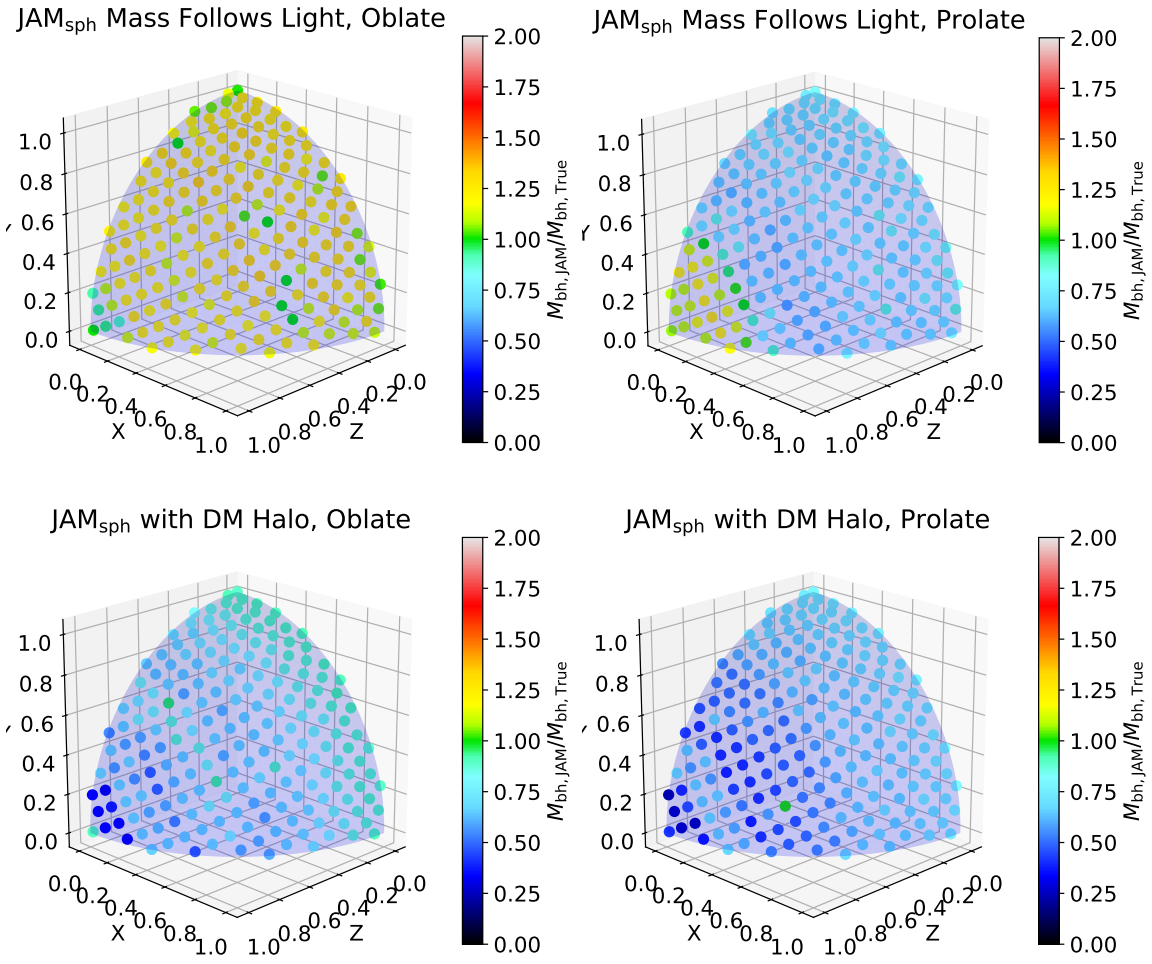

Figure 5.5: Three dimensional plot showing the recovered supermassive black hole mass for different JAM$_{sph}$ model assumptions. The meaning of the three dimensional plot is that each point on the surface of the sphere corresponds to the projection of the galaxy such that the major axis is oriented in that direction where the z axis is the line-of-sight. We see that the recovery of the black hole mass is close to 0.8 for most cases, but is constantly elevated for oblate symmetry with a mass-follows-light potential.

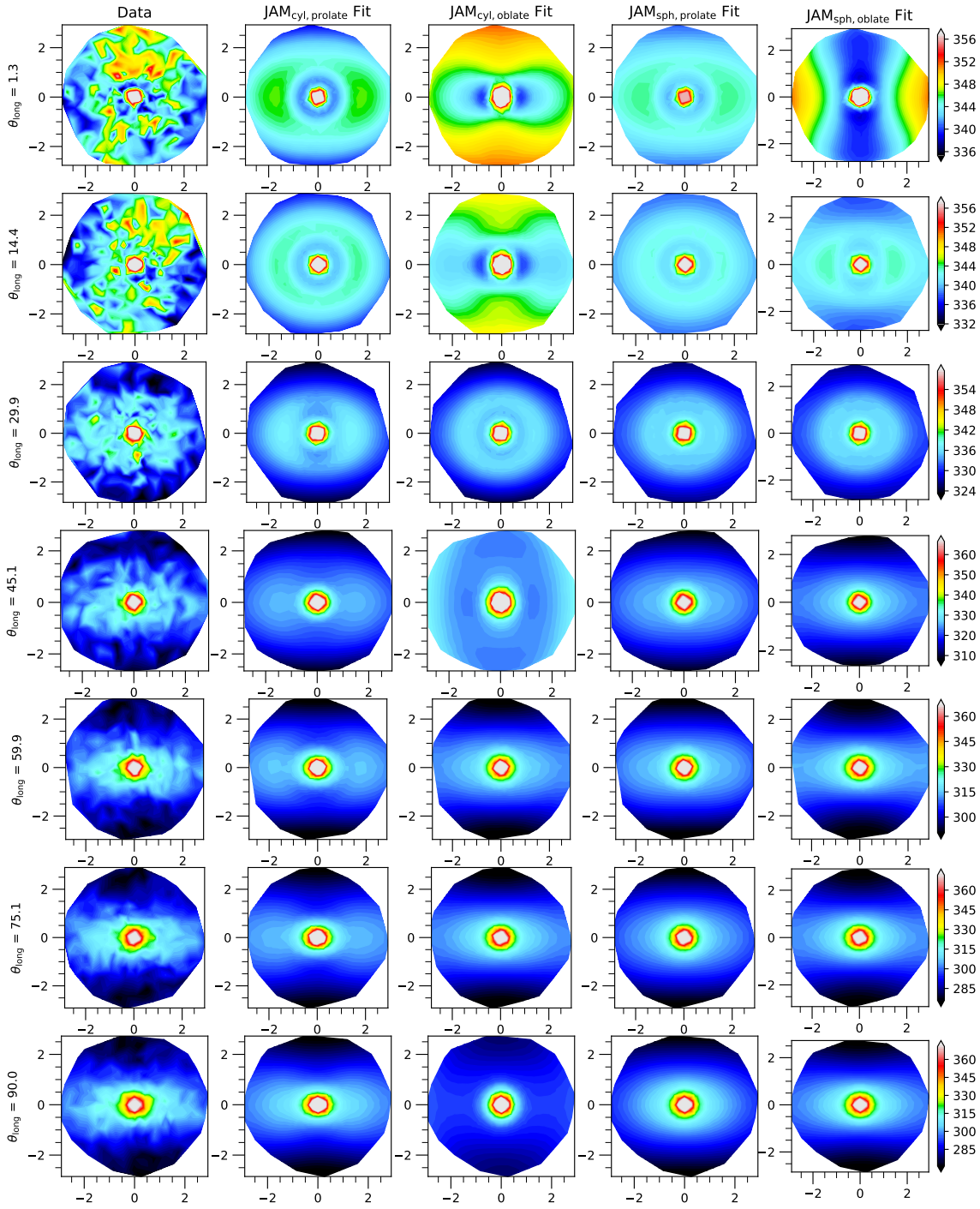

Figure 5.6: Performance of JAM models across different modelling assumptions and θ_{long} where θ_{long} is the angle between the line-of-sight and the intrinsic long axis of the galaxy. We see that the worst model recoveries all occur when projecting along the long axis of the galaxy. We also clearly see that JAM$_{\mathrm{cyl}}$ performs significantly better assuming prolate symmetry than oblate symmetry.

Several features immediately stand out. First, we see that for $\mathrm{JAM_{cyl}}$, the field of view that does the best job of recovering the black hole mass tends to be between 2 and 3 times the sphere of influence depending on whether prolate or oblate symmetry is assumed. The overall trends of the recovery as a function of the radius vary greatly depending on the assumptions made. $\mathrm{JAM_{cyl}}$ with a halo and prolate symmetry has a best-fit black hole that increases with the field of view, while assuming a mass-follows-light model with prolate symmetry gives a black hole that initially increases but later decreases.

For $\mathrm{JAM_{sph}}$, the situation is somewhat more complicated. The model assumption that does the best job of recovering the black hole mass is the oblate mass-follows-light model. For the remaining models, the best-fit black hole tends to hover around 0.8 times the true value, before changing significantly when a large field of view is encompassed.

To better understand the results, it is useful not just to see the recovered black hole mass, but the values of the other parameters as well. We show the recovery for all parameters for prolate $\mathrm{JAM_{cyl}}$ models with dark matter and prolate $\mathrm{JAM_{sph}}$ models with dark matter in Figure 5.9 and Figure 5.10, respectively. Here we see that there is a trade-off between M/L and f_{dm} depending on the field of view. As we might expect, increasing the field of view improves the accuracy of the recovered f_{dm}. For $\mathrm{JAM_{cyl}}$, the anisotropy is always close to isotropic, so much so that it could be set as a fixed parameter without significantly impacting the results. For $\mathrm{JAM_{sph}}$, the anisotropy is almost always up against the prior boundaries. As a result, both M_{bh} and M/L are somewhat underestimated.

These results confirm the intuition that JAM can accurately model black holes locally, but struggle to fit every feature when trying to constrain both wide scale structures as well as the central region.

5.4.3 PSF Uncertainty

So far in this work we have used the simulation data without applying any point spread function (PSF). PSFs are challenging to measure, especially for integral field studies of galaxies where there are often no stars in the field of view. To better understand what impact uncertainty in the PSF has on recovering supermassive black hole masses, we perform an experiment where we convolve a mock observation of the y projection with a PSF with $\sigma_{\mathrm{psf}}^{\mathrm{truth}} = 0.5, 1, 2, 3$ kpc. We then perform JAM best-fits using $\mathrm{JAM_{cyl}}$ with constant anisotropy, but varying the JAM PSF from 0 to $2\sigma_{\mathrm{psf}}^{\mathrm{truth}}$ to represent some uncertainty in the PSF modelling. We plot the results of this in Figure 5.11.

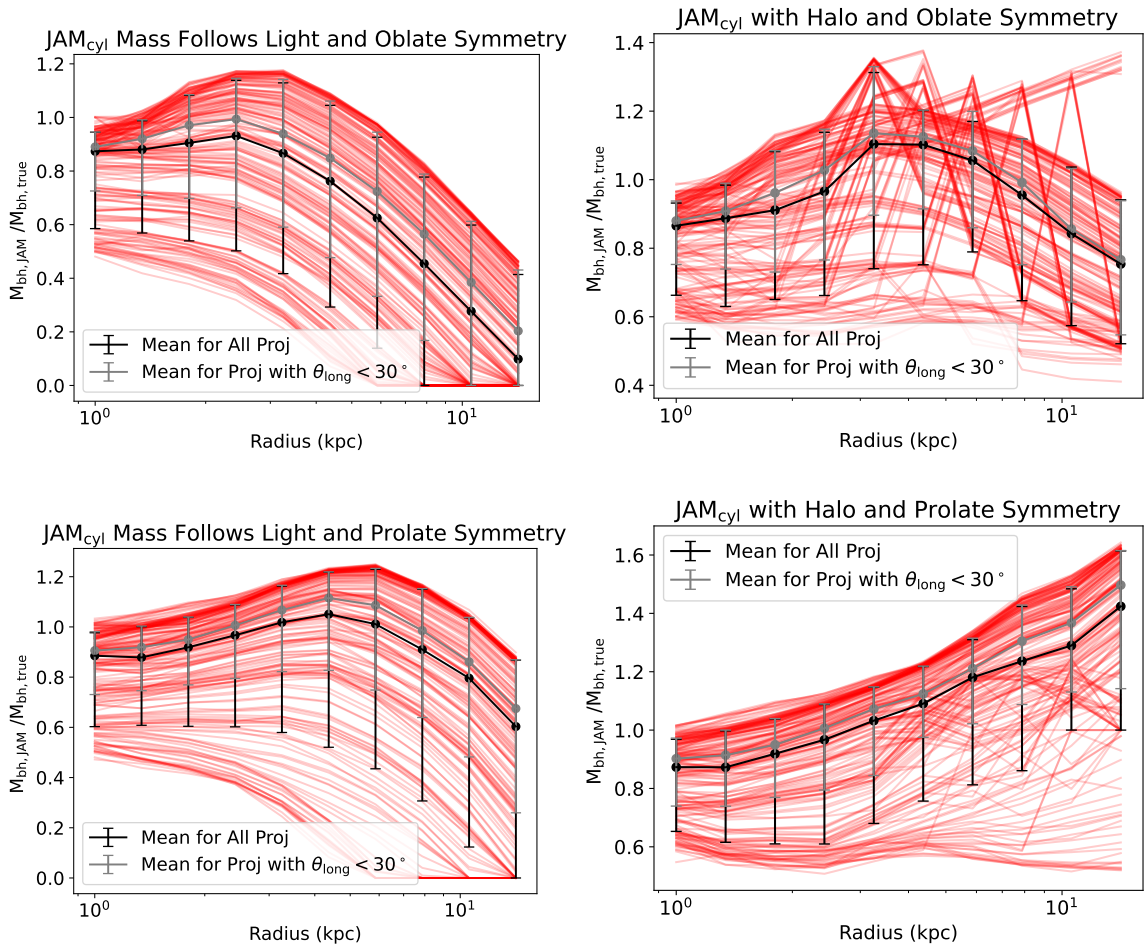

Figure 5.7: Accuracy of the recovery of the black hole mass for $\mathrm{JAM_{cyl}}$ as a function of the field of view. We show the average of all projections binned for each choice of field of view in black, and the same but excluding projections within 30 degrees of the long axis in gray. We clearly see that the best-fit supermassive black hole mass is obtained when a field of view of around 3 times the sphere of influence is assumed.

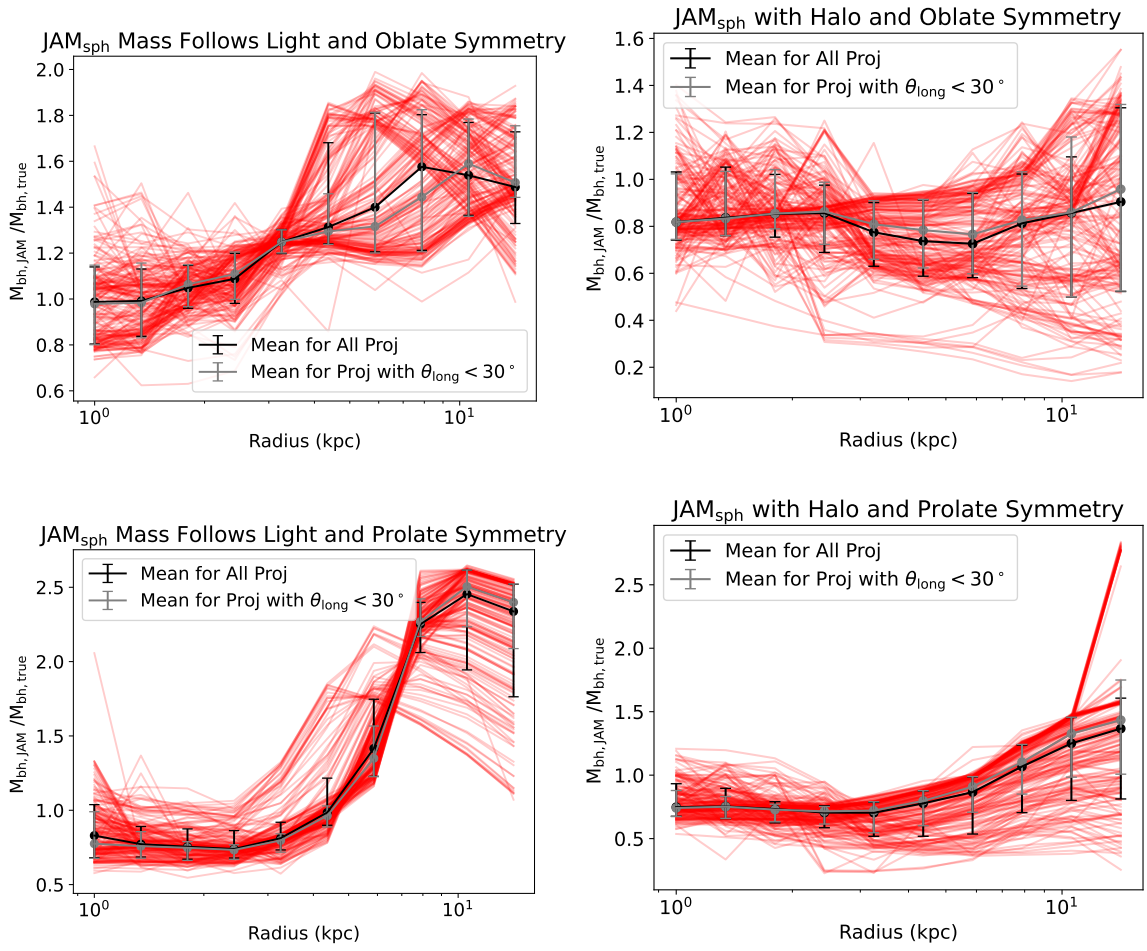

Figure 5.8: Accuracy of the recovery of the black hole mass for JAM$_{sph}$ as a function of the field of view. We show the average of all projections binned for each choice of field of view in black, and the same but excluding projections within 30 degrees of the long axis in gray. We see that the best-fit black hole masses tend to be achieved with a wider field of view for the prolate case. For oblate symmetry there is no clear trend, with the mass-follows-light case being close to the truth with small fields of view and the case with a halo always being close to 0.8.

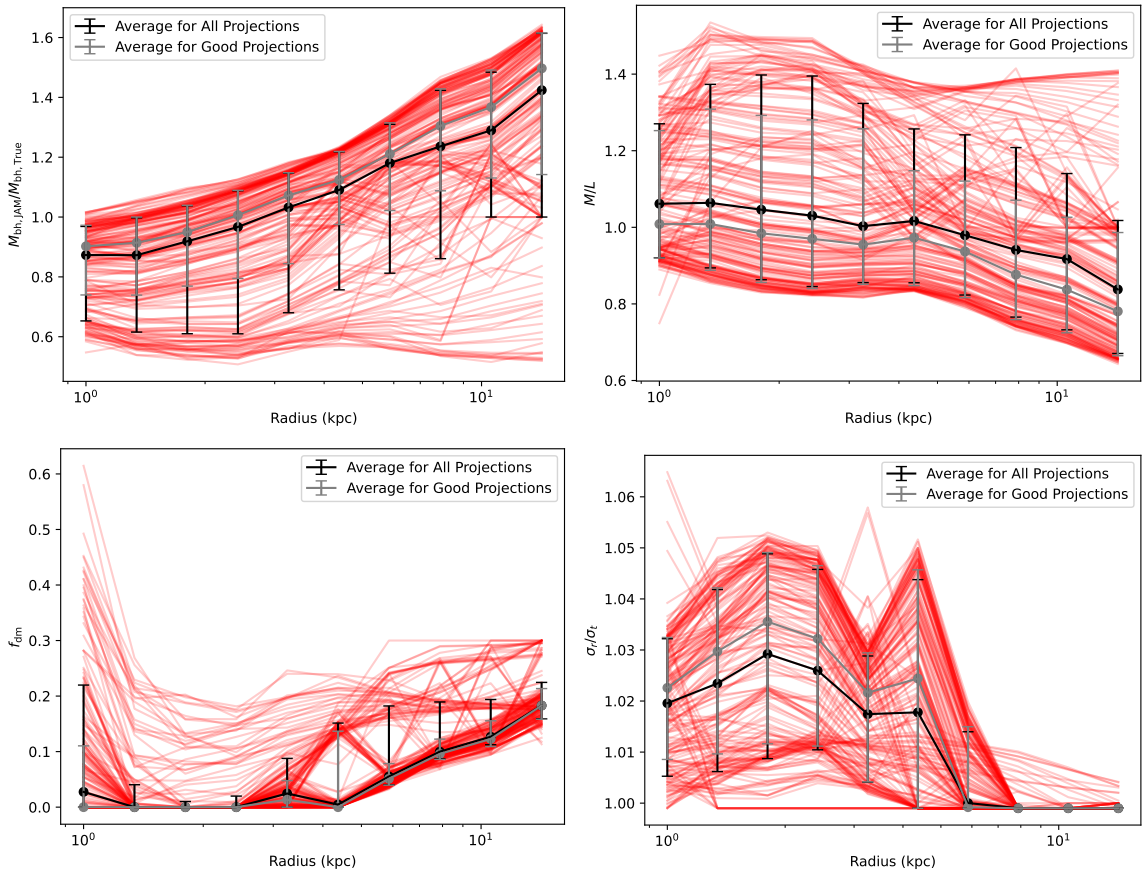

Figure 5.9: Best-fit $\mathrm{JAM_{cyl}}$ models with a dark halo and prolate symmetry as a function of radius for all projections. The top left panel shows the recovery of the black hole mass, top right panel shows the recovery of the M/L ratio, and the bottom left panel shows the recovery of the dark matter fraction all as a function of the field of view, and the bottom right panel shows the value of the constant anisotropy σ_r/σ_t. The black points are the median across all projections and the gray points are the median across all projections where the intrinsic major axis is further than 30 degrees away from the line-of-sight. We see that all models are close to isotropic. Additionally, we observe a degeneracy between the stellar mass and dark matter as the stellar M/L decreases with increased radius while the dark matter fraction increases.

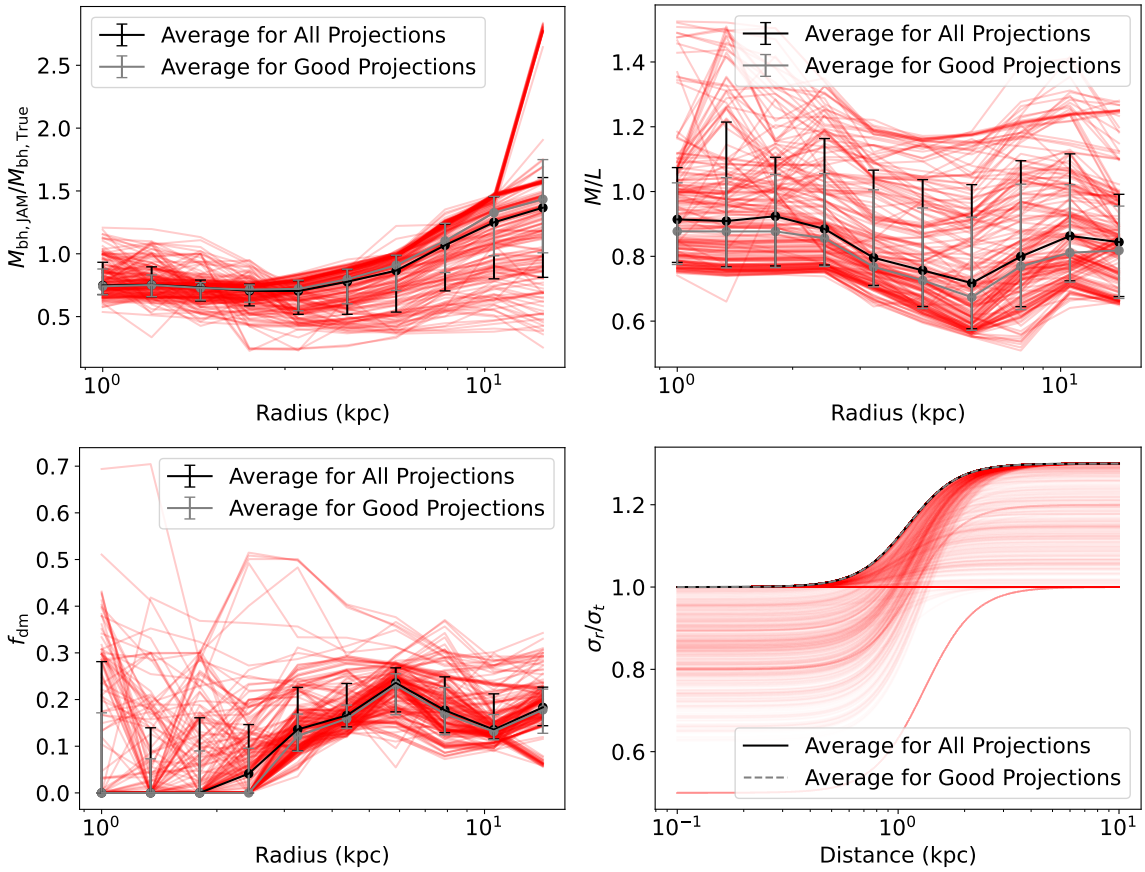

Figure 5.10: Best-fit JAM_{sph} models with a dark halo and prolate symmetry as a function of radius for all projections. The top left panel shows the recovery of the black hole mass, top right panel shows the recovery of the M/L ratio, and the bottom left panel shows the recovery of the dark matter fraction, all as a function of the field of view. The bottom right panel shows all of the anisotropy profiles for each projection and choice of field of view. Note that the x axis here is different from the other figures as this is not the enclosed field of view but the shape of the anisotropy as a function of the position in the galaxy. The black points/curves are the median across all projections and the gray points/curves are the median across all projections where the intrinsic major axis is further than 30 degrees away from the line-of-sight. We see that all models tend to have anisotropy profiles that push against the prior boundaries. Additionally, we observe a degeneracy between the stellar mass and dark matter as the stellar M/L decreases with increased radius while the dark matter fraction increases.

There we see that the sensitivity of the black hole mass to the choice of PSF increases as the true size of the PSF increases. For well-resolved observations (PSF sigma of half the black hole sphere of influence) a ten per-cent deviation in the PSF roughly corresponds to a ten per-cent deviation in the black hole mass. However, in the unresolved case (PSF sigma equal to twice the black hole sphere of influence) a ten per-cent deviation in the PSF corresponds to a thirty per-cent deviation in the black hole mass. Once the PSF becomes sufficiently large ($\sigma_{\rm psf}^{\rm truth} = 3$), the model fails to accurately recover the true black hole mass even when the PSF is well-modelled. This illustrates the importance of accurately determining the PSF for black hole studies and the strong relationship between uncertainties in measured black holes and the resolution of the data.

It is worth emphasising, however, that this test is somewhat simplified as it does not account for the wavelength dependence of the PSF. In fact, no modern dynamical modelling method takes this factor into account. The reason for it is that dynamical models are designed around fitting a set of kinematics. Modelling a wavelength dependent PSF requires predicting an IFS datacube, and thus is significantly more difficult to do. For instruments with small wavelength ranges this likely will not have a significant effect. However, for larger wavelength ranges, it is possible to imagine that this can have a significant impact. For now, this is outside the scope of this study, though it is worth a study of its own.

5.4.4 Markov Chain Monte Carlo

In order to better understand the uncertainties and degeneracies in our JAM models we use the code EMCEE (Foreman-Mackey et al., 2013) to perform an MCMC analysis of our JAM models. For the errors, we take the uncertainties in V and σ to be 5 km s^{-1} and 5 per-cent of σ, respectively. We show the results for a mass-follows-light model in Figure 5.12 and the results for models with dark matter in Figure 5.13. We choose as our projection the galaxy aligned such that the intermediate axis is along the line-of-sight. We see that our use of priors is quite important as many of the models run up against the boundaries. Despite this, the recovery of the black hole mass is always reliable. One important feature is that the JAM$_{\rm sph}$ with logistic anisotropy results show a fairly significant degeneracy. This is to be expected as slow rotators tend to be triaxial but close to spherical symmetry with observed ellipticity $\epsilon < 0.4$ and as a result are more susceptible to the mass-anisotropy degeneracy (Binney & Mamon, 1982). This degeneracy is also a result of the limited field of view. We show the results comparing two fields of view, one with 3 kpc and one with 10 kpc

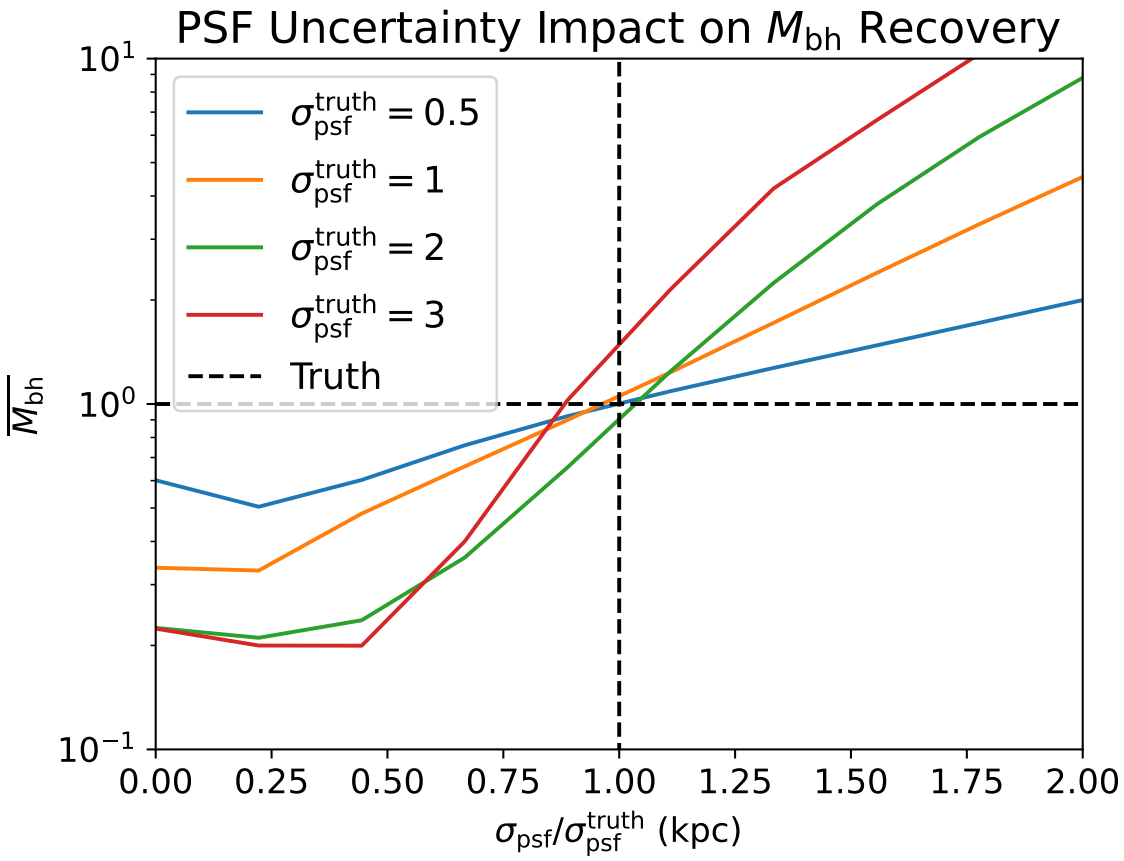

Figure 5.11: Best-fit JAM black hole masses (normalised to one) assuming cylindrical alignment of the velocity ellipsoid, constant anisotropy, and mass-follows-light gravitational potential as a function of the JAM PSF relative to the true PSF. The recovered black hole masses depend weakly on the PSF when the sphere of influence is well-resolved, but depend heavily on it when poorly resolved. In the highly unresolved case with $\sigma_{\rm psf}^{\rm truth} = 3$ using the true PSF value returns a black hole mass that substantially deviates from the truth.

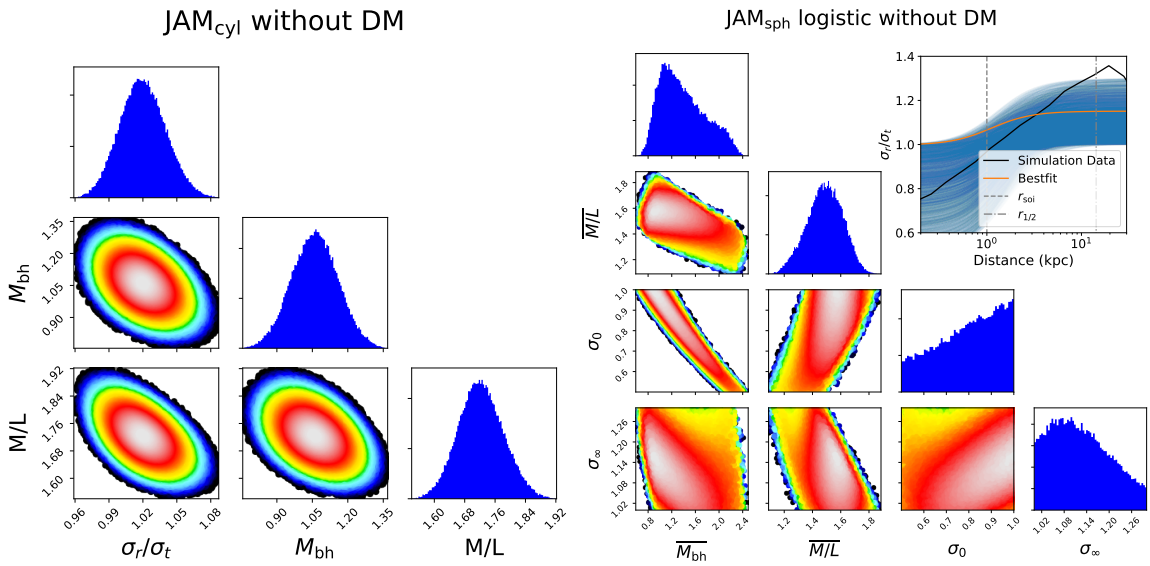

Figure 5.12: MCMC chains for the two different JAM models using a mass-follows-light model with data out to 3 kpc. In the right model with varying anisotropy we show a random sampling of the anisotropy from the posterior in the top right corner.

in Figure 5.14. We see that, while there does still remain some degeneracy between the parameters, going to a larger field of view pins down the dark matter fraction, removing a significant range of parameter space.

5.5 Schwarzschild Modelling Results

5.5.1 Axisymmetric and Triaxial Black Hole Masses

For all of our Schwarzschild models our goal is to determine the best-fit black hole masses under the most generous assumptions in order to determine the systematics of the method. With this in mind, for our kinematic data we use GH moments determined via the histogram fitting method in section 3.5 up to h_{12}.

We show a projection of our sampled black hole masses and their $\lg \chi^2$ values in Figure 5.15. In each panel we show the values for DYNAMITE on the left and for the axisymmetric code on the right. The projection number corresponds to the numbers in Figure 5.1. For these figures we only include the kinematic component of the χ^2. Practically, we find that including other terms in the χ^2, such as the fit to the tracer, results in the black hole mass being recovered more poorly, in some cases being off by up to 60 per-cent. Looking at Figure 5.15, several interesting features immediately stand out. First, we see that all models produce what appears to be a

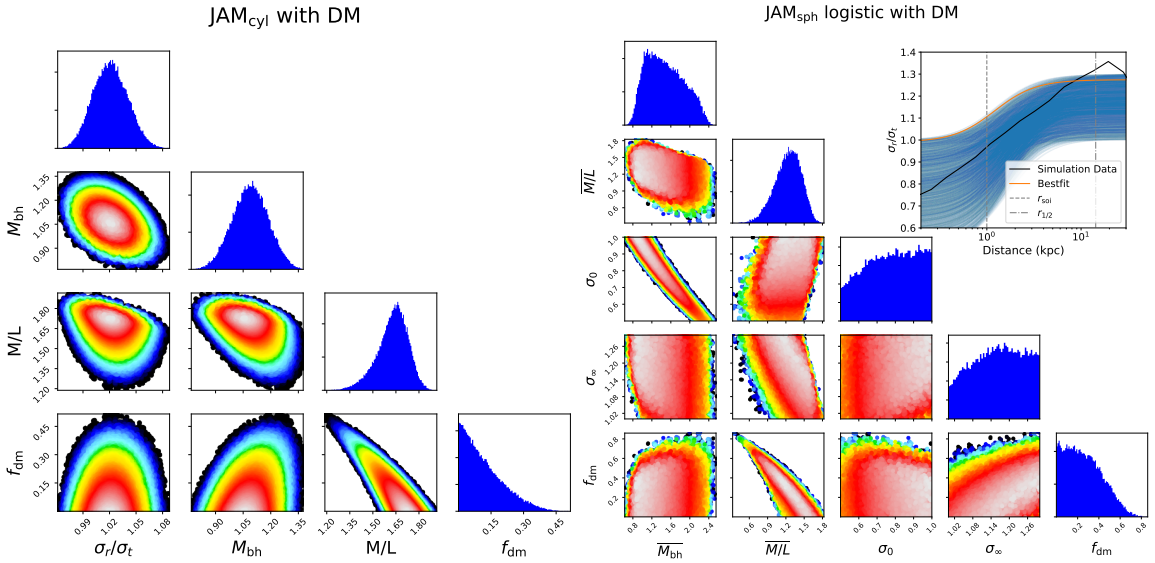

Figure 5.13: MCMC chains for the two different JAM models including dark matter with data out to 3 kpc. In the right model with varying anisotropy we show a random sampling of the anisotropy from the posterior in the top right corner.

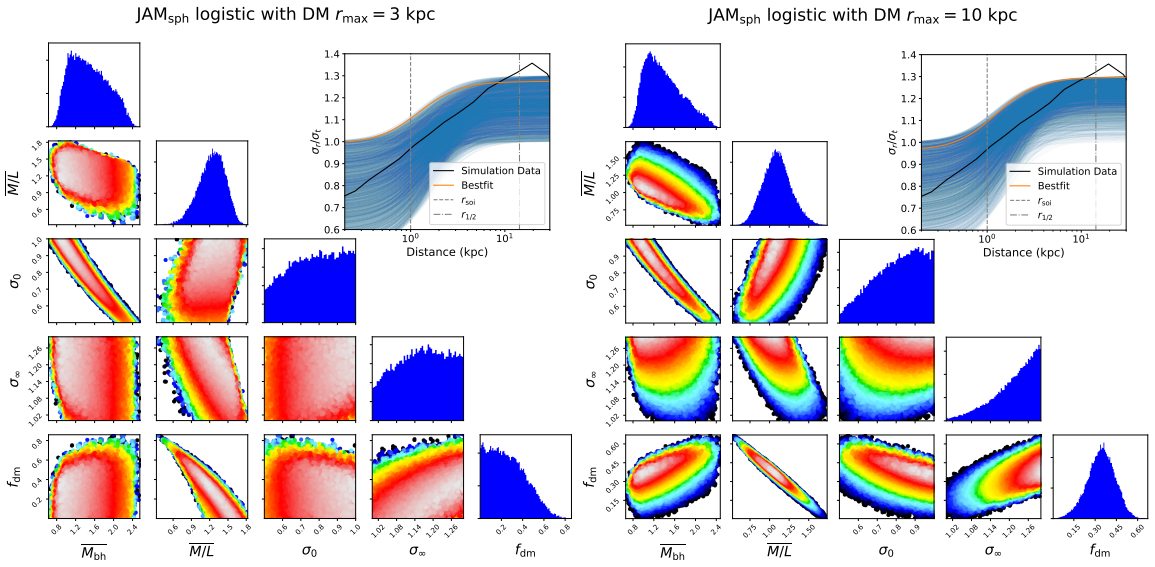

Figure 5.14: MCMC chains for the JAM$_{sph}$ logistic model with dark matter. The left hand panel includes data out to 3 kpc while the right panel includes data out to 10 kpc. We show a random sampling of the anisotropy from the posterior in the top right corner. We see that the inclusion of the extra spatial data allows us to do a much better job breaking the mass-anisotropy degeneracy.

clear quadratic-shaped contour in the χ^2 plot. Additionally, there is generally a good agreement between the axisymmetric and triaxial models, with most models agreeing to within 10 per-cent.

To understand the relationship between the best-fit black hole mass and the projection, we plot the recovered black hole masses with different methods as a function of the angle between the intrinsic long axis and the line-of-sight ($\theta_{\rm long}$) in Figure 5.16. We immediately see several key features. The first is that both codes present an even scatter of points around the true black hole mass, with both appearing to do a good job of recovering the black hole within 20 per-cent. The largest outliers appear on the right-hand side of the plot, corresponding to the projection along the intrinsic long axis of the galaxy. Similar to JAM, the transition from accurate recovery to inaccurate recovery appears to occur close to $\theta_{\rm long} = 30$ degrees. The case where the projection is aligned with the long axis corresponds to the face on case for axisymmetric galaxies since this triaxial system is close to prolate.

The best-fit black hole masses on their own are not sufficient as we also require the quality of the fit to the data to be high. As such, we plot some of the best-fit maps from our axisymmetric Schwarzschild models in Figure 5.17 and for our triaxial Schwarzschild models in Figure 5.18. Note that in the code we are fitting up to h_{12} but here we only show up to h_6. Several features immediately stand out. First, we see that our axisymmetric Schwarzschild models do a good job fitting even terms in the LOSVD (σ, h_4, h_6, etc.), but, in some cases, fail to replicate some of the odd features, especially the velocity. This is due to the fact that the triaxial shape of the galaxy results in significant twists and non-axisymmetric features in the velocity maps which the model is not able to recover. In the triaxial case, we see that these features are reproduced well.

Similar to the case with JAM, we see significant qualitative changes in the maps as they approach the intrinsic long axis. Unlike JAM, however, the quality of the fits here appears on the surface to be acceptable. It is only by knowing the true values of the black hole mass that we know that these projections do not accurately recover the mass.

Across all projections, the accuracy of the recovery for axisymmetric Schwarzschild is $M_{\rm bh}/M_{\rm bh,true} = 1.09^{+0.13}_{-0.08}$ and for Triaxial Schwarzschild is $M_{\rm bh}/M_{\rm bh,true} = 1.08^{+0.06}_{-0.24}$. If we exclude the projections within $\theta_{\rm long}$ of 30 degrees then this changes to $M_{\rm bh}/M_{\rm bh,true} = 1.09^{+0.08}_{-0.13}$ for axisymmetric Schwarzschild and $M_{\rm bh}/M_{\rm bh,true} = 1.09^{+0.06}_{-0.06}$ for Triaxial Schwarzschild. We thus see that the accuracy of the recovery of the black hole mass is similar for both triaxial and axisymmetric Schwarzschild modelling, though the

spread is slightly smaller in the triaxial case. Interestingly, both methods tend to overestimate the black hole mass by 9 per-cent.

5.6 Discussion

5.6.1 Representativeness of the Simulation

In order to justify extrapolating the results of our tests to real galaxies, we have to carefully determine how representative our galaxy is of the slow rotator population. We investigate four separate parameters, the projected ellipticity of the galaxy ϵ, the specific angular momentum λ_r, the photometric-kinematic misalignment angle Ψ, and the dark matter fraction within one effective radius $f_{\rm dm}$.

The criterion for slow rotators, as given in Cappellari (2016), is that $\lambda_r < 0.08+\epsilon/4$ and $\epsilon < 0.4$. We calculate ϵ using the FIT_GALAXY routine of MGEFIT, and we calculate λ_r from the definition

$$\lambda_r = \frac{\sum_{i=1}^{N} F_i R_i |V_i|}{\sum_{i=1}^{N} F_i R_i \sqrt{V_i^2 + \sigma_i^2}} \tag{5.40}$$

where we take F_i equal to the number of particles in each Voronoi bin, R_i the radius at each bin, and V_i, σ_i the mean and standard deviation of the LOSVD in each bin calculated using individual particles (see also Equation 1.5). We show the histograms for this in Figure 5.19. We see that λ_r is always very small, with the maximum value attained being 0.017. Additionally, the distribution of ϵ is always below 0.3, guaranteeing that all projections of our galaxy satisfy the standard criterion for slow rotators.

We define the misalignment angle Ψ as the difference between the photometric and kinematic major axes. We calculate the photometric major axis using the FIND_GALAXY routine of MGEFIT and the kinematic major axis using PAFIT[6] (Krajnović et al., 2006, appendix C) and plot the results in Figure 5.19. We see that, typical for slow rotators, there is a great deal of misalignment in this galaxy, with Ψ uniformly populating the range of angles.

We calculate the dark matter fraction directly from the simulation using individual particles. We see that the dark matter fraction is very large, ranging between 0.3 and 0.4. Statistics on the dark matter fraction distribution for slow rotator galaxies have previously been studied in MaNGA (Cappellari, 2026, Fig. 11). In this work, the maximum dark matter fraction observed was 0.3, with the vast majority closer to

[6]Available at `https://pypi.org/project/pafit/`

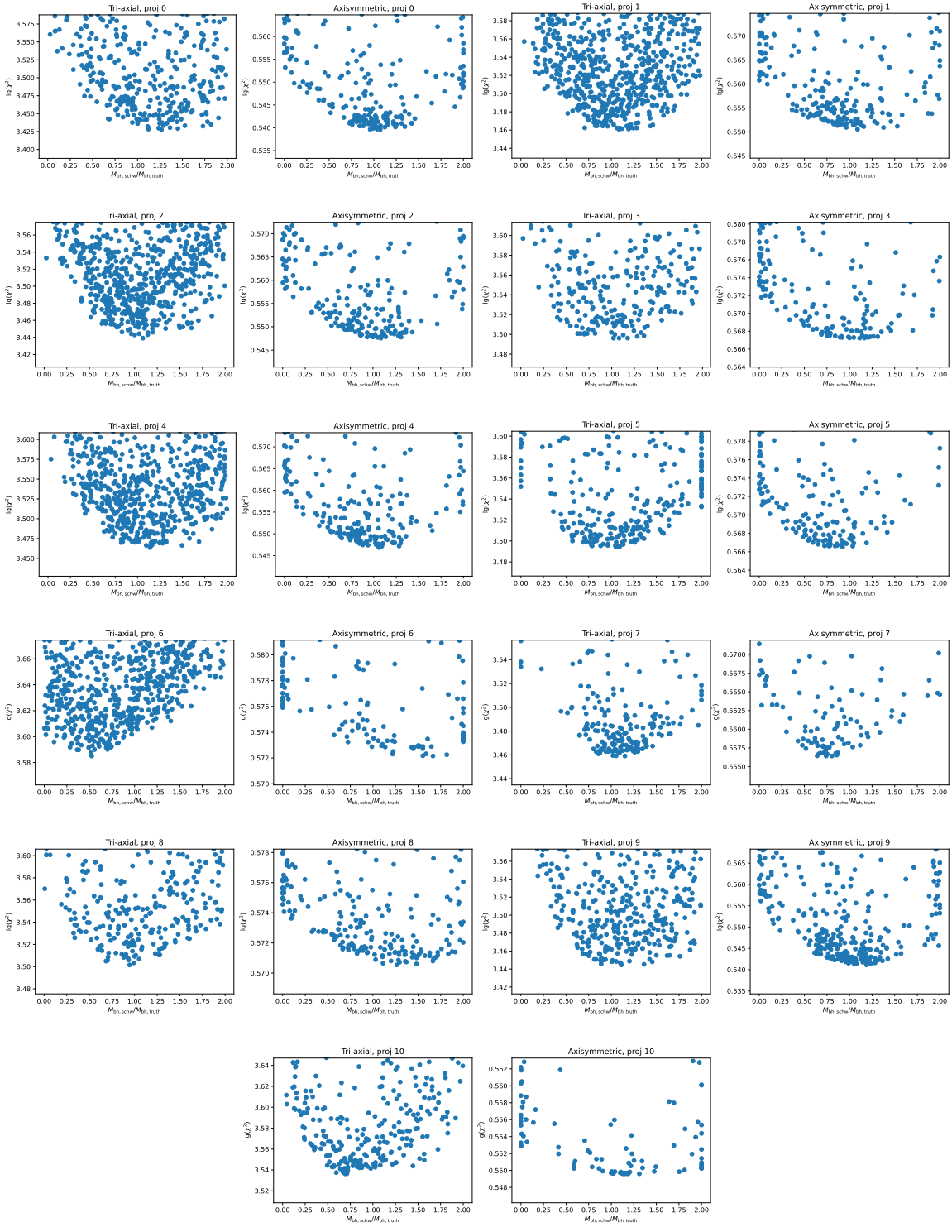

Figure 5.15: Plot of the black hole mass versus log χ^2 for sampled points with triaxial and axisymmetric Schwarzschild codes as a function of the normalised black hole mass. The projection number corresponds to the right panel of Figure 5.1. Both codes always feature a clear quadratic-shaped distribution of points with a best-fit black hole mass within a factor of two of the truth, and often much closer than this. Axisymmetric and triaxial models tend to return a best fit black hole mass that agree to within 20 per-cent. Crucially, There are no models where there appears to be any preference for a zero black hole mass.

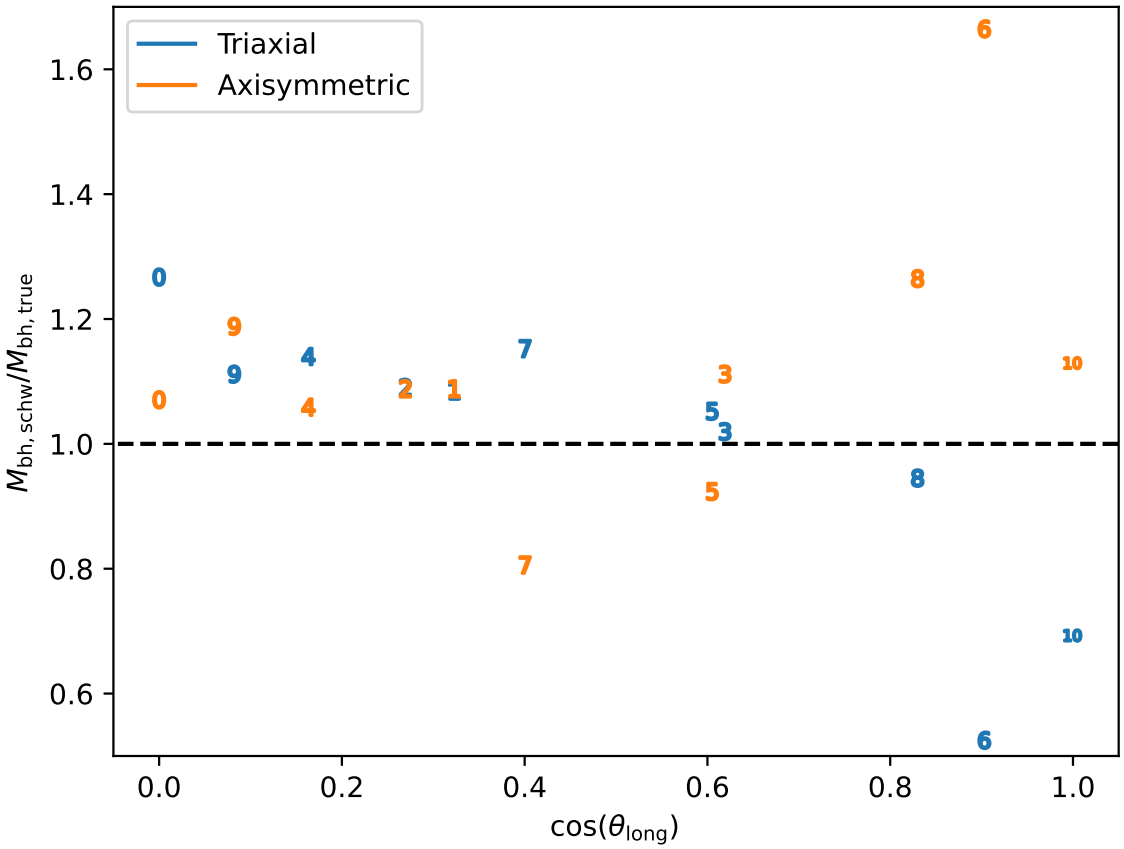

Figure 5.16: Plot of the best-fit black hole masses as a function of the cosine of the angle between the intrinsic long axis of the galaxy and the line-of-sight (θ_{long}). We see that both methods start to perform very poorly as the long axis is increasingly aligned with the line-of-sight.

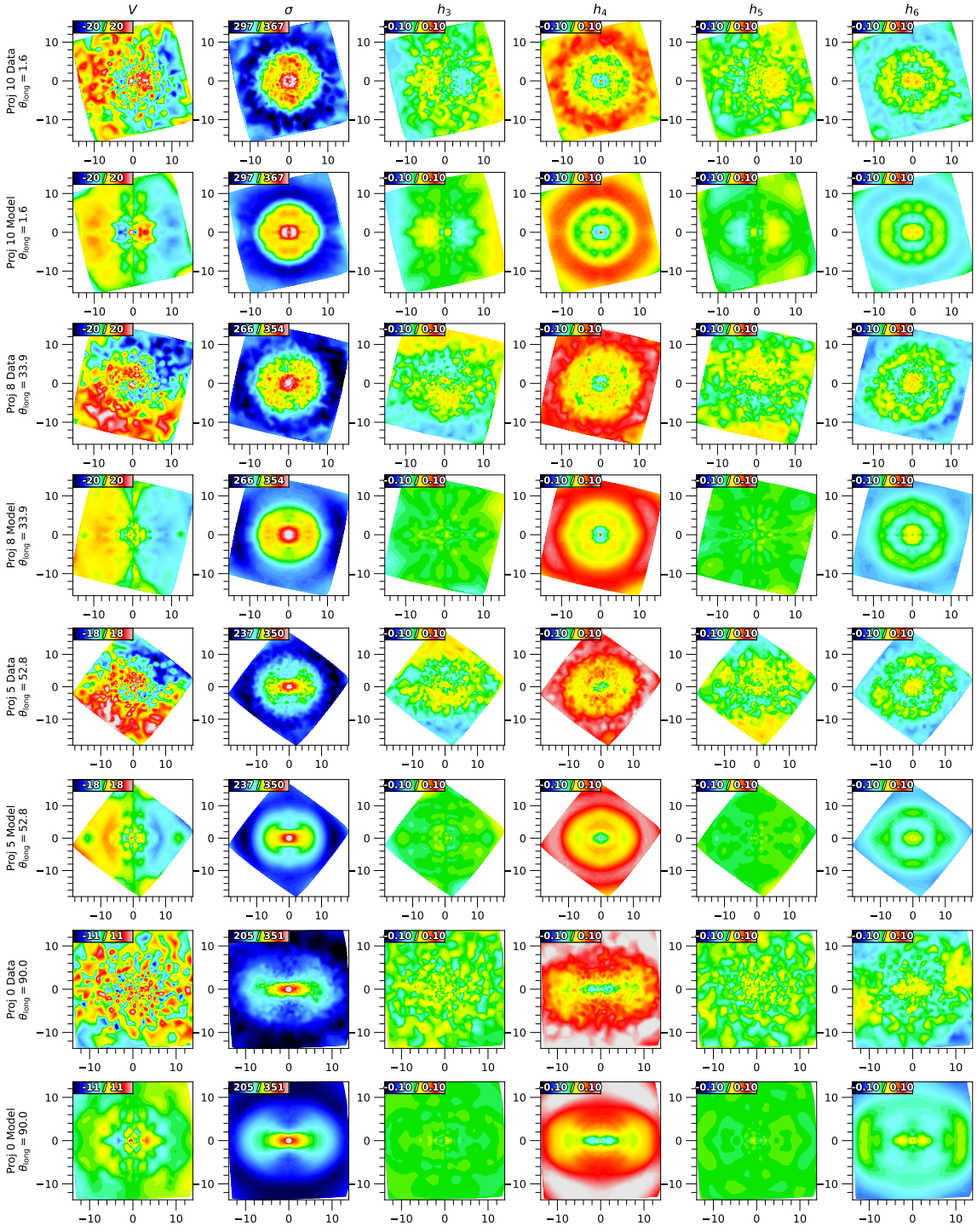

Figure 5.17: Model and data maps for the best-fit axisymmetric Schwarzschild models. Different projections are shown in order of increasing $\theta_{\rm long}$. Unlike the case of JAM, here we are able to produce believable models close to the intrinsic long axis, though we know these do not accurately recover the black hole mass.

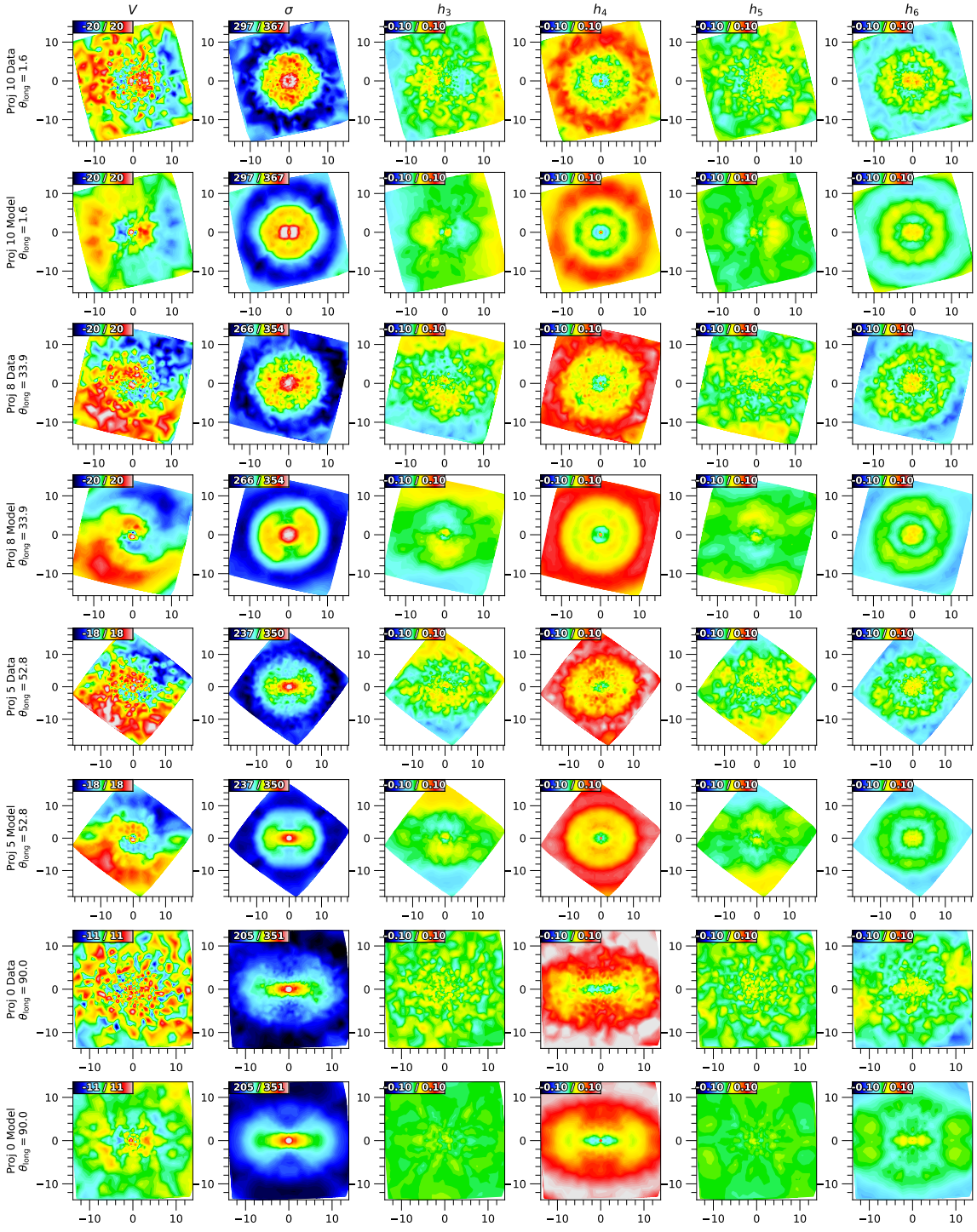

Figure 5.18: Model and data maps for the best-fit DYNAMITE models. Different projections are shown in order of increasing $\theta_{\rm long}$. Unlike the case of axisymmetric Schwarzschild, here we are able to reproduce the twists in the velocity and h_3 maps. We also see that models close to the intrinsic long axis appear unreliable.

0.1. In our tests comparing mass-follows-light models to models with a dark matter halo, we find that, for JAM, models with a dark matter halo tend to do a better job of recovering the supermassive black hole mass than without. Our Schwarzschild models without a dark halo often fail to recover a black hole mass at all. This stands in stark contrast to the situation in real galaxies. Many of the initial black hole mass determinations using stellar kinematics did so assuming a mass-follows-light model and still found evidence of a supermassive black hole. This provided strong evidence that every galaxy has a supermassive black hole in its center. That our galaxy is so sensitive to the dark halo is likely a result of the unusually large dark matter fraction and suggests that pure mass-follows-light models would likely perform much better with real data. It is worth emphasising that the motivation for the simulation came from trying to explain the large core in the galaxy NGC 1600. An initial Schwarzschild model of this galaxy revealed a large amount of dark matter similar to that used in the simulation (Thomas et al., 2016). This creates a significant caveat to our results. Namely, we expect our results to be more sensitive to the modelling of the dark matter than typical slow rotator galaxies would be.

5.6.2 Dark Matter Halo Assumptions

The choice of dark matter halo plays a central role in the interpretation of our dynamical modelling results and therefore warrants a dedicated discussion. Because this work is based on a simulated galaxy for which the true mass distribution is known, we can explore modelling strategies that are not available when analysing real galaxies. While this provides a useful controlled experiment, it also implies that some of our assumptions make the modelling problem artificially easier than in real applications.

As shown in Figure 3.1, the dark matter distribution in the simulation deviates significantly from a standard NFW profile in both the inner and outer regions. The outer deviation arises because the simulated halo is initialised as a Hernquist profile (Hernquist, 1990). For this reason, adopting a fixed NFW halo, as is commonly done in studies of real galaxies, is not well justified in this case. Instead, in order to isolate the performance of the modelling methods from biases introduced by an incorrect halo parametrisation, we adopt the exact Nuker profile fitted (both inner and outer halo slopes) in Section 3.3, leaving only its overall normalisation as a free parameter. A similar approach was taken by (Neureiter et al., 2023, Section 4.3) who assumed a steep outer halo slope of γ = -4.5.

It is important to emphasise that this strategy is not feasible for real galaxies, where the true dark matter distribution is unknown. By construction, our choice

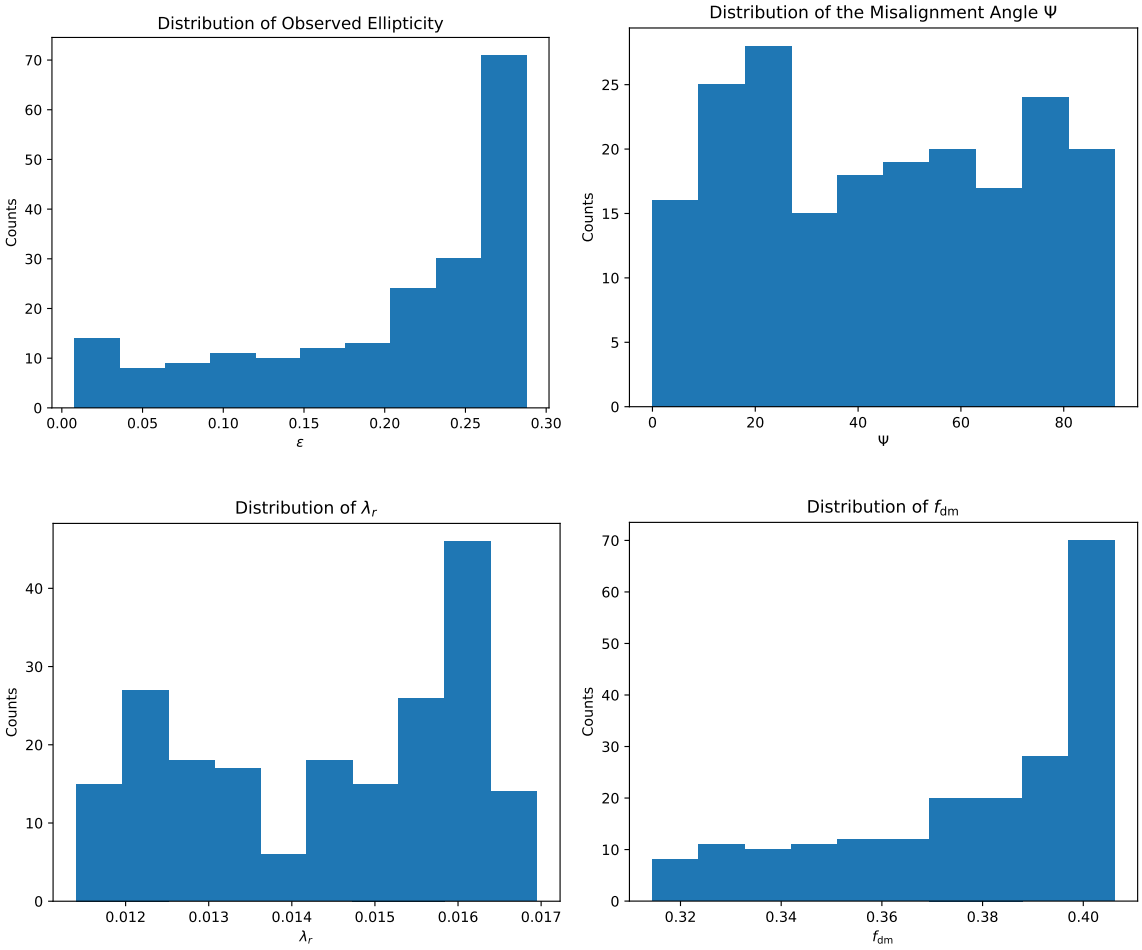

Figure 5.19: Distributions of key slow rotator parameters for all of the 202 projections used in the JAM modelling. All properties, apart from the dark matter fraction $f_{\rm dm}$, fall in the typical range for slow rotator galaxies.

removes a major source of uncertainty and therefore makes the recovery of the black hole mass easier for the models.

This point becomes particularly clear when we adopt an approach applied to many real galaxies. Namely, fitting with a mass-follows-light model. Under this assumption, both the axisymmetric and triaxial Schwarzschild models recover a black hole mass that is smaller than the true black hole by over a factor of 2, and in some cases consistent with zero. This result is surprising, as dynamical studies of massive galaxies never found zero black hole masses when the halo is not explicitly fitted. Although we do not wish to generalise based on a single object, one possible explanation is that the simulated galaxy contains substantially more dark matter than typical real massive galaxies (see subsection 5.6.1). In this case, the total mass density would differ strongly from the stellar density alone, which is inconsistent with observational results indicating little dark matter within the central regions of high velocity dispersion galaxies. Additionally, previous studies of M87 found that including a dark halo increased the black hole mass by a factor of 2 (Gebhardt & Thomas, 2009). This dependence of the black hole mass on the halo is typically understood as resulting from the degeneracy between dark halo and stellar M/L. Increasing the halo corresponds to a decrease in the M/L which, in the center of the galaxy, must be made up for by the black hole. Our dark matter fraction within one effective radius is larger than that of M87, and thus it is possible that our models returning zero black hole mass are just the end point of the black hole mass decreasing as the halo decreases.

Another key factor is the role that the geometry of the system plays in the recovery of the gravitational potential. Thomas et al. (2007) found that axisymmetric Schwarzschild modelling of triaxial or prolate galaxies systematically miss the total mass within one $R_{\rm e}$ by up to 20 per-cent. Our axisymmetric models tend to over-estimate the dark matter fraction and under-estimate the stellar mass-to-light ratio, whereas our triaxial models are closer to the truth. The reason for this likely has to do with systematics associated to the deprojection. As we saw in our JAM modelling, prolate dynamical models of this galaxy seem more reliable than oblate axisymmetric models. However, all of our axisymmetric Schwarzschild models assume oblate symmetry. Additionally, van den Bosch & de Zeeuw (2010) found that some galaxies can have an increase in the black hole mass by a factor of two when triaxial Schwarzschild modelling is performed rather than axisymmetric modelling. Here, we don't find a significant offset between the axisymmetric and triaxial models, though it should be

noted that van den Bosch & de Zeeuw (2010) did not include a halo. Further determining the relationship between the recovered black hole mass, geometry, and halo, will form the basis for future work.

The strong sensitivity of the Schwarzschild models to the assumed halo profile also suggests that, in real galaxies, the unknown form of the dark matter halo may introduce an additional source of systematic uncertainty or bias in black hole mass measurements using these models.

In contrast, JAM appears significantly more robust to the choice of halo. Reasonable black hole masses can be recovered even when no dark matter halo is included at all. This robustness likely reflects the fact that JAM does not require, and in general should not attempt, to fit kinematic data over a large radial extent, and is therefore less sensitive to uncertainties in the outer mass distribution.

Taken together, these considerations emphasise that JAM and Schwarzschild modelling probe the problem in complementary ways and respond differently to assumptions about the dark matter halo. Schwarzschild methods offer great flexibility but can be highly sensitive to the halo while JAM appears more robust to uncertainties in the outer mass distribution.

5.6.3 Previous Comparisons

There have previously been several papers that model the same kinematic data using both Schwarzschild codes and JAM. We summarise these in Figure 5.20. Of these, only one uses triaxial Schwarzschild modelling (Feldmeier-Krause et al., 2017), and only two use $\mathrm{JAM_{sph}}$ (Thater et al., 2019, 2022a). The remaining all use axisymmetric Schwarzschild or $\mathrm{JAM_{cyl}}$ (van den Bosch et al., 2012; Emsellem, 2013; Dominiak et al., 2025; Rusli et al., 2013; Krajnović et al., 2018; Drehmer et al., 2015; Siopis et al., 2009; Walsh et al., 2016).

To better understand the relationship between JAM and Schwarzschild black hole masses, we fit a line to the relationship in Figure 5.20 and find that, on average, JAM models produce supermassive black hole masses that are larger than those determined with Schwarzschild codes by about 40 per-cent. A natural question is which choice of modelling better reflects the true values. There are a few special cases with reliable black hole masses that are worth comparing.

One benchmark case is the maser galaxy NGC 4258, for which a black hole mass of $(3.9\pm0.1) \times 10^7\ M_\odot$ was determined in Herrnstein et al. (1999). An initial Schwarzschild model gave a black hole mass that was 15% too small, while a JAM model gave a black hole mass that was 22% too large (Siopis et al., 2009; Drehmer

et al., 2015). However, more recent work by Nguyen et al. (2025) performs both $\mathrm{JAM_{cyl}}$ and $\mathrm{JAM_{sph}}$ modelling for several different PSF values and the inclusion (or not) of a varying M/L ratio, which they marginalise over. They find an ensemble average that is just 5% larger than the Maser value.

Another such case is the galaxy NGC 4751 for which reliable gas kinematics are available and have been presented in Osorno et al. (2025); Dominiak et al. (2025). Both analyses find a black hole mass around 3-4$\times 10^9$ $M_\odot$. Dominiak et al. (2025) also does $\mathrm{JAM_{cyl}}$ and $\mathrm{JAM_{sph}}$ modelling and finds a supermassive black hole mass of $(2.52 \pm 0.36) \times 10^9$ $M_\odot$, and $(3.24 \pm 0.87) \times 10^9$ $M_\odot$, respectively. A previous Schwarzschild model, however, found a black hole mass of $(1.4^{+1.5}_{-1.3}) \times 10^9$ $M_\odot$, significantly less than either the gas or JAM measurements.

One especially interesting case is NGC 1277, for which an initial Schwarzschild model suggested a significantly overmassive black hole (van den Bosch et al., 2012). An N-body realisation with moments identical to a $\mathrm{JAM_{cyl}}$ model on the same data, however, suggested that an equally good fit to the data could be found with a significantly smaller black hole mass (Emsellem, 2013). This smaller black hole mass was later confirmed with higher resolution data using both axisymmetric Schwarzschild modelling and JAM modelling (Krajnović et al., 2018; Walsh et al., 2016). The reason for this discrepancy is primarily due to the fact that van den Bosch et al. (2012) measured an h_4 that was systematically too large (up to values of 0.2) and fit Schwarzschild models that perfectly matched the data. The later analysis by Walsh et al. (2016) found values of h_4 that were less than half of those from before. This case emphasises the challenge of Schwarzschild modelling. Higher order GH moments can often be systematically offset due to template mismatch which can have significant impacts on the extracted parameters in dynamical models.

Another case is that of the Milky Way. The Milky Way supermassive black hole has one of the most well measured black hole masses, with the most recent value from the GRAVITY collaboration of $(4.297 \pm 0.012) \times 10^6$ $M_\odot$ (Gravity Collaboration et al., 2023). Feldmeier-Krause et al. (2017) performed both triaxial Schwarzschild modelling as well as $\mathrm{JAM_{cyl}}$ modelling of the Milky Way nuclear star cluster and found an almost identical mass in both cases, though both values have a mean that is around 30% less than the Gravity Collaboration et al. (2023) value (though their error bars are also close to 30%).

One final case is that of M87. There have been three previous black hole mass measurements with Schwarzschild models, returning 6.0 ± 0.5, 6.2 ± 0.4, and $(5.37^{+0.37}_{-0.25} \pm 0.22) \times 10^9$ $M_\odot$ respectively (Gebhardt & Thomas, 2009; Gebhardt et al., 2011; Liepold

et al., 2023b). The benchmark measurement with the event horizon telescope found a value of $(6.5 \pm 0.7) \times 10^9\ M_\odot$ (Event Horizon Telescope Collaboration et al., 2019). In chapter 2, we measure the supermassive black hole mass in M87 and determine a number of values, but have a preferred one of $(8.7 \pm 1.2 \pm 1.3) \times 10^9\ M_\odot$. Interestingly, we recover a value of $(5.5^{+0.5}_{-0.3}) \times 10^9\ M_\odot$ if we assume the same stellar tracer distribution as in previous works, suggesting that in this case JAM and Schwarzschild are consistent when provided the same data.

Ultimately, more data and more black hole mass measurements with identical data using multiple methods will be necessary in order to better understand the strengths and weaknesses associated with each method and determine which methods, under what circumstances, out perform the other.

5.7 Conclusions

We summarise our key results as follows:

- Our galaxy is representative of the slow rotator population except for its dark matter halo which has a dark matter fraction within one effective radius that is about twice that of a typical slow rotator, and has an outter slope that is steeper than a typical NFW profile.

- Dynamical models can accurately model our triaxial galaxy for all projections where the intrinsic long axis is greater than 30 degrees away from the line-of-sight.

- The best JAM models for recovering black hole masses are $\mathrm{JAM_{cyl}}$ with constant anisotropy and prolate symmetry in a small field of view (3 times r_{soi}) or $\mathrm{JAM_{sph}}$ in a similar field of view (Figure 5.7,Figure 5.8).

- We find a mean black hole mass recovery across all reliable projections of $\mathrm{JAM_{cyl}}$ of $M_{\mathrm{bh,JAM}}/M_{\mathrm{bh,true}} = 1.04^{+0.10}_{-0.32}$ and for $\mathrm{JAM_{sph}}$ of $M_{\mathrm{bh,JAM}}/M_{\mathrm{bh,true}} = 0.80^{+0.21}_{-0.06}$.

- Axisymmetric Schwarzschild modelling recovers black hole masses with an average accuracy of $M_{\mathrm{bh}}/M_{\mathrm{bh,true}} = 1.09^{+0.08}_{-0.13}$ and $M_{\mathrm{bh}}/M_{\mathrm{bh,true}} = 1.09^{+0.06}_{-0.06}$ for DYNAMITE.

- In the literature, JAM models tend to measure supermassive black holes that are 40 per-cent larger than those measured with Schwarzschild models (Figure 5.20).

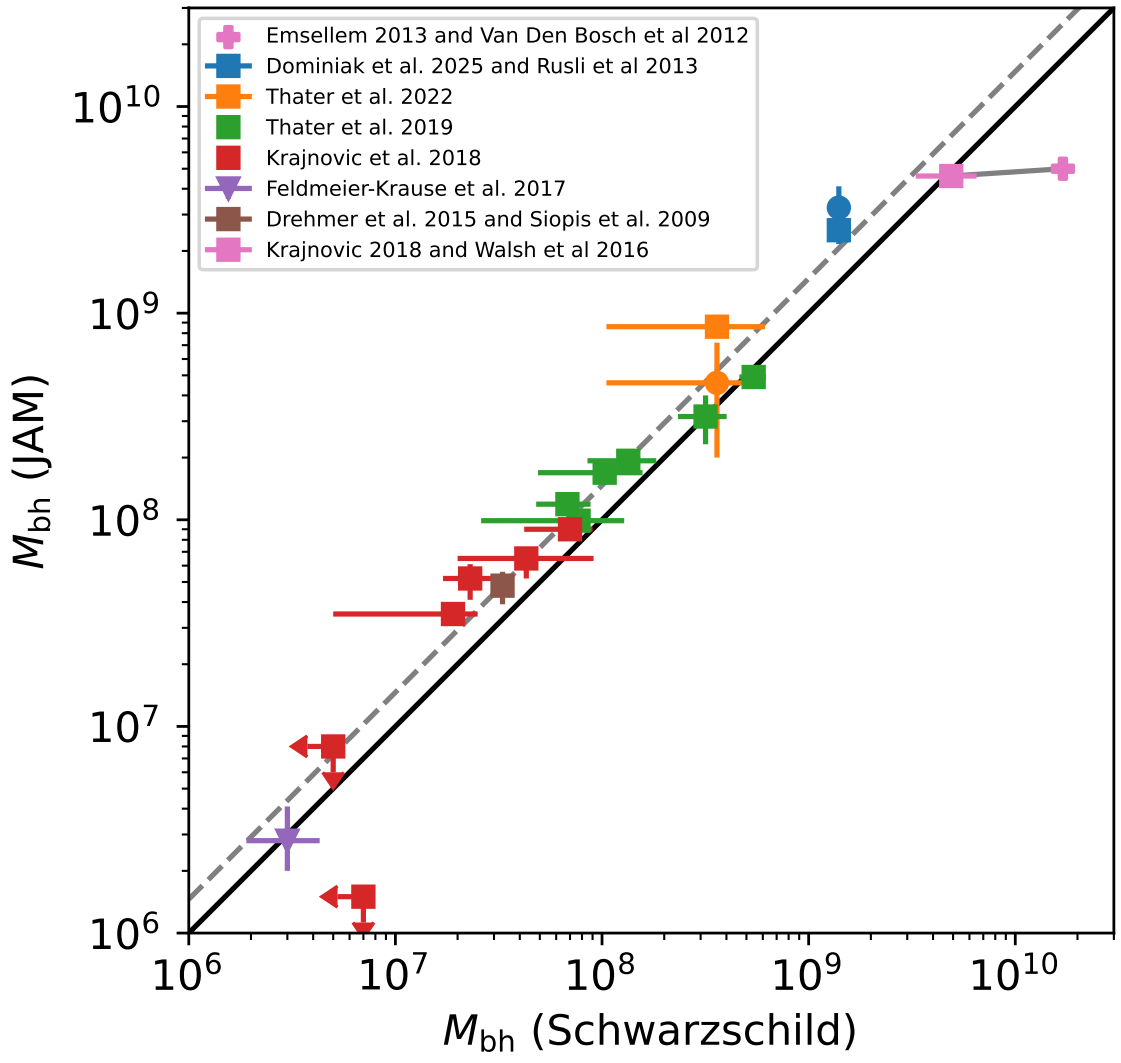

Figure 5.20: Comparison of measurements of supermassive black holes using JAM and Schwarzschild codes. The best-fit line to the data excluding upper limits gives that JAM models tend to have supermassive black holes that are about 40 per-cent larger than their Schwarzschild counterparts. Square and circle markers denote $\mathrm{JAM_{cyl}}$ and $\mathrm{JAM_{sph}}$, respectively, being compared with an axisymmetric Schwarzschild code. The triangle denotes triaxial Schwarzschild being compared with $\mathrm{JAM_{cyl}}$. The plus denotes an N-body model with moments identical to a $\mathrm{JAM_{cyl}}$ model. The two pink data points correspond to two separate measurements of NGC 1277 (the line is drawn to emphasise that this is the exact same galaxy, but with different data).

Chapter 6

Machine Learning Dynamical Models

6.1 Summary

The advent of integral field data has revolutionised the study of galaxy evolution. A key component of this is dynamical modelling methods which have allowed for crucial insights to be made from kinematic data. Despite this importance, most dynamical models make a number of key assumptions which do not hold for real galaxies. At the same time, machine learning methods are becoming increasingly powerful, with many applications appearing in astronomy. These have the potential to be used both to improve existing dynamical modelling methods as well as to build new ones. To investigate this, we construct a training set of dynamical models of early-type galaxies using Jeans Anisotropic Modelling (JAM). We then train a neural network on this data using the parameters of JAM and mock photometry as the input. We find that we are able to speed up JAM to a remarkable degree while maintaining accuracy in our models. We conclude with some remarks on the application of this to cosmological hydrodynamical simulations and the prospects of producing the next-generation of dynamical models.

6.2 Introduction

6.2.1 Limitations of Dynamical Modelling

In chapter 5, we saw that JAM and Schwarzschild dynamical modelling can recover the supermassive black hole mass of a mock galaxy with a bias from 4 per-cent up to 20 per-cent. This is in the generous case where we assume the underlying kinematics are known without any bias. One possible reason for this persistant bias is that

dynamical modelling codes make a number of assumptions which are well motivated but don't entirely hold for real galaxies. These include assumptions on the shape of the stellar distribution (often described using a multi-Gaussian expansion), assumptions on the shape of the velocity ellipsoid (spherical/cylindrical for Jeans modeling), and assumptions on the underlying geometry of the stellar system (spherical, axisymmetric, or triaxial). Advances in machine learning and artificial intelligence provide a unique opportunity for developing next-generation dynamical modelling methods that can break some of these assumptions. One important first step towards realising this goal is determining what machine learning architectures are best suited for dynamical modelling.

6.2.2 Dynamical Modelling Workflows

The workflow for dynamical modeling techniques depends significantly on the method. Jeans modelling, as described in previous chapters, proceeds by two independent streams. First, photometry must be observed and turned into an MGE to describe the geometry of the stellar system. Next, instrument specific parameters, such as the PSF, must be determined. Then, properties of the galaxy, such as its distance from earth, anisotropy, M/L ratio, supermassive black hole mass, etc. are chosen. All of these are then taken as function inputs for JAM modelling and the output is a unique prediction for the first and second velocity moments according to the Jeans equations. This predicted map is then compared with the observed map, and the parameters for the gravitational potential and anisotropy are optimised for, typically using a MCMC. A diagrammatic representation of this is given in Figure 6.1.

This approach is fundamentally different from that of Schwarzschild modelling. Jeans modelling produces a unique prediction of the velocity second moment from a set of parameters, which is then compared to observations via an MCMC. In Schwarzschild modelling, model predictions cannot be made without prior knowledge of the data you are fitting. That is to say, a Schwarzschild model requires a set of observations in order for the set of orbits to be constrained to the data. Then, when a model has been fit, it can be compared with the original data and optimised for its parameters (typically M/L, black hole mass, etc.). A diagrammatic representation of this is given in Figure 6.2.

Each of these workflows represents a different approach to dynamical modelling. Additionally, each of these approaches have the potential to be augmented with machine learning (by replacing chunks of the workflow with neural networks). An example of this is Gomer et al. (2023), which took the JAM workflow and replaced

the MGE and JAM portions of the workflow with a neural network. Another approach, similar to the Schwarzschild workflow, is to take observed photometry and kinematics as an input, and then directly output the parameters of the gravitational potential. This was the approach taken by Chu et al. (2024). Another possibility for Schwarzschild modelling would be to replace part of the internal workflow, such as the generation of the orbit library or the determination of orbit starting conditions with a neural network. Each of these approaches is interesting and warrants an independent study. In this work, we consider a machine learning approach similar to Gomer et al. (2023) where we augment the JAM workflow by replacing the MGE and JAM parts of the workflow with a neural network.

It is worth emphasising that the above list of workflows is not exhaustive and other, potentially more interesting, possibilities remain. The above deals primarily with dynamical modelling, but one could imagine pushing back the problem one step further and directly using the observed IFS cube as input and then taking the output to be the parameters of the gravitational potential. This is an attractive possibility as it directly connects the lowest level data product before any processing has been done (full spectral fitting for the kinematics), with the final output (gravitational potential). While interesting, this faces a number of challenging issues, not the least being how to generate representative training data that is applicable to real galaxies. A naive implementation would likely involve a number of significant limitations, such as which stellar templates were trained on, which additive and multiplicative Legendre polynomials were used for the full spectrum fitting, and more (see chapter 4 for some issues). Such an investigation is beyond the scope of this work. For now, we continue with the aforementioned modified JAM workflow.

6.3 Machine Learning Architecture

The purpose of our neural network is to replace the part of the JAM workflow involving fitting the MGE component and solving the Jeans equations to produce the output second moment map. We show a diagram of our neural network architecture in Figure 6.3. For the first part of the machine learning architecture, we use a convolutional neural network (CNN) to transform the mock photometry into a feature vector of length 64. We experimented with the size of the feature vector and found diminishing returns beyond 64. The network has a total of 6 layers: four of these are convolutional layers with hidden layer size of 128 and 2 of these are 2 by 2 max pool layers. The convolution layers have kernel size 2 and a stride of 2.

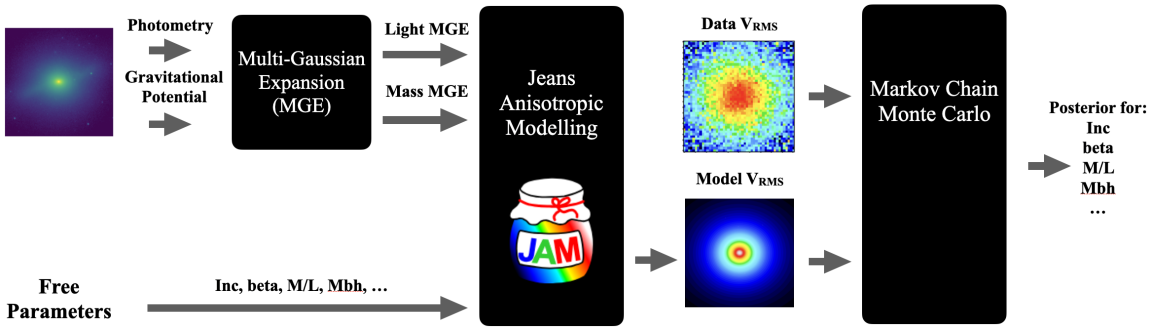

Figure 6.1: Schematic of the typical JAM data pipeline. Information about the photometry and gravitational potential are parameterised using an MGE expansion which is then fed into JAM. Other JAM parameters are directly fed into the model. An MCMC is then used to determine posteriors for the parameters in the gravitational potential.

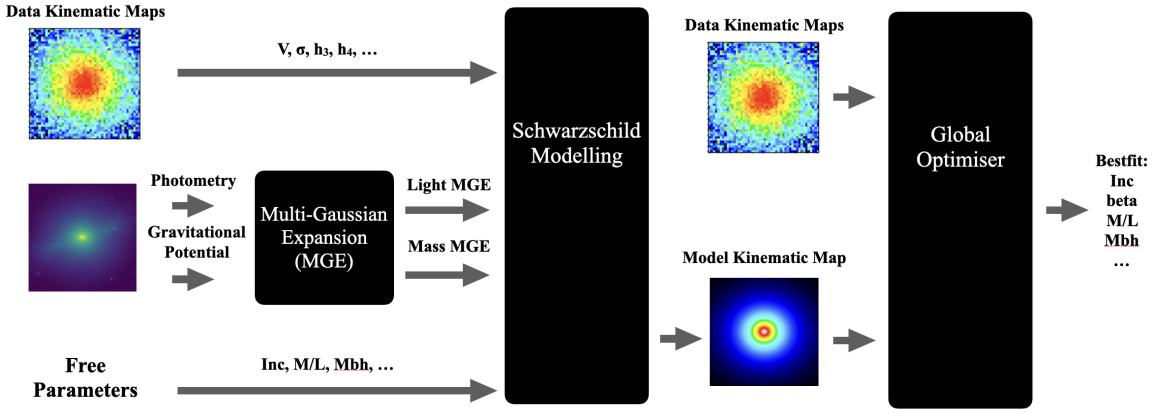

Figure 6.2: Schematic of the typical Schwarzschild data pipeline. Information about the photometry and gravitational potential are parameterised using an MGE expansion (or other parameterisation) which is then fed into the Schwarzschild model. Other parameters are directly fed into the model. Additionally, knowledge of the observed maps must be provided to allow for the orbits to be fit. Parameters for the gravitational potential are then sampled on a grid or optimised for with a black box global optimiser.

For the second section, we use a 7 layer MLP with input size of 67 (3 for the non-photometry input parameters of Table 6.1 and 64 from the output of the CNN), hidden layer size of 4096, and output size of 200 to match the grid of the JAM training data (see section 6.4 for more details). We chose LeakyReLU with a slope of 0.01 as our activation function. Note that this architecture exactly copies the workflow of JAM. Another valid choice would have been to directly feed in the second moment maps and photometry and extract the physical parameters. In the standard JAM picture these parameters would be extracted from observations using something like MCMC. Note that our architecture for reconstructing the second moment maps does not use convolutions or anything else which explicitly takes into account information from neighboring grid points. In principle this would increase the accuracy of the model (at the cost of the performance). However, as we will see in subsection 6.6.1, the current architecture achieves sufficient accuracy.

The choice of activation function was also tested extensively. We initially considered the ReLU activation function. However, we frequently found that this resulted in dead neurons randomly dispersed across our output images. We also tested LeakyReLU (Maas et al., 2013) and PReLU (He et al., 2015). We found a notable improvement with both LeakyReLU and PReLU with there being little difference between these two. Since PReLU introduces an additional parameter for training, we chose to proceed just using LeakyReLU with a slope of 0.01.

In an attempt to improve the speed with which the model was trained we tested two different choices of initial conditions. The first choice (the default PyTorch implementation) is Glorot/Xavier initialisation (Glorot & Bengio, 2010). In this case the weights are distributed according to a uniform distribution. The other choice we tested was Kaiming/He initialisation (He et al., 2015). In this case the weights are randomly distributed according to normal distribution. We empirically found that the final distribution of trained weights tended to be Gaussian, though sometimes with a small bias. We tried both of these initializations but found that they make a negligible impact on the accuracy of the final trained model. As such, we proceeded with the default PyTorch implementation of Glorot/Xavier initialisation.

6.4 Training Data

The choice of training data determines the scope and limitations of the learned machine learning model. Our primary interest is determining which neural network architectures are well suited for performing dynamical modelling. For this reason,

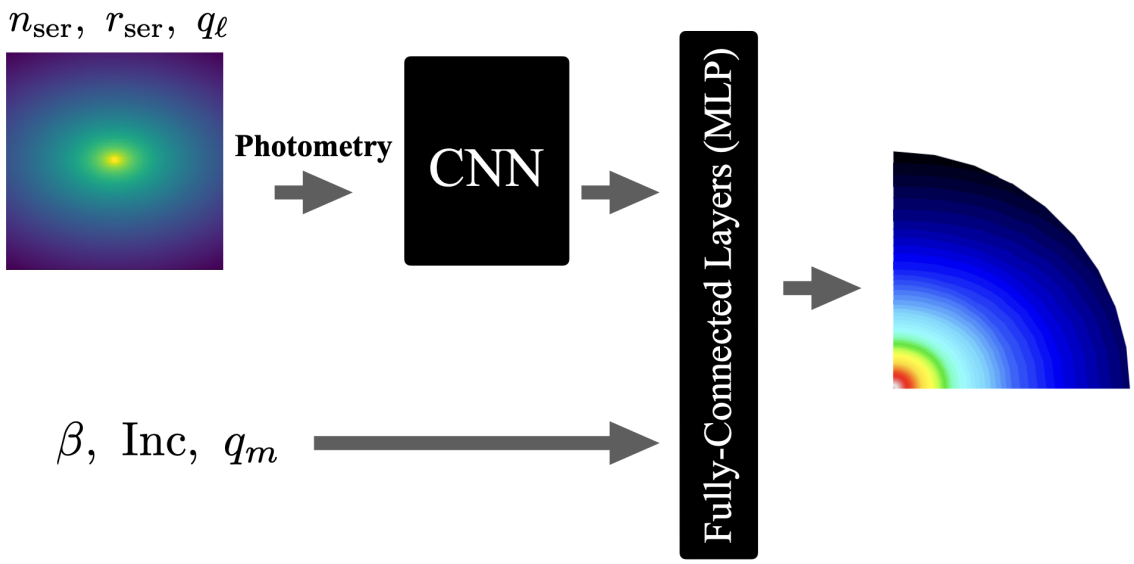

Figure 6.3: We show a schematic of our neural network architecture. This broadly mimics JAM, but with the multi-Gaussian expansion replaced by a convolutional neural network and the final stage replaced with a multi-layer perceptron.

following Gomer et al. (2023), we construct our training data using galaxies with Sérsic photometry as the input and $\mathrm{JAM}_{\rm cyl}$ second moments as the output of the model. We show the parameters that we vary in Table 6.1. The data is taken at fixed distance = 1 Mpc and with a total field of view of 70″. The choice of distance is not relevant since the impact of the distance d is to scale the second moment by a factor $d^{-1/2}$. In order to train the machine learning architecture more effectively, it is important to only sample parameters corresponding to physical solutions to the Jeans equation. With this in mind, we add an additional constraint to our parameter sampling from Wang et al. (2021). There, they show that galaxies with $\beta_z > 0.7\epsilon_{\rm intr}$ correspond to non-physical solutions of the Jeans equation and no galaxies have been observed in this region of parameter space. In practice these correspond to $V_{\rm rms}$ maps that exhibit extremely sharp behavior.

One key difference between this work and Gomer et al. (2023) is that we do not directly feed all of the parameters in Table 6.1 into our machine learning architecture. Rather, we use the parameters $n_{\rm ser}$, $R_{\rm ser}$, and $q_{\rm L}$ to generate mock photometry which is then fed into the machine learning algorithm. We generate this mock photometry using the ASTROPY function SERSIC2D. The dimensions of the images are 64×64.

After randomly sampling the parameters above we generate the corresponding JAM models. To do this, we first generate an MGE corresponding to each set of mock-photometry. This is done using the routine MGE_FIT_1D from the MGEFIT

Table 6.1: Randomly sampled parameters used to generate the training data. Parameters are sampled according to a uniform distribution within the bounds.

Parameter	Description	Bounds
$n_{\rm ser}$	Sérsic Index	[2, 4]
$R_{\rm ser}$	Sérsic Radius	[5″, 20″]
$q_{\rm L}$	Light Axial Ratio	[0.6, 1]
$q_{\rm M}$	Mass Axial Ratio	[0.6, 1]
$\beta_{\rm z}$	Anisotropy	[-0.4, 0.4]
i	Inclination	[arccos(0.6), 90°]

package of Cappellari (2002) to the Sérsic profile with axial ratio equal to 1, and then setting the axial ratio for each MGE element equal to $q_{\rm L}$. We then use the routine JAM_AXI_PROJ from Cappellari (2008) under the assumption of cylindrical symmetry to generate the velocity maps. One key difference between this work and Gomer et al. (2023) is that rather than sample the kinematic maps on a regular square grid, we sample the maps on a grid uniformly spaced in θ and logarithmically spaced in radius. We do this partly to improve the speed of the architecture since it dramatically reduces the number of data points output by the algorithm, but also because JAM internally operates by evaluating the Jeans equation on this θ-log radius grid and then interpolates to the desired points. Thus, sampling on this grid is well motivated from the JAM method, is faster than sampling on a square grid, and can be readily applied to real observations since it is straightforward to interpolate on this grid to whatever data points are required by observations . Due to axisymmetry, we only sample these points on one quarter of the field of view. We show an example of one such map in Figure 6.4.

We use a single CPU core to generate 10000 samples for the training data. This took around 3 hours. This is significantly faster than the time it took to generate the training data in Gomer et al. (2023). This is because in that work they determine the MGE profile for the JAM models by fitting MGE models to their 2D mock photometry. In this work, however, we perform a fit to the 1D Sérsic profile and then adjust to the choice of $q_{\rm L}$.

It is worth emphasising the constraints imposed by using this set of training data. The use of JAM models for the training data means that the range of applications of this machine learning architecture will suffer the same restrictive assumptions of JAM modelling. In this case, our goal is a preliminary study to determine properties of machine learning neural networks that are suitable for dynamical models. For this reason, we accept the limitations of our choice of training data, and instead focus

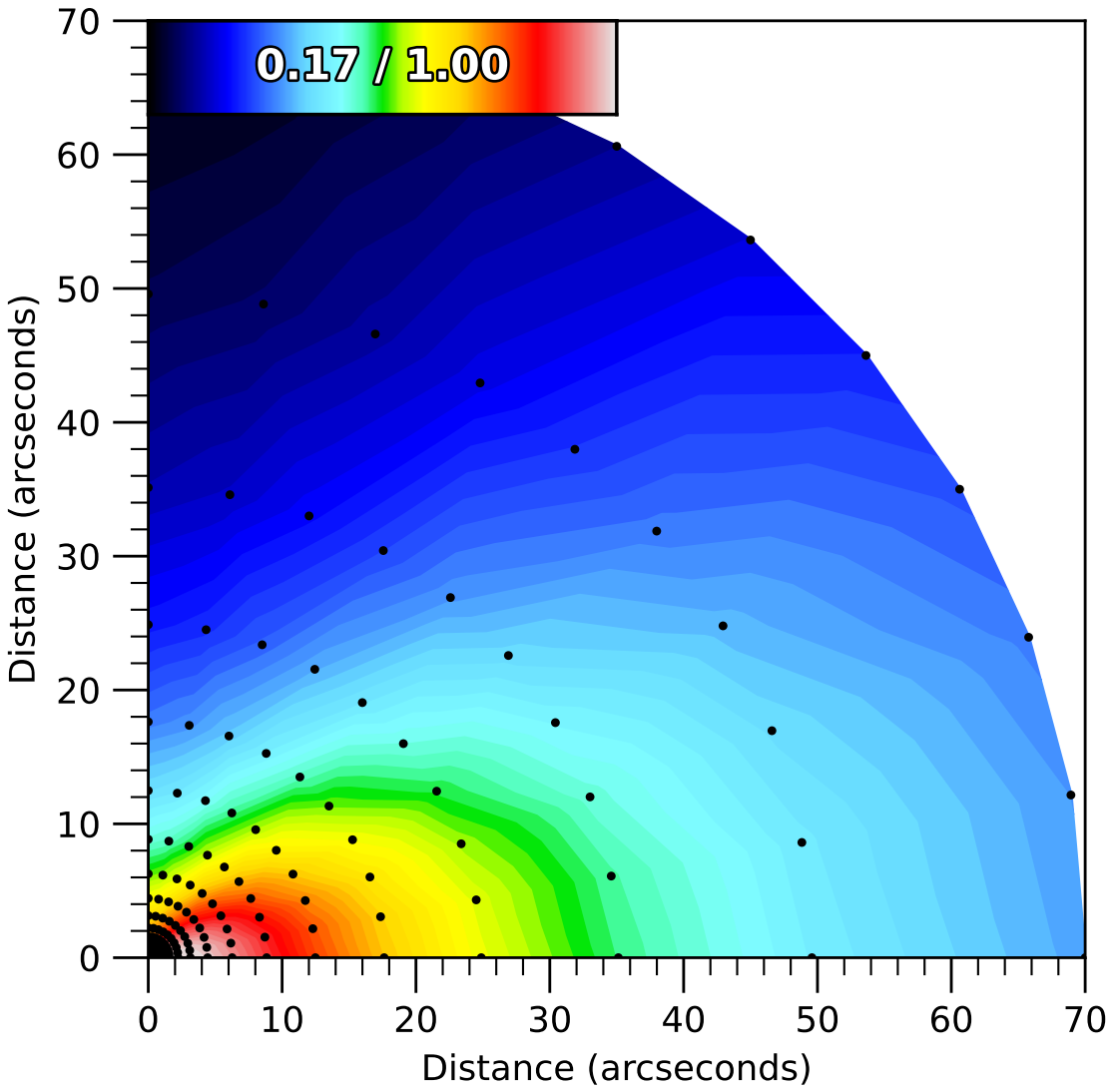

Figure 6.4: Sample training set galaxy V_{rms}. Note that due to axisymmetry it is only necessary to sample data from one quarter of the total field of view. The radial points are sampled logarithmically while the angles are sampled uniformly. The maximum second moment is normalised to 1.

on how to best design machine learning neural networks to match this data. Once a suitable machine learning architecture has been found then it is straightforward to change the training set to another, more general set of data.

6.5 Training

We train with the mean squared error loss function and the ADAM optimizer using an initial learning rate of 0.001. We train using the PyTorch scheduler ReduceLROnPlateau. This works by decreasing the learning rate by a factor λ after the loss has failed to decrease after a certain number of epochs called patience. We set the patience to 20 and the factor to $10^{-1/8}$. We use a batch size of 50 and train for 800 epochs. We reshuffle the training galaxies at the start of each epoch. We train on 8000 galaxies and retain 2000 for the test set. We train using a single NVIDIA V100 GPU. The time it takes to train is approximately 30 minutes.

6.6 Performance

6.6.1 Accuracy

In principle, a machine learning model can be trained to arbitrary accuracy given sufficient training data and model complexity. Real integral field derived kinematics have an accuracy of 6-7 km s^{-1} Cappellari et al. (2011b), so this is the benchmark accuracy we aim for. Any machine learning model which achieves an accuracy significantly lower than this is unnecessarily trading performance for accuracy. We show some outcomes of our model for the test set in Figure 6.5. The top row shows the mock photometry, the central row the true JAM model, and the bottom row the ML prediction. We see that they are identical by eye.

We quantify the accuracy of our model on a pixel-by-pixel level using the metric $\mathcal{E}$ = (Model - Truth)/Truth. We show the results for this for our test data in Figure 6.6. In the top panel we show a histogram of the results for every data point in the test data. In the bottom panel we show the results for this where we only consider the maximum of $|\mathcal{E}|$ over each $V_{\rm rms}$ map in the test data. From this we see that the vast majority of the maps are accurate to within 1 per-cent and 90 per-cent of the maps have a max-error less than 2 per-cent. The maximum $V_{\rm rms}$ one is likely to encounter for a typical galaxy is close to 100-300 km s^{-1}, so 2 per-cent accuracy corresponds with an error of 2-6 km s^{-1}. It is worth noting that this is not as accurate as that of Gomer et al. (2023) which trains using fewer galaxies, but uses a

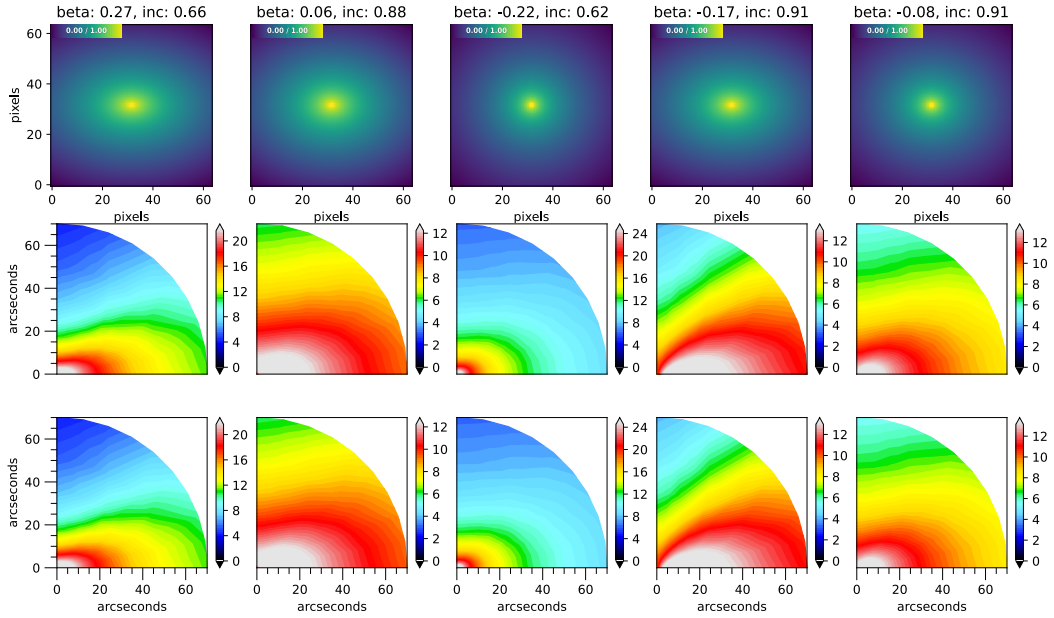

Figure 6.5: This figure shows the performance of our machine learning model for a random selection of test galaxies sampling different balues of β and the inclination. The top row shows the mock photometry, the middle row shows the true $V_{\rm rms}$ map, and the bottom row shows the output of the neural network. The two appear almost visually indistinguishable.

significantly more complicated machine learning architecture and does not use mock photometry. This demonstrates that there are many potential viable machine learning architectures that can be used for dynamical modelling which can be selected on the basis of whether or not one is trying to optimize for performance, training set size, speed, or accuracy.

One final test we perform is how the accuracy depends on the total number of galaxies in the training set for our given choice of machine learning architecture. To do this, we perform training where the only variable we change is the number of galaxies in the training set. In Figure 6.7 we show how the error at the 50th and 90th percentile changes as a function of training size. We compare this to an averaging strategy, that is, the neural network that outputs the mean pixel value of all of the training set galaxies. This serves as a lower benchmark, showing when the model starts to learn non-trivial features. We see from the figure that the training of

non-trivial features starts around a training size of 500 galaxies and decreases with diminishing returns to less than one per-cent at a training set size of 10,000. This shows that, for our choice of neural network architecture, close to 10,000 galaxies is the lower bound to achieve 1 per-cent accuracy.

6.6.2 Speed

In addition to accuracy, it is important to carefully benchmark the speed of the model. Because the machine learning architecture we use is a simple convolutional neural network fed into a multi-layer perceptron, we expect significant speed improvements over previous work. To quantify this, we randomly sample 1000 input parameters and record the time it takes the model to run for each set of parameters. A histogram of this is shown in the left-hand side of Figure 6.8. We see that it takes on average 1 millisecond to run the model. However, in general, this over-estimates the total amount of time required to run the model. This is because the CNN part of our machine learning architecture only needs to be evaluated one time for each choice of photometry. Thus, for practical uses, such as running MCMC chains, it is only necessary to run the CNN for each set of photometry once, and then one can repeatedly vary parameters in the MLP. We show the time it takes to run just the MLP part of the architecture in the right-hand side of Figure 6.8. This shows that a majority of the time spent evaluating the model is spent in the CNN, with the remaining spent by the MLP adding to only 0.3 milliseconds. This is faster than the model in Gomer et al. (2023) by around a factor of 300.

6.7 Adapting to Non-Sérsic Photometry

So far we have only been concerned with training data that solely uses Sérsic photometry. In principle we would like to expand this to a larger class of input photometry, ultimately using original unprocessed telescope images as the input. One challenge for this is how to generate a representative training set for a broad class of possible input photometry. We attempt to solve this problem by taking photometry from the MaNGA DynPop project of Zhu et al. (2023). This project uses SDSS r-band photometry to generate MGEs for each of the galaxies in the sample.

We construct our training data similar to the previous case by randomly sampling one of the galaxies from the MaNGA DynPop project and creating a $\mathrm{JAM}_{\mathrm{cyl}}$ model with constant anisotropy and inclination using the MGE for that galaxy. Using this, we generate 50000 models. There are two natural choices of photometry input

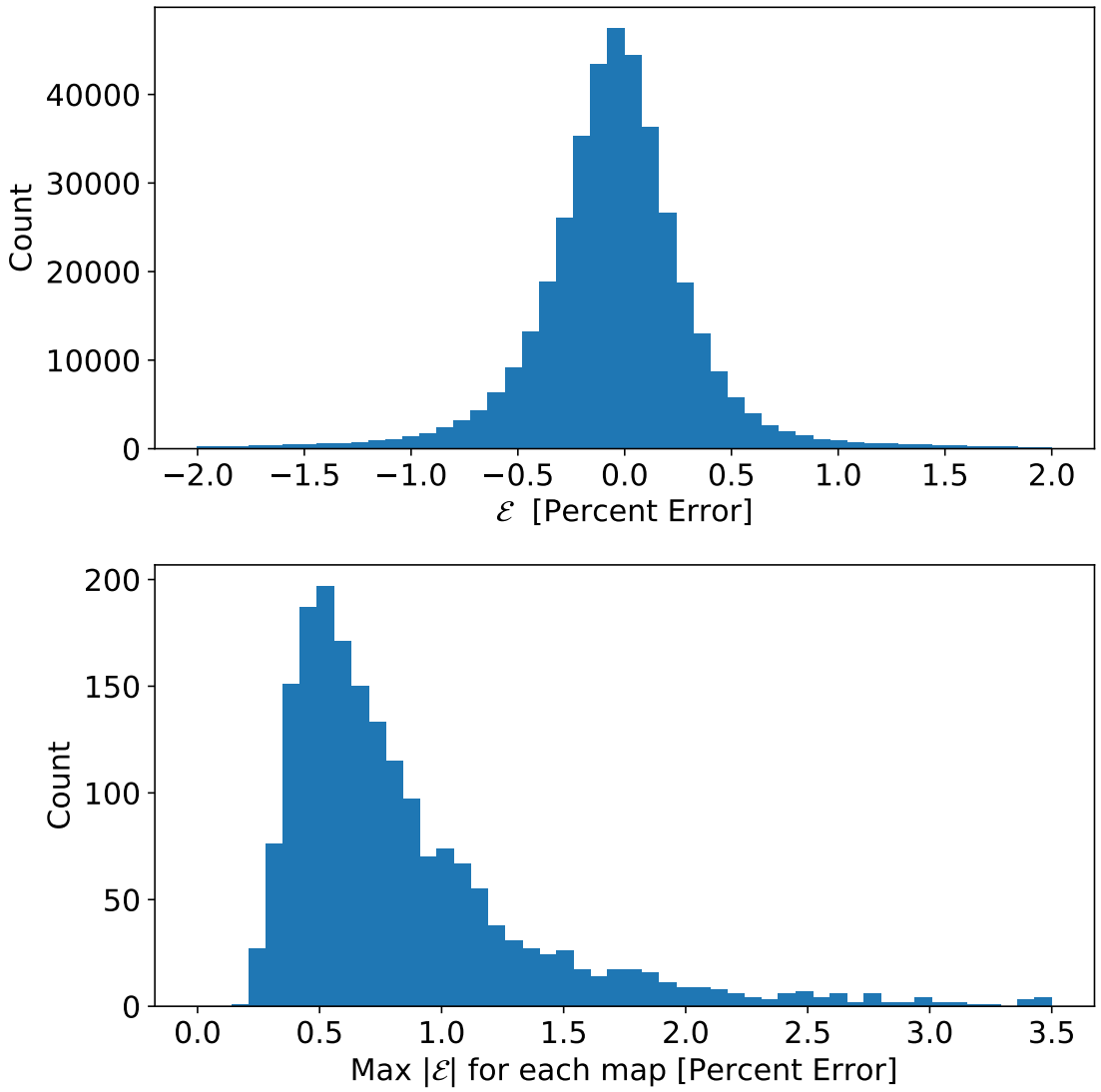

Figure 6.6: Top panel shows a histogram of the per-cent error of every pixel for every galaxy in the test set of galaxy kinematics. Almost all of the errors are less than 1 per-cent. Bottom panel shows a histogram of the maximum error for each galaxy in the test data. Here we see that the max error for almost every galaxy is less than 3 per-cent.

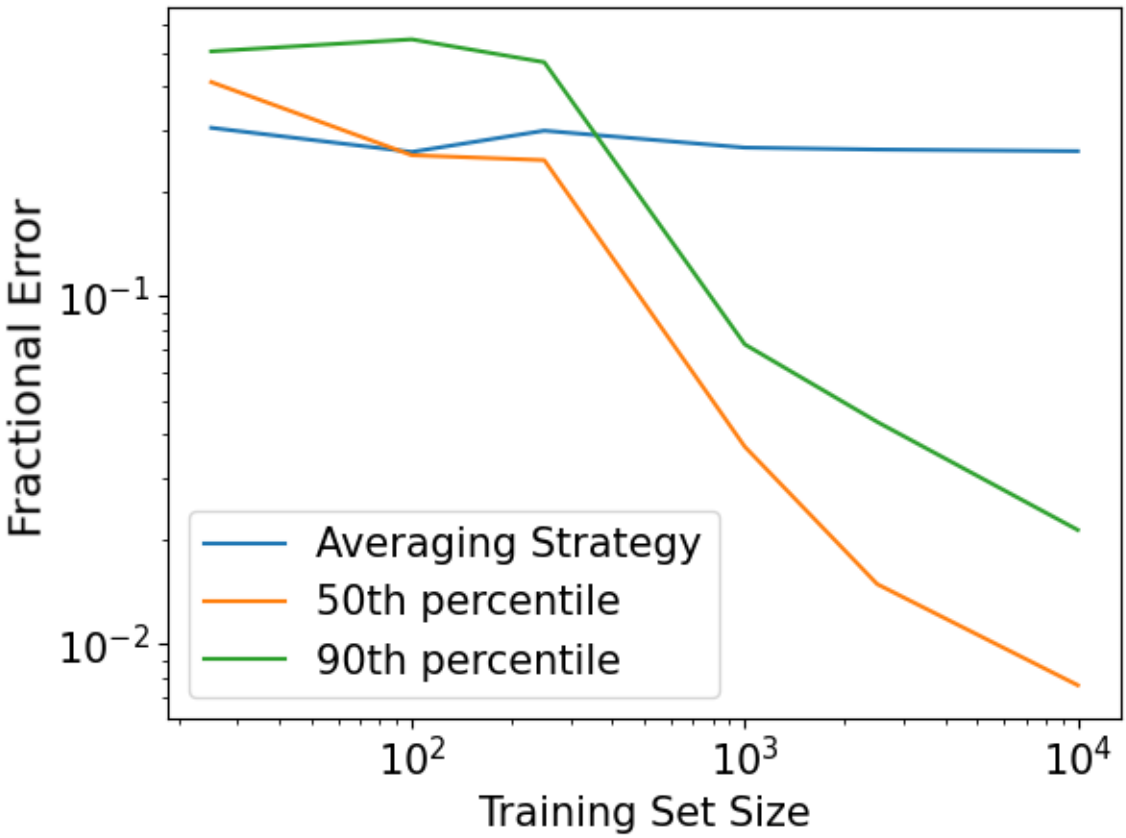

Figure 6.7: This figure shows the relationship between the training set size and the per-cent error of the model on the test data at the 50th and 90th percentile. We see that, for our machine learning architecture, close to 10,000 galaxies are necessary in order to reach close to 1 per-cent accuracy.

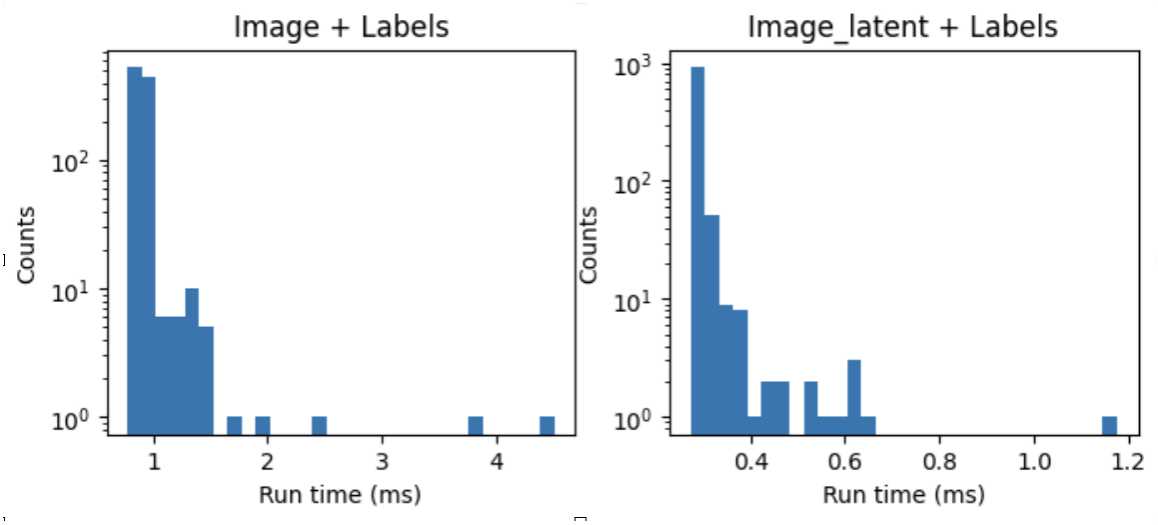

Figure 6.8: Left panel: histogram of the time it takes to run our full trained machine learning algorithm for one thousand randomly sampled sets of parameters. Right panel: histogram of the time it takes to run the MLP part of our neural network. This is the part that dominantly contributes when doing an MCMC chain. We see that it is the CNN part of our neural network that dominates the total run time.

in this case. One is to use the original galaxy photometry, along with some data augmentation (rotations by 90 degrees). The other is to create a pristine galaxy image using the MGE. We show an example of this in Figure 6.9. We proceed by testing each of these cases. Special care must be taken to avoid train-test set contamination in this case. We perform our training set/test set split by selecting a small subset of galaxy photometry and collecting all of the maps based on that photometry to serve as the test set. This ensures that our test set uses galaxy photometry that the model has not previously been trained on.

We find that using the real SDSS photometry as the input image results in significant inaccuracies in the trained model, with the typical error at the 90th percentile equal to over 30 per-cent. For the case of the MGE photometry, this is 7 per-cent. This suggests that the simple CNN currently used for extracting features from the photometry is only well optimized for relatively simple photometry. Once realistic features such as foreground stars or background objects are included, the performance diminishes substantially.

6.8 Cosmological Simulations

So far, the problem we have been concerned with is how to create a machine learning implementation of JAM. This is useful because there are some potential practical

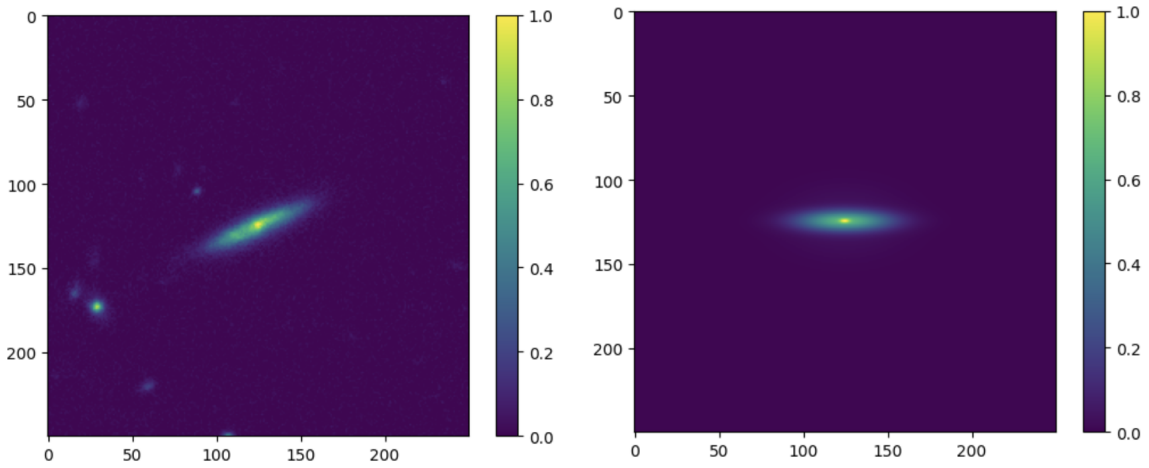

Figure 6.9: The left panel shows a normalised SDSS r-band image used in the MaNGA DynPop project. The right panel shows the MGE of the same galaxy but rotated so that the major axis is aligned with the x axis. This simplifies the photometry significantly by effectively masking all of the non-galactic light and removing one dimension (position angle) that the model needs to fit.

advantages to machine learning implementations of JAM, especially when it comes to decreasing the time it takes to do dynamical modelling. A more important point, however, is that once an adequate machine learning architecture for dynamical modelling has been determined, it is merely a matter of changing the training data to create something that applies to a more general class of problems. It is to this question that we turn our attention now.

The machine learning model described previously inherits all of the advantages and disadvantages of JAM modelling. This is because we created our training data generated using JAM. One could just as well use the same machine learning architecture but use training data without the same assumptions and limitations that JAM has. There are a number of different ways to do this. One such way is to use mock observations of cosmological simulations.

One difference of fundamental importance between these two approaches is that using mock observations of cosmological simulations introduces errors into the training data. This is because the kinematic maps are constructed out of a finite number of particles along the line of sight in the simulation which are subject to both Poisson noise, as well as systematic offsets due to the simulation resolution (Harborne et al., 2024). Ideally one would have several thousand particles along this line of sight in order to minimize the uncertainty in the shape of the line of sight velocity distribution. However, in cosmological simulations, there is a trade off between the

simulation resolution and the overall volume of the simulation and as a result it is often only possible to have on the order of tens of particles along a single line of sight in order to adequately resolve a mock observation.

We perform a test using mock observations from Chu et al. (2024), which is based on the TNG100 simulation (Nelson et al., 2018; Springel et al., 2018; Marinacci et al., 2018; Pillepich et al., 2018; Naiman et al., 2018). A complete description of the mock observations are available in Chu et al. (2024) but we describe some key features of the data. Galaxies are identified using the SUBFIND algorithm of Springel et al. (2001); Dolag et al. (2009). We select galaxies at redshift $z = 0$ with a total mass less than 10^{14} solar masses and a mass within 30 kpc greater than 5×10^9 solar masses. This gives us around 5000 galaxies. This is somewhat smaller than the number of galaxies used in Chu et al. (2024) as we also only select galaxies for which a gNFW halo has been fitted. Mock observations of the kinematics are made out to a field of view of 3 effective radii with a 48 by 48 pixel size. r-band mock photometry is produced on a denser 150 by 150 grid (corresponding to a spatial size of 0.04 $R_{\rm eff}$ per pixel) and is processed for dust attenuation effects using the semi-analytic approach of Xu et al. (2017). We saw in Figure 6.7 that close to 10000 galaxies are required for per-cent accuracy. To close this gap, we perform data augmentation by including projections of these galaxies across the x, y, and z axes of the TNG simulation snapshots.

Next, we need to choose which parameters to treat as the input to the neural network. It is worth emphasizing how this step is unique from the JAM case. JAM is effectively a mathematical function that connects given inputs to a unique output. In the case of cosmological simulations, we don't exactly know all of the parameters necessary to uniquely specify a given velocity map. Furthermore, it is usually not single parameters that are necessary but entire functions describing things like the orbital anisotropy or varying M/L ratio. As such, the question of what parameters to choose is very interesting and worthy of a study in deeper detail. For now, we follow the JAM case and choose the equivalent parameters from the TNG100 simulation (with the addition of the parameters from a gNFW halo and the orientation of the galaxy). These are summarized in Table 6.2.

Determining the accuracy of the neural network is difficult due to the noise inherent in the training data. One way that we account for this is by using a modified version of the loss function. We continue to use the mean squared error, but we weigh each pixel by the Poisson noise due to the finite number of particles included. For now, we perform a visual inspection of the models performance on the test data. Some samples of this can be seen in Figure 6.10. From top to bottom we have the mock

Table 6.2: Parameters taken from the cosmological simulation that are used for the training of the neural network.

Parameter	Description
$M/L(r < R_{\rm e})$	M/L within 1 $R_{\rm e}$
$\beta_{\rm z}(r < R_{\rm e})$	Anisotropy within 1 $R_{\rm e}$
$f_{\rm dm}(r < R_{\rm e})$	$f_{\rm dm}$ within 1 $R_{\rm e}$
γ	DM halo inner slope
$r_{\rm s}$	DM halo break radius
i	Inclination
θ	Orientation

photometry, the kinematic maps from the simulation, our predicted maps over the range of valid data, our predicted maps over the full field of view, and the residuals. The valid data is determined by whether or not there are more than two particles along the line of sight, all other data points are masked (the black pixels). We see that the major features of the maps, such as the orientation and overall shape are well recovered by the network. Important details like the amplitude of the maps, however, is not always accurate. In JAM modelling this would suggest that there is an issue with the overall mass normalization or details of the MGE parameterization. For our neural network, it suggests that improvements should be made in the way that we handle the overall normalization of the data, as well as suggesting that some changes to our CNN could lead to improvements.

6.9 Conclusion

This work shows that modest neural networks can, under the right assumptions, act as faithful surrogates for classical dynamical codes, enabling orders-of-magnitude speed-ups. Training a general purpose dynamical modelling machine learning model will require expanding training sets to include realistic non-Sérsic photometry and simulation mocks. Depending on the limitations of the available data, different modelling strategies will be appropriate.

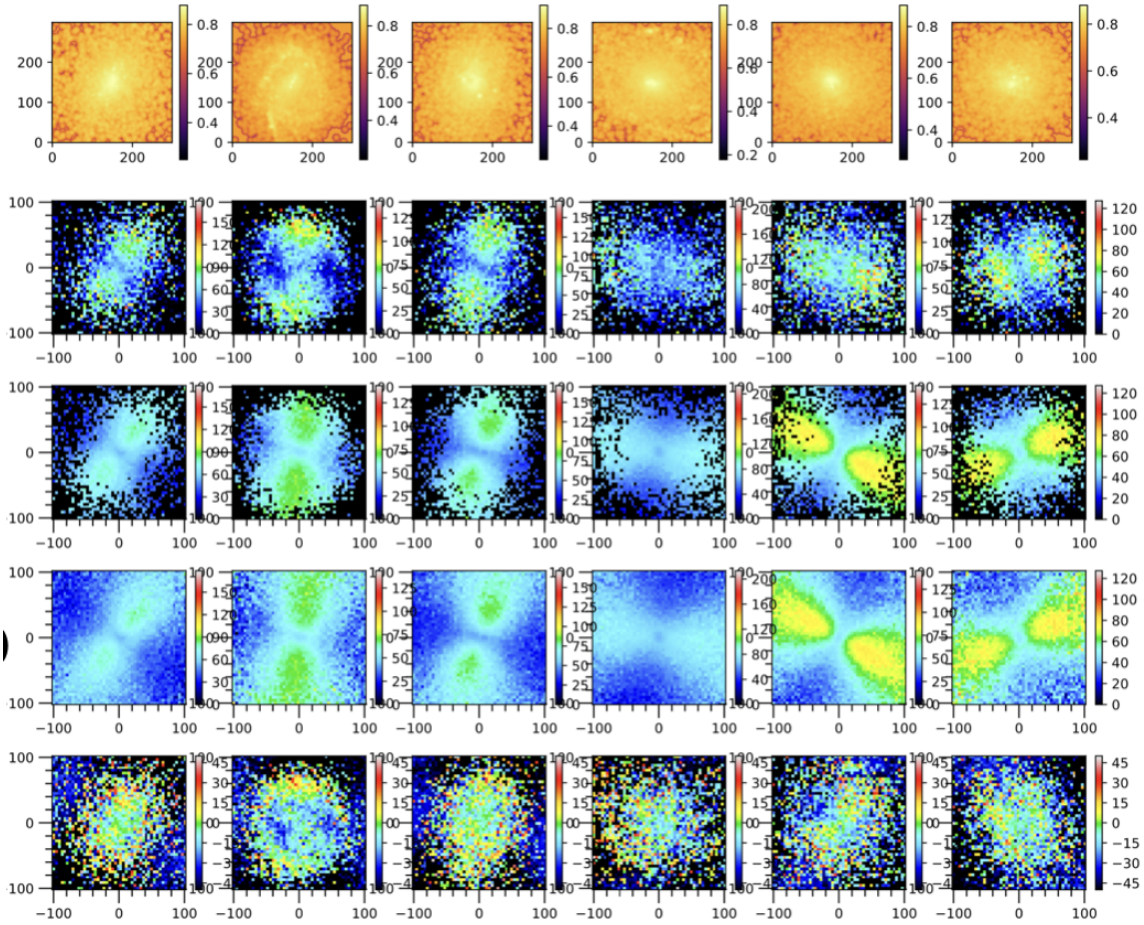

Figure 6.10: This figure shows the performance of our machine learning model for a random selection of test galaxies. The top row shows the mock photometry, the second row shows the true $V_{\rm rms}$ map, the third row shows output of the neural network with a mask applied to match the same mask of the input data, the fourth row shows the true output of the neural network without the mask, and the final row shows the residuals of the model. We see that the major features of the maps, such as their general shape and orientation, are recovered. Important details, such as the absolute magnitude of the maps, however, are not perfect.

Chapter 7

Conclusions

7.1 Summary

This thesis has been primarily focused on integral field spectroscopy and how to best use this tool to better understand the Universe. In chapter 2 we investigated the supermassive black hole mass in M87 using high resolution integral field data from MUSE. We developed a new modification to the JAM code allowing for radially varying anisotropy and found a supermassive black hole mass of $M_{\rm bh} = 8.7 \pm 1.2$ [random] ± 1.3 [systematic])$\times 10^9 M_\odot$. This approach has since been used in several other publications (Nguyen et al. (2025, 2026), Thater et al. in prep., Haitong Wang et al. in prep.) Additionally, we fully exploited our IFS data by measuring both the PSF and stellar distribution from the datacube by treating the additive Legendre polynomials of our pPXF fits as representative of the AGN spectrum in the center of the galaxy. This approach has also now been used by another group in Nguyen et al. (2026).

In chapter 3 we focused on the extraction of stellar kinematics from high a resolution N-body simulation and studied the systematic uncertainties imposed by approximating the stellar line of sight velocity distribution as a partial sum of a Gauss-Hermite series. We find that the impact on the second velocity moment is highly non-trivial, with it converging in an oscillatory way as a function of the number of Gauss-Hermite terms included. These results suggest that not including higher order moments for Jeans modelling will result in an average bias of the gravitational potential of around 10 per-cent. Additionally, I derived closed form expressions for the normalisation and higher order moments of the GH series in section A.1 and section A.2, respectively.

In chapter 4 we produce highly realistic mock datacubes of our simulated galaxy and study the impact of the recovered kinematics. The key development here is the

addition of observed spectra residuals to our mock spectra, while retaining their same spatial information by imposing the binning scheme of a real galaxy on top of our simulation. With this we find that, contrary to the previous chapter, the optimal way to measure the second moment is using a pure gaussian LOSVD. We also see that additive, not multiplicative, Legendre polynomials are useful for improving the quality of the fit and show that the ζ diagram is a useful way to determine what polynomial degree to include. Additionally, we show that assuming template mismatch typical for MaNGA, it is more accurate to measure the second moment using pPXF fits that only fit a Gaussian, not including higher order moments. This is an important result as a number of studies in the past have assumed this to be the case, but it has never been shown in the literature before that this is the optimal choice.

We venture into dynamical models in chapter 5, describing both Jeans and Schwarzschild modeling techniques, and applying them to ideal data from chapter 3. This is a useful contribution to the literature as there have been very few prior studies comparing different Schwarzschild codes and JAM models on exactly the same data. We find that dynamical models tend to fail to recover the supermassive black hole mass when the projected line of sight is within 30° of the intrinsic long axis of the galaxy. Our $\mathrm{JAM_{cyl}}$ models and Schwarzschild models all recover the black hole mass on average within 10 per-cent. $\mathrm{JAM_{sph}}$ recovers it within 20 per-cent. We also compare previous results from the literature and uncover a 40 per-cent offset between JAM and Schwarzschild black hole masses. These results emphasise the importance and complementary nature of using different dynamical modelling approaches to study the same data.

Lastly, in chapter 6 we investigate different machine learning architectures and how well they perform as a replacement for JAM modelling. This is an especially interesting question as machine learning methods have the potential to overcome limitations of traditional dynamical modelling approaches. We find that relatively simple models can perform very well when the training data consists of Sersic galaxies. Galaxies with more realistic photometry, however, struggle to be described using these neural networks. Future work will focus on extending to training sets made from cosmological simulations.

7.2 Future Directions

Following this thesis, there is a large number of potential follow up research questions. Measurements of the supermassive black hole mass in M87 depended on the shape

of the stellar distribution in the center of the galaxy. Follow up measurements of this distribution using high resolution photometry, such as that with JWST, would serve as a strong independent test of the reliability of the pPXF measurement of the central stellar distribution.

We spent some time investigating the systematics associated with using the Gauss-Hermite series parameterisation of the line of sight velocity distribution in chapter 3 and chapter 4. A natural question is how this compares to "non-parametric" methods and if there are other approaches for decreasing the systematic errors. A study comparing GH kinematics with non-parametric kinematics was performed in Reiter et al. (2025). They found less than a 5 per-cent difference in the M/L ratio for a galaxy with a single kinematic component. However, they performed their analysis on real data and thus do not have acccess to the truth and cannot say which approach was more accurate. A follow up study using simulated data for which the truth is available would provide more insight into this situation.

The results in chapter 5 raises some important questions about the reliability of JAM and Schwarzschild modelling. This work focused on the optimistic case where the underlying kinematics are known exactly. For real data this is never the case. Extending these results to more realistic kinematics would provide a better idea of the reliability of dynamical models for real galaxies. Additionally, We find in the literature that there is a 40 per-cent offset between JAM and Schwarzschild models on the same data. Several aspects of this are poorly understood. One is if this trend changes when looking only at axisymmetric or only at triaxial Schwarzschild modelling. Likewise, further breaking the trend down into spherically aligned or cylindrically aligned JAM models would also prove interesting. This is difficult to do now since there is a lack of studies applying different dynamical modelling approaches to the same data. Future studies would help to flesh out these trends and better understand their origins, and how best to mitigate them in the future.

The work in this thesis on machine learning based dynamical modelling has only scratched the surface of the potential for this method to revolutionise dynamical modelling. Our approach showed that relatively small neural networks are already very powerful for describing kinematic data. The future of this field, however, will be using larger models and data from cosmological simulations to directly measure kinematic properties from real galaxies. An interesting case to investigate is MaNGA, for which representative mock observations from simulations have already been produced. A comparison of JAM or Schwarzschild models of mock MaNGA data with machine

learned dynamical models would be a very interesting test and would open the door to using machine learning based dynamical models for more cases.

Appendix A

Gauss-Hermite Series Integrals

A.1 Gauss-Hermite Series Normalisation

The normalisation for the Gauss-Hermite parameterisation for the LOSVD is

$$C = \int_{-\infty}^{\infty} \frac{e^{-y^2/2}}{\sigma\sqrt{2\pi}} \left[1 + \sum_{m=3}^{M} h_m H_m(y)\right] dv \tag{A.1}$$

Note that $dv = \sigma dy$, giving

$$\int_{-\infty}^{\infty} \frac{e^{-y^2/2}}{\sqrt{2\pi}} \left[1 + \sum_{m=3}^{M} h_m H_m(y)\right] dy \tag{A.2}$$

We can use the fact that Hermite functions are eigenfunctions of the Fourier transform (van der Marel & Franx, 1993, Eq. A7)

$$\int_{-\infty}^{\infty} \frac{H_m(y)}{\sqrt{2\pi}} e^{2\pi i f y - y^2/2} dy = i^m H_m(2\pi f) e^{-(2\pi f)^2/2} \tag{A.3}$$

The integral of each term in the sum corresponds to the $f = 0$ case, giving

$$\int_{-\infty}^{\infty} \frac{e^{-y^2/2}}{\sqrt{2\pi}} \left[1 + \sum_{m=3}^{M} h_m H_m(y)\right] dy = 1 + \sum_{m\in\{4,6,\ldots\}}^{M} i^m h_m H_m(0) \tag{A.4}$$

using the fact that

$$H_m(0) = \frac{i^m (m-1)!!}{\sqrt{m!}} \frac{(1 + (-1)^m)}{2} \tag{A.5}$$

gives

$$C = \left(1 + \sum_{m\in\{4,6,\ldots\}}^{M} h_m \frac{(m-1)!!}{\sqrt{m!}}\right) \tag{A.6}$$

We thus write the normalised LOSVD as

$$\mathcal{L}_{\rm norm}(v) = \frac{e^{-y^2/2}}{C\sigma\sqrt{2\pi}} \left[1 + \sum_{m=3}^{M} h_m H_m(y)\right] \tag{A.7}$$

A.2 Analytic LOSVD Moments

The expression for the nth moment of the LOSVD is

$$\langle v^n \rangle = \int_{-\infty}^{\infty} (\sigma y + V)^n \frac{e^{-y^2/2}}{C\sqrt{2\pi}} \left[1 + \sum_{m=3}^{M} h_m H_m(y) \right] dy \tag{A.8}$$

For simplicity, let's take $h_0 = 1, h_1 = 0, h_2 = 0$ so that we can rewrite the above more compactly as

$$\langle v^n \rangle = \int_{-\infty}^{\infty} (\sigma y + V)^n \frac{e^{-y^2/2}}{C\sqrt{2\pi}} \left[\sum_{m=0}^{M} h_m H_m(y) \right] dy \tag{A.9}$$

We can expand the first term as a binomial series and move the sums to get

$$\langle v^n \rangle = \sum_{\ell=0}^{n} \sum_{m=0}^{M} \binom{n}{\ell} h_m \sigma^\ell V^{n-\ell} \int_{-\infty}^{\infty} y^\ell \frac{e^{-y^2/2}}{C\sqrt{2\pi}} H_m(y) dy \tag{A.10}$$

All that remains is to evaluate the integral. This can be done using differentiation under the integral sign of the fourier transform property. Namely, note that

$$\int_{-\infty}^{\infty} y^\ell \frac{e^{-y^2/2}}{\sqrt{2\pi}} H_m(y) dy \tag{A.11}$$

$$= \frac{1}{(2\pi i)^\ell} \frac{d^\ell}{df^\ell} \int_{-\infty}^{\infty} \frac{H_m(y)}{\sqrt{2\pi}} e^{2\pi i f y - y^2/2} dy \Big|_{f=0} \tag{A.12}$$

$$= \frac{\sqrt{2\pi} i^m}{(2\pi i)^\ell} \frac{d^\ell}{df^\ell} H_m(2\pi f) e^{-(2\pi f)^2/2} \Big|_{f=0} \tag{A.13}$$

We can use the general Leibniz rule to get

$$\frac{d^\ell}{df^\ell} \left(H_m(2\pi f) e^{-(2\pi f)^2/2} \right) \Big|_{f=0} \tag{A.14}$$

$$= \sum_{k=0}^{\ell} \binom{\ell}{k} \left(\frac{d^k}{df^k} H_m(2\pi f) \right) \left(\frac{d^{(\ell-k)}}{df^{(\ell-k)}} e^{-(2\pi f)^2/2} \right) \Big|_{f=0} \tag{A.15}$$

For the first term we have

$$\frac{d^k}{df^k} H_m(2\pi f) = (2\pi)^k H_m^{(k)}(2\pi f) \tag{A.16}$$

Evaluating at $f = 0$ and using

$$H_m^{(k)}(y) = \begin{cases} \sqrt{2^k \frac{m!}{(m-k)!}} H_{(m-k)}(y) & k \leq m \\ 0 & k > m \end{cases} \tag{A.17}$$

gives

$$(2\pi)^k \sqrt{2^k \frac{m!}{(m-k)!}} H_{(m-k)}(0) \tag{A.18}$$

$$= (2\pi)^k \sqrt{2^k \frac{m!}{(m-k)!}} \frac{i^{(m-k)}(m-k-1)!!}{\sqrt{(m-k)!}} \frac{(1+(-1)^{m-k})}{2} \tag{A.19}$$

$$= (2\pi)^k \sqrt{2^k m!} \frac{i^{(m-k)}(m-k-1)!!}{(m-k)!} \frac{(1+(-1)^{m-k})}{2} \tag{A.20}$$

$$= \begin{cases} (2\pi)^k \sqrt{2^k m!} \frac{i^{(m-k)}(m-k-1)!!}{(m-k)!} & (m-k) \geq 0 \text{ and even} \\ 0 & \text{else} \end{cases} \tag{A.21}$$

and for the second term, if we let $u = \sqrt{2}\pi f$ and $\ell - k = n$

$$\frac{d^{(\ell-k)}}{df^{(\ell-k)}} e^{-(2\pi f)^2/2} = (\sqrt{2}\pi)^n \frac{d^n}{df^n} e^{-u^2} \tag{A.22}$$

$$= (-\sqrt{2}\pi)^n e^{-u^2} H_n^{\text{phys}}(u) \tag{A.23}$$

$$= (-\sqrt{2}\pi)^{(\ell-k)} e^{-(2\pi f)^2/2} H_{\ell-k}^{\text{phys}}(\sqrt{2}\pi f) \tag{A.24}$$

$$= (-\sqrt{2}\pi)^{(\ell-k)} e^{-(2\pi f)^2/2} \sqrt{(\ell-k)! 2^{(\ell-k)}} H_{\ell-k}(\sqrt{2}\pi f) \tag{A.25}$$

evaluating at $f = 0$ gives

$$(-\sqrt{2}\pi)^{(\ell-k)} \sqrt{(\ell-k)! 2^{(\ell-k)}} H_{\ell-k}(0) \tag{A.26}$$

$$= (-\sqrt{2}\pi)^{(\ell-k)} \sqrt{(\ell-k)! 2^{(\ell-k)}} i^{(\ell-k)} \frac{(\ell-k-1)!!}{\sqrt{(\ell-k)!}} \frac{(1+(-1)^{\ell-k})}{2} \tag{A.27}$$

$$= (-2\pi)^{(\ell-k)} i^{(\ell-k)} (\ell-k-1)!! \frac{(1+(-1)^{\ell-k})}{2} \tag{A.28}$$

$$= \begin{cases} (-2\pi i)^{(\ell-k)} (\ell-k-1)!! & (\ell-k) \geq 0 \text{ and even} \\ 0 & \text{else} \end{cases} \tag{A.29}$$

Putting it all together gives

$$\frac{\sqrt{2\pi} i^n}{(2\pi i)^\ell} \sum_{k=0}^{\ell} \binom{\ell}{k} (2\pi)^k \sqrt{2^k n!} \frac{i^{(n-k)}(n-k-1)!!}{(n-k)!} (-2\pi)^{(\ell-k)} i^{(\ell-k)} (\ell-k-1)!! \tag{A.30}$$

$$= \frac{\sqrt{2\pi} i^n}{(2\pi i)^\ell} \sum_{k=0}^{\ell} \binom{\ell}{k} (2\pi)^k \sqrt{2^k n!} \frac{(n-k-1)!!}{(n-k)!} (-2\pi)^{(\ell-k)} i^{(n+\ell-2k)} (\ell-k-1)!! \tag{A.31}$$

$$= \frac{\sqrt{2\pi} i^n}{(2\pi i)^\ell} \sum_{k=0}^{\ell} \binom{\ell}{k} \sqrt{2^k n!} \frac{(n-k-1)!!}{(n-k)!} (-2\pi)^{\ell} i^{(n+\ell)} (\ell-k-1)!! \tag{A.32}$$

$$= \sqrt{2\pi} \sum_{k=0}^{\ell} \binom{\ell}{k} \sqrt{2^k n!} \frac{(n-k-1)!!}{(n-k)!} (-1)^{\ell+n} (\ell-k-1)!! \tag{A.33}$$

Combing these results, we see that the general expression for the nth moment is

$$\langle v^n \rangle = \frac{1}{C} \sum_{\ell=0}^{n} \sum_{m=0}^{M} \sum_{k=0}^{\ell} h_m \sigma^\ell V^{n-\ell} A_{k\ell mn} \tag{A.34}$$

with

$$A_{k\ell mn} = \frac{n! \sqrt{2^k m!}}{k!(n-\ell)!(m-k)!!(\ell-k)!!} \tag{A.35}$$

where $m - \ell, \ell - k \geq 0$ and $m - k$,$\ell - k$ are even, else $A_{k\ell mn}$ returns zero.

Bibliography

Agnello A., Evans N. W., Romanowsky A. J., Brodie J. P., 2014, MNRAS, 442, 3299

Ahn C. P., et al., 2018, ApJ, 858, 102

Al-Amri R. M., Walsh J. L., Liepold E. R., Ma C.-P., Greene J. E., 2025, arXiv e-prints, p. arXiv:2510.02439

An J. H., Evans N. W., 2006, ApJ, 642, 752

Arsenault R., et al., 2008, in Hubin N., Max C. E., Wizinowich P. L., eds, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series Vol. 7015, Adaptive Optics Systems. p. 701524

Astropy Collaboration et al., 2013, A&A, 558, A33

Astropy Collaboration et al., 2018, AJ, 156, 123

Babcock H. W., 1939, Lick Observatory Bulletin, 498, 41

Bacon R., 1985, A&A, 143, 84

Bacon R., 2024, Ap&SS, 369, 111

Bacon R., Simien F., Monnet G., 1983, A&A, 128, 405

Bacon R., et al., 1988, in Ulrich M. H., ed., European Southern Observatory Conference and Workshop Proceedings Vol. 30, Very Large Telescopes and their Instrumentation, Volume 2. p. 1185

Bacon R., Emsellem E., Monnet G., Nieto J. L., 1994, A&A, 281, 691

Bacon R., et al., 1995, A&AS, 113, 347

Bacon R., et al., 2001, MNRAS, 326, 23

Bacon R., et al., 2010, in McLean I. S., Ramsay S. K., Takami H., eds, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series Vol. 7735, Ground-based and Airborne Instrumentation for Astronomy III. p. 773508

Baes M., van Hese E., 2007, A&A, 471, 419

Bartusiak M., 2010, The day we found the universe. Vintage

Bender R., 1988, A&A, 193, L7

Bender R., 1990, A&A, 229, 441

Bender R., Doebereiner S., Moellenhoff C., 1988, A&AS, 74, 385

Bertola F., Capaccioli M., 1975, ApJ, 200, 439

Binney J., 1976, MNRAS, 177, 19

Binney J., 1978, MNRAS, 183, 501

Binney J., 1982, ARA&A, 20, 399

Binney J., Mamon G. A., 1982, MNRAS, 200, 361

Binney J., Tremaine S., 2011, Galactic dynamics. Princeton university press

Bottema R., 1988, A&A, 197, 105

Boulesteix J., Georgelin Y., Marcelin M., Monnet G., 1984, in Boksenberg A., Crawford D. L., eds, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series Vol. 445, Instrumentation in astronomy V. pp 37–41, doi:10.1117/12.966123

Bronkalla W., Oleak H., 1987, Publications of the Astronomical Institute of the Czechoslovak Academy of Sciences, 69, 449

Bruzual G., Charlot S., 2003, MNRAS, 344, 1000

Bufano F., Sciacca E., Riggi S., eds, 2026, Machine Learning for Astrophysics 2024: Proceedings of the 2nd ML4ASTRO International Conference 8-12 July 2024. Astrophysics and Space Science Proceedings Vol. 62, Springer Cham

Cappellari M., 2002, MNRAS, 333, 400

Cappellari M., 2008, MNRAS, 390, 71

Cappellari M., 2016, ARA&A, 54, 597

Cappellari M., 2017, MNRAS, 466, 798

Cappellari M., 2020, MNRAS, 494, 4819

Cappellari M., 2023, MNRAS, 526, 3273

Cappellari M., 2026, in Encyclopedia of Astrophysics, Volume 4. pp 122–152 (arXiv:2503.02746), doi:10.1016/B978-0-443-21439-4.00109-7

Cappellari M., Copin Y., 2003, MNRAS, 342, 345

Cappellari M., Emsellem E., 2004, PASP, 116, 138

Cappellari M., McDermid R. M., 2005, Classical and Quantum Gravity, 22, S347

Cappellari M., Verolme E. K., van der Marel R. P., Verdoes Kleijn G. A., Illingworth G. D., Franx M., Carollo C. M., de Zeeuw P. T., 2002, ApJ, 578, 787

Cappellari M., et al., 2006, MNRAS, 366, 1126

Cappellari M., et al., 2007, MNRAS, 379, 418

Cappellari M., et al., 2008, in Bureau M., Athanassoula E., Barbuy B., eds, Vol. 245, Formation and Evolution of Galaxy Bulges. pp 215–218 (arXiv:0709.2861), doi:10.1017/S1743921308017687

Cappellari M., Neumayer N., Reunanen J., van der Werf P. P., de Zeeuw P. T., Rix H. W., 2009, MNRAS, 394, 660

Cappellari M., et al., 2010, in Debattista V. P., Popescu C. C., eds, American Institute of Physics Conference Series Vol. 1240, Hunting for the Dark: the Hidden Side of Galaxy Formation. pp 211–214 (arXiv:1001.3233), doi:10.1063/1.3458489

Cappellari M., et al., 2011a, MNRAS, 413, 813

Cappellari M., et al., 2011b, MNRAS, 416, 1680

Cappellari M., et al., 2012, Nature, 484, 485

Cappellari M., et al., 2013, MNRAS, 432, 1709

Cappellari M., et al., 2015, ApJ, 804, L21

Cenarro A. J., Cardiel N., Gorgas J., Peletier R. F., Vazdekis A., Prada F., 2001, MNRAS, 326, 959

Cherinka B., et al., 2019, AJ, 158, 74

Chu J., Tang H., Xu D., Lu S., Long R., 2024, MNRAS, 528, 6354

Cid Fernandes R., Mateus A., Sodré L., Stasińska G., Gomes J. M., 2005, MNRAS, 358, 363

Conroy C., Gunn J. E., 2010, ApJ, 712, 833

Conroy C., Gunn J. E., White M., 2009, ApJ, 699, 486

Côté P., et al., 2004, ApJS, 153, 223

Courtes G., Georgelin Y., Monnet R. B. G., Boulesteix J., 1988, in Robinson L. B., ed., Instrumentation for Ground-Based Optical Astronomy. p. 266

Courtès G., 1980, Application de la Photométrie Bidimensionelle à l'Astronomie

Cretton N., de Zeeuw P. T., van der Marel R. P., Rix H.-W., 1999, ApJS, 124, 383

Curtis H., Campbell W., Moore J., Wilson R., Wright W., 1918, Studies of the Nebulae: Made at the Lick Observatory, University of California, at Mount Hamilton, California, and Santiago, Chile. No. v. 13 in Publications of the Lick Observatory of the University of California, University of California Press, `https://books.google.co.uk/books?id=HYnnAAAAMAAJ`

Davidson R., Flachaire E., 2008, Journal of Econometrics, 146, 162

Davies R. L., Illingworth G., 1983, ApJ, 266, 516

Davies R. L., Efstathiou G., Fall S. M., Illingworth G., Schechter P. L., 1983, ApJ, 266, 41

Dolag K., Borgani S., Murante G., Springel V., 2009, MNRAS, 399, 497

Dominiak P., et al., 2025, MNRAS, 542, 2039

Drehmer D. A., Storchi-Bergmann T., Ferrari F., Cappellari M., Riffel R. A., 2015, MNRAS, 450, 128

Emsellem E., 2013, MNRAS, 433, 1862

Emsellem E., Monnet G., Bacon R., 1994, A&A, 285, 723

Emsellem E., et al., 2004, MNRAS, 352, 721

Emsellem E., et al., 2007, MNRAS, 379, 401

Emsellem E., et al., 2011, MNRAS, 414, 888

Emsellem E., Krajnovic D., Sarzi M., 2014, MNRAS, 445, L79

Evans N. W., Sanders J. L., Williams A. A., An J., Lynden-Bell D., Dehnen W., 2016, MNRAS, 456, 4506

Event Horizon Telescope Collaboration et al., 2019, ApJ, 875, L6

Faber S. M., Jackson R. E., 1976, ApJ, 204, 668

Falcón-Barroso J., Martig M., 2021, A&A, 646, A31

Falcón-Barroso J., Sánchez-Blázquez P., Vazdekis A., Ricciardelli E., Cardiel N., Cenarro A. J., Gorgas J., Peletier R. F., 2011, A&A, 532, A95

Feldmeier-Krause A., Zhu L., Neumayer N., van de Ven G., de Zeeuw P. T., Schödel R., 2017, MNRAS, 466, 4040

Ferrarese L., Merritt D., 2000, ApJ, 539, L9

Ferrarese L., van den Bosch F. C., Ford H. C., Jaffe W., O'Connell R. W., 1994, AJ, 108, 1598

Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, PASP, 125, 306

Franx M., Illingworth G. D., 1988, ApJ, 327, L55

Freeman K., 2013, arXiv preprint arXiv:1301.7509

Freudling W., Romaniello M., Bramich D. M., Ballester P., Forchi V., García-Dabló C. E., Moehler S., Neeser M. J., 2013, A&A, 559, A96

Frigo M., Naab T., Rantala A., Johansson P. H., Neureiter B., Thomas J., Rizzuto F., 2021, MNRAS, 508, 4610

Fu Y., Cappellari M., Zhu K., Mao S., Lu S., 2025, MNRAS, 544, 1038

Gebhardt K., Thomas J., 2009, ApJ, 700, 1690

Gebhardt K., et al., 2000a, AJ, 119, 1157

Gebhardt K., et al., 2000b, ApJ, 539, L13

Gebhardt K., et al., 2003, ApJ, 583, 92

Gebhardt K., Adams J., Richstone D., Lauer T. R., Faber S. M., Gültekin K., Murphy J., Tremaine S., 2011, ApJ, 729, 119

Gerhard O. E., 1993, MNRAS, 265, 213

Gerhard O., 1996, in Minniti D., Rix H.-W., eds, Spiral Galaxies in the Near-IR. p. 138, doi:10.1007/978-3-540-49739-4˙20

Gerhard O. E., Binney J. J., 1996, MNRAS, 279, 993

Glorot X., Bengio Y., 2010, in Proceedings of the thirteenth international conference on artificial intelligence and statistics. pp 249–256

Gomer M. R., et al., 2023, A&A, 679, A59

Gontcharov G. A., Khovritchev M. Y., Mosenkov A. V., 2020, MNRAS, 497, 3674

González Á., 2010, Mathematical geosciences, 42, 49

Gravity Collaboration et al., 2023, A&A, 677, L10

Harborne K. E., del P Lagos C., Croom S. M., van de Sande J., Ludlow A., Remus R. S., Kimmig L. C., Power C., 2024, MNRAS,

Harms R. J., et al., 1994, ApJ, 435, L35

Harris C. R., et al., 2020, Nature, 585, 357

Hau G., 2017, SINFONI User Manual. European Southern Observatory, `https://www.eso.org/sci/facilities/paranal/decommissioned/sinfoni/doc/VLT-MAN-ESO-14700-3517_v101.0.pdf`

He K., Zhang X., Ren S., Sun J., 2015, in Proceedings of the IEEE international conference on computer vision. pp 1026–1034

Head T., Kumar M., Nahrstaedt H., Louppe G., Shcherbatyi I., 2021, scikit-optimize/scikit-optimize v0.9.0, doi:10.5281/zenodo.5565057, https://doi.org/10.5281/zenodo.5565057

Hernquist L., 1990, ApJ, 356, 359

Herrnstein J. R., et al., 1999, Nature, 400, 539

Herschel W., 1786, Philosophical Transactions of the Royal Society of London, pp 457–499

Hubble E. P., 1922, ApJ, 56, 162

Hubble E. P., 1925, Popular Astronomy, 33, 252

Hubble E. P., 1926a, ApJ, 63, 236

Hubble E. P., 1926b, ApJ, 64, 321

Hubble E., 1929a, Proceedings of the National Academy of Science, 15, 168

Hubble E. P., 1929b, ApJ, 69, 103

Hubble E. P., 1936, Realm of the Nebulae

Hunter J. D., 2007, Computing in Science and Engineering, 9, 90

Hunter C., de Zeeuw P. T., 1992, ApJ, 389, 79

Illingworth G., 1977, ApJ, 218, L43

Ivanov V. D., Coccato L., Neeser M. J., Selman F., Pizzella A., Dalla Bontà E., Corsini E. M., Morelli L., 2019, A&A, 629, A100

Jeans J. H., 1922, MNRAS, 82, 122

Jeans J. H., 1928, Astronomy and cosmogony, doi:10.1017/CBO9780511694363.

Jeter B., Broderick A. E., 2021, ApJ, 908, 139

Jeter B., Broderick A. E., McNamara B. R., 2019, ApJ, 882, 82

Jethwa P., Thater S., Maindl T., Van de Ven G., 2020, DYNAMITE: DYnamics, Age and Metallicity Indicators Tracing Evolution, Astrophysics Source Code Library, record ascl:2011.007 (ascl:2011.007)

Jin Y., Zhu L., Long R. J., Mao S., Xu D., Li H., van de Ven G., 2019, MNRAS, 486, 4753

Jin Y., Zhu L., Long R. J., Mao S., Wang L., van de Ven G., 2020, MNRAS, 491, 1690

Jones D. R., Schonlau M., Welch W. J., 1998, Journal of Global optimization, 13, 455

Kelson D. D., Illingworth G. D., van Dokkum P. G., Franx M., 2000, ApJ, 531, 159

Kluyver T., et al., 2016, in , IOS Press. pp 87–90, doi:10.3233/978-1-61499-649-1-87

Knowles A. T., Sansom A. E., Vazdekis A., Allende Prieto C., 2023, MNRAS, 523, 3450

Kochanek C. S., Rybicki G. B., 1996, MNRAS, 280, 1257

Kormendy J., 1982, ApJ, 257, 75

Kormendy J., Bender R., 1996, ApJ, 464, L119

Kormendy J., Ho L. C., 2013, ARA&A, 51, 511

Kormendy J., Illingworth G., 1982, ApJ, 256, 460

Kormendy J., Richstone D., 1995, ARA&A, 33, 581

Kormendy J., Fisher D. B., Cornell M. E., Bender R., 2009, ApJS, 182, 216

Krajnović D., Cappellari M., de Zeeuw P. T., Copin Y., 2006, MNRAS, 366, 787

Krajnović D., McDermid R. M., Cappellari M., Davies R. L., 2009, MNRAS, 399, 1839

Krajnović D., et al., 2011, MNRAS, 414, 2923

Krajnović D., et al., 2018, MNRAS, 477, 3030

Krist J., Hook R., Stoehr F., 2010, Astrophysics Source Code Library, record ascl:1010.057

Kuijken K., Merrifield M. R., 1993, MNRAS, 264, 712

Kuntschner H., et al., 2010, MNRAS, 408, 97

Lablanche P.-Y., et al., 2012, MNRAS, 424, 1495

Lacerda E. A. D., Sánchez S. F., Mejía-Narváez A., Camps-Fariña A., Espinosa-Ponce C., Barrera-Ballesteros J. K., Ibarra-Medel H., Lugo-Aranda A. Z., 2022, New Astron., 97, 101895

Lauer T. R., et al., 1992, AJ, 103, 703

Lauer T. R., et al., 1995a, AJ, 110, 2622

Lauer T. R., et al., 1995b, AJ, 110, 2622

Lauer T. R., et al., 2007, ApJ, 664, 226

Lawson C. L., Hanson R. J., 1995, Solving least squares problems. SIAM

Leavitt H. S., 1907, Annals of Harvard College Observatory, 60, 87

Leavitt H. S., Pickering E. C., 1912, Harvard College Observatory Circular, 173, 1

Leung G. Y. C., et al., 2018, MNRAS, 477, 254

Li H., Li R., Mao S., Xu D., Long R. J., Emsellem E., 2016, MNRAS, 455, 3680

Li H., et al., 2018a, MNRAS, 476, 1765

Li H., Mao S., Cappellari M., Graham M. T., Emsellem E., Long R. J., 2018b, ApJ, 863, L19

Li C., et al., 2020, MNRAS, 492, 2775

Liepold E. R., Quenneville M. E., Ma C.-P., Walsh J. L., McConnell N. J., Greene J. E., Blakeslee J. P., 2020, ApJ, 891, 4

Liepold E. R., Ma C.-P., Walsh J. L., 2023a, ApJ, 945, L35

Liepold E. R., Ma C.-P., Walsh J. L., 2023b, ApJ, 945, L35

Light E. S., Danielson R. E., Schwarzschild M., 1974, ApJ, 194, 257

Lundmark K., 1925, MNRAS, 85, 865

Lundmark K., 1926, Arkiv for Matematik, Astronomi och Fysik, 19, 1

Maas A. L., Hannun A. Y., Ng A. Y., et al., 2013, in Proc. icml. p. 3

Macchetto F., Marconi A., Axon D. J., Capetti A., Sparks W., Crane P., 1997, ApJ, 489, 579

Magorrian J., 1999, MNRAS, 302, 530

Magorrian J., et al., 1998, AJ, 115, 2285

Mamon G. A., Łokas E. L., 2005, MNRAS, 363, 705

Marinacci F., et al., 2018, MNRAS, 480, 5113

Mayall N. U., 1951, Publications of Michigan Observatory, 10, 19

McConnell N. J., Ma C.-P., Murphy J. D., Gebhardt K., Lauer T. R., Graham J. R., Wright S. A., Richstone D. O., 2012, ApJ, 756, 179

McDermid R. M., et al., 2006, MNRAS, 373, 906

Mehrgan K., Thomas J., Saglia R., Mazzalay X., Erwin P., Bender R., Kluge M., Fabricius M., 2019, ApJ, 887, 195

Mehrgan K., Thomas J., Saglia R., Parikh T., Bender R., 2023, ApJ, 948, 79

Merritt D., 1985, MNRAS, 214, 25P

Merritt D., 1997, AJ, 114, 228

Milosavljević M., Merritt D., 2001, ApJ, 563, 34

Milosavljević M., Merritt D., Rest A., van den Bosch F. C., 2002, MNRAS, 331, L51

Mitzkus M., Cappellari M., Walcher C. J., 2017, MNRAS, 464, 4789

Monnet G., Bacon R., Emsellem E., 1992, A&A, 253, 366

Murphy J. D., Gebhardt K., Adams J. J., 2011, ApJ, 729, 129

Naiman J. P., et al., 2018, MNRAS, 477, 1206

Navarro J. F., Frenk C. S., White S. D. M., 1997, ApJ, 490, 493

Nelson D., et al., 2018, MNRAS, 475, 624

Neumayer N., Cappellari M., Reunanen J., Rix H. W., van der Werf P. P., de Zeeuw P. T., Davies R. I., 2007, ApJ, 671, 1329

Neureiter B., et al., 2021, MNRAS, 500, 1437

Neureiter B., de Nicola S., Thomas J., Saglia R., Bender R., Rantala A., 2023, MNRAS, 519, 2004

Nguyen D. D., et al., 2025, arXiv e-prints, p. arXiv:2509.20519

Nguyen D. D., et al., 2026, arXiv e-prints, p. arXiv:2601.17439

Nitschai M. S., Cappellari M., Neumayer N., 2020, MNRAS, 494, 6001

Nitschai M. S., Eilers A.-C., Neumayer N., Cappellari M., Rix H.-W., 2021, ApJ, 916, 112

Oldham L. J., Auger M. W., 2016, MNRAS, 457, 421

Oldham L., Auger M., 2018, MNRAS, 474, 4169

Osipkov L. P., 1979, Pisma v Astronomicheskii Zhurnal, 5, 77

Osorno J., Nagar N., Richtler T., Humire P., Gebhardt K., Gultekin K., 2023, arXiv e-prints, p. arXiv:2304.11264

Osorno J., et al., 2025, A&A, 695, A72

Parikh T., Saglia R., Thomas J., Mehrgan K., Bender R., Maraston C., 2024, MNRAS, 528, 7338

Paszke A., et al., 2019, Advances in neural information processing systems, 32

Pease F. G., 1916, Proceedings of the National Academy of Science, 2, 517

Pease F. G., 1918, Proceedings of the National Academy of Science, 4, 21

Pedregosa F., et al., 2011, Journal of Machine Learning Research, 12, 2825

Pérez F., Granger B. E., 2007, Computing in Science and Engineering, 9, 21

Pillepich A., et al., 2018, MNRAS, 475, 648

Press W. H., 2007, Numerical recipes 3rd edition: The art of scientific computing. Cambridge university press

Price-Whelan A. M., Foreman-Mackey D., 2017, The Journal of Open Source Software, 2

Python Software Foundation 2024, Python, https://www.python.org

Qian E. E., de Zeeuw P. T., van der Marel R. P., Hunter C., 1995, MNRAS, 274, 602

Quenneville M. E., Liepold E. R., Ma C.-P., 2021, ApJS, 254, 25

Quenneville M. E., Liepold E. R., Ma C.-P., 2022, ApJ, 926, 30

Rantala A., Pihajoki P., Johansson P. H., Naab T., Lahén N., Sawala T., 2017, ApJ, 840, 53

Rantala A., Johansson P. H., Naab T., Thomas J., Frigo M., 2018, ApJ, 864, 113

Rantala A., Johansson P. H., Naab T., Thomas J., Frigo M., 2019, ApJ, 872, L17

Reiter S., Jethwa P., van de Ven G., Thater S., Leaman R., 2025, A&A, 701, A12

Reynolds J. H., 1920, MNRAS, 80, 746

Richstone D. O., 1980, ApJ, 238, 103

Richstone D. O., Tremaine S., 1984, ApJ, 286, 27

Richstone D. O., Tremaine S., 1988, ApJ, 327, 82

Rix H.-W., White S. D. M., 1990, ApJ, 362, 52

Rix H.-W., White S. D. M., 1992, MNRAS, 254, 389

Rix H.-W., de Zeeuw P. T., Cretton N., van der Marel R. P., Carollo C. M., 1997, ApJ, 488, 702

Rix H.-W., Carollo C. M., Freeman K., 1999, ApJ, 513, L25

Roberts C. A., Bentz M. C., Vasiliev E., Valluri M., Onken C. A., 2021, ApJ, 916, 25

Romanowsky A. J., Kochanek C. S., 1997, MNRAS, 287, 35

Rubin V. C., 1965, ApJ, 142, 934

Rubin V. C., 1995, ApJ, 451, 419

Rubin V. C., Ford Jr. W. K., 1970, ApJ, 159, 379

Rubin V. C., Burley J., Kiasatpoor A., Klock B., Pease G., Rutscheidt E., Smith C., 1962, AJ, 67, 491

Rubin V. C., Ford Jr. W. K., Thonnard N., 1978, ApJ, 225, L107

Rusli S. P., et al., 2013, AJ, 146, 45

Rybicki G. B., 1987, in de Zeeuw P. T., ed., IAU Symposium Vol. 127, Structure and Dynamics of Elliptical Galaxies. p. 397, doi:10.1007/978-94-009-3971-4˙41

Saha P., Williams T. B., 1994, AJ, 107, 1295

Sánchez-Blázquez P., et al., 2006, MNRAS, 371, 703

Sánchez S. F., et al., 2016, Rev. Mex. Astron. Astrofis., 52, 171

Sandage A., 2005, ARA&A, 43, 581

Santucci G., et al., 2022, ApJ, 930, 153

Sargent W. L. W., Schechter P. L., Boksenberg A., Shortridge K., 1977, ApJ, 212, 326

Sarzi M., Spiniello C., La Barbera F., Krajnović D., van den Bosch R., 2018, MNRAS, 478, 4084

Schechter P. L., Gunn J. E., 1979, ApJ, 229, 472

Scheiner J., 1899, ApJ, 9, 149

Schlafly E. F., Finkbeiner D. P., 2011, ApJ, 737, 103

Schwarzschild M., 1979, ApJ, 232, 236

Scott N., et al., 2015, MNRAS, 451, 2723

Sérsic J. L., 1963, Boletin de la Asociacion Argentina de Astronomia La Plata Argentina, 6, 41

Seth A. C., et al., 2014, Nature, 513, 398

Shapiro K. L., Cappellari M., de Zeeuw T., McDermid R. M., Gebhardt K., van den Bosch R. C. E., Statler T. S., 2006, MNRAS, 370, 559

Shapley H., 1918, ApJ, 48, 154

Silge J. D., Gebhardt K., Bergmann M., Richstone D., 2005, AJ, 130, 406

Simkin S. M., 1974, A&A, 31, 129

Simon D. A., Cappellari M., Hartke J., 2024, MNRAS, 527, 2341

Siopis C., et al., 2009, ApJ, 693, 946

Slipher V. M., 1913, Lowell Observatory Bulletin, 2, 56

Slipher V. M., 1914, Lowell Observatory Bulletin, 2, 66

Slipher V. M., 1917, Proceedings of the American Philosophical Society, 56, 403

Soto K. T., Lilly S. J., Bacon R., Richard J., Conseil S., 2016, MNRAS, 458, 3210

Springel V., 2005, MNRAS, 364, 1105

Springel V., White S. D. M., Tormen G., Kauffmann G., 2001, MNRAS, 328, 726

Springel V., et al., 2018, MNRAS, 475, 676

Statler T., 1995, AJ, 109, 1371

Ströbele S., et al., 2012, in Ellerbroek B. L., Marchetti E., Véran J.-P., eds, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series Vol. 8447, Adaptive Optics Systems III. p. 844737, doi:10.1117/12.926110

Tahmasebzadeh B., Zhu L., Shen J., Gerhard O., van de Ven G., 2022, ApJ, 941, 109

Tahmasebzadeh B., et al., 2024, MNRAS, 534, 861

Thater S., et al., 2017, A&A, 597, A18

Thater S., Krajnović D., Cappellari M., Davis T. A., de Zeeuw P. T., McDermid R. M., Sarzi M., 2019, A&A, 625, A62

Thater S., et al., 2022a, MNRAS, 509, 5416

Thater S., et al., 2022b, A&A, 667, A51

The New York Times 1924, The New York Times

Thomas J., Saglia R. P., Bender R., Thomas D., Gebhardt K., Magorrian J., Corsini E. M., Wegner G., 2005, MNRAS, 360, 1355

Thomas J., Jesseit R., Naab T., Saglia R. P., Burkert A., Bender R., 2007, MNRAS, 381, 1672

Thomas J., Saglia R. P., Bender R., Erwin P., Fabricius M., 2014, ApJ, 782, 39

Thomas J., Ma C.-P., McConnell N. J., Greene J. E., Blakeslee J. P., Janish R., 2016, Nature, 532, 340

Tonry J., Davis M., 1979, AJ, 84, 1511

Valluri M., Merritt D., Emsellem E., 2004, ApJ, 602, 66

Vanderriest C., 1980, PASP, 92, 858

Vasiliev E., 2013, MNRAS, 434, 3174

Vasiliev E., Valluri M., 2020, ApJ, 889, 39

Vazdekis A., Sánchez-Blázquez P., Falcón-Barroso J., Cenarro A. J., Beasley M. A., Cardiel N., Gorgas J., Peletier R. F., 2010, MNRAS, 404, 1639

Vazdekis A., et al., 2015, MNRAS, 449, 1177

Vazdekis A., Koleva M., Ricciardelli E., Röck B., Falcón-Barroso J., 2016, MNRAS, 463, 3409

Verro K., et al., 2022, A&A, 660, A34

Virtanen P., et al., 2020, Nature Methods, 17, 261

Vlasov A., 1938

Walsh J. L., Barth A. J., Ho L. C., Sarzi M., 2013, ApJ, 770, 86

Walsh J. L., van den Bosch R. C. E., Gebhardt K., Yildirim A., Gültekin K., Husemann B., Richstone D. O., 2015, ApJ, 808, 183

Walsh J. L., van den Bosch R. C. E., Gebhardt K., Yıldırım A., Richstone D. O., Gültekin K., Husemann B., 2016, ApJ, 817, 2

Wang B., Cappellari M., Peng Y., 2021, MNRAS, 500, L27

Way M. J., 2013, arXiv preprint arXiv:1301.7294

Weilbacher P. M., et al., 2020, A&A, 641, A28

Westfall K. B., et al., 2019, The Astronomical Journal, 158, 231

Wilkinson A., Sharples R. M., Fosbury R. A. E., Wallace P. T., 1986, MNRAS, 218, 297

Wilkinson D. M., Maraston C., Goddard D., Thomas D., Parikh T., 2017, MNRAS, 472, 4297

Willmer C. N. A., 2018, ApJS, 236, 47

Wolfram Research Inc. 2025, Mathematica, Version 14.3, `https://www.wolfram.com/mathematica`

Woo J., Walters D., Archinuk F., Faber S. M., Ellison S. L., Teimoorinia H., Iyer K., 2024, MNRAS, 530, 4260

Wyithe J. S. B., Turner E. L., Spergel D. N., 2001, ApJ, 555, 504

Xu D., Springel V., Sluse D., Schneider P., Sonnenfeld A., Nelson D., Vogelsberger M., Hernquist L., 2017, MNRAS, 469, 1824

Zettl A., 2005, Sturm-liouville theory. No. 121, American Mathematical Soc.

Zhao H., 1996, MNRAS, 278, 488

Zhu L., et al., 2018, Nature Astronomy, 2, 233

Zhu K., Lu S., Cappellari M., Li R., Mao S., Gao L., 2023, MNRAS, 522, 6326

de Nicola S., Neureiter B., Thomas J., Saglia R. P., Bender R., 2022, MNRAS, 517, 3445

de Nicola S., Thomas J., Saglia R. P., Kluge M., Snigula J., Bender R., 2025, arXiv e-prints, p. arXiv:2512.04178

de Zeeuw P. T., Evans N. W., Schwarzschild M., 1996, MNRAS, 280, 903

de Zeeuw P. T., et al., 2002, MNRAS, 329, 513

van Houdt J., et al., 2021, ApJ, 923, 11

van Maanen A., 1916, ApJ, 44, 210

van Maanen A., 1935, ApJ, 81, 336

van de Sande J., et al., 2021, MNRAS, 505, 3078

van den Bergh S., 1976, ApJ, 206, 883

van den Bosch F. C., 1997, MNRAS, 287, 543

van den Bosch R. C. E., de Zeeuw P. T., 2010, MNRAS, 401, 1770

van den Bosch R. C. E., van de Ven G., Verolme E. K., Cappellari M., de Zeeuw P. T., 2008, MNRAS, 385, 647

van den Bosch R. C. E., Gebhardt K., Gültekin K., van de Ven G., van der Wel A., Walsh J. L., 2012, Nature, 491, 729

van der Marel R. P., 1991, MNRAS, 253, 710

van der Marel R. P., 1994, MNRAS, 270, 271

van der Marel R. P., Franx M., 1993, ApJ, 407, 525

van der Marel R. P., Cretton N., de Zeeuw P. T., Rix H.-W., 1998, ApJ, 493, 613