Exploring the use of Machine Learning with extragalactic emission-line surveys, in preparation for the Square Kilometre Array.

by

James Dawson

A Thesis submitted to Cardiff University
for the degree of Doctor of Philosophy

01 October 2021

"Decide in your heart of hearts what really excites and challenges you, and start moving your life in that direction. Every decision you make, from what you eat to what you do with your time tonight, turns you into who you are tomorrow, and the day after that. Look at who you want to be, and start sculpting yourself into that person. You may not get exactly where you thought you'd be, but you will be doing things that suit you in a profession you believe in. Don't let life randomly kick you into the adult you don't want to become."

Chris Hadfield
NASA Astronaut

Acknowledgements

When I started my PhD I had no idea of the impending roller-coaster of people and experiences that, when looking back, seemed to bombard me throughout my 4 years. The funny thing is, the more people I've met along the way and who have been integral to my PhD experience, the more people have entered my life... like some strange Hydra of enrichment. The PhD has been great academically, but it's been priceless in terms of getting to know the myriad of people who've helped me along the way.

I arrived in Cardiff in 2013, naive, tidy, and hungry to learn astrophysics. Well it's been 8 years since, and I can say with a heavy heart that I'm finally leaving Cardiff University, hopefully still tidy, but certainly still somewhat naive. When I first bluffed Dr. Timothy Davis that I knew anything about ALMA, I didn't realise I'd landed myself a PhD with one of the best supervisors on the planet in every respect. When you couple that with the fact that Dr. Edward Gomez too embraced me as one of his students, I doubt you will find another PhD student whose studies were guided quite so gloriously. PhDs can be hard, yet it's only when I look back at my time under the guidance of Tim and Edward that I realise why my experience seemed more fun than gruelling. It's only in hindsight that I can clearly see just how many opportunities they opened for me, subtly guiding me into a position where the world truly is out there should I just reach out and employ what they have both taught me.

I'm sure I could have spent the second 4 years of my career at Cardiff University as a playstation herm... err I mean academic hermit; but I didn't and I'm glad I didn't because if I had I would never have met some amazing people. 53 the Parade was an incredibly motivating and amusing place to work (office golf anyone?) and I am thankful that I was a part of the 53 galaxy group before we all moved back into the physics buildings and had to show at least some measure of decorum and professionalism. A special thank you of course to Nikki for making the PhD and

everyday since more like a daily lesson in how to have fun; and to Davey for putting up with my champagne problems and whining about lack of GPUs.

Thank you to all my friends in the astro group for the Friday pub excursions, the trips, hikes, and parties to stay sane. Thank you also, to those outside of the astronomy group who played a major part in my studies. Those of you who work on those things we can't legally talk about in that place we're not supposed to mention... I am forever grateful that I crossed paths with you more times than once.

I always thank these specific people in particular for getting me where I am in my academic career and I'm going to do it again... sorry. To Dr. Edward Gillen and Prof. Didier Queloz (congratulations on the Nobel prize!), if you hadn't allowed me to research machine learning with you at Cambridge I would not be in such a privileged position to this day. I hope to one day thank you both in person again but safe to say my career exists only because of you both. To Dr Hardie Pienaar, you kickstarted my life of searching for adventure and saying yes to new ideas and experiences. I will always be in your debt as a functioning human being and a thesis quote just didn't seem good enough to emphasise your influence on my ability to knuckle down and get solving problems. Last but definitely not least, it goes without saying but, thanks to my mother and father, (known to myself as Mum and Dad) for your help and eternal support.

Abstract

This thesis investigates the use of machine learning for analysing the kinematics of galaxies in a time efficient manner. The application of machine learning in astronomy is arguably nascent, and very much so in the case of galaxy kinematics. Being able to extract kinematic information at speed will be important come the advent of next generation telescopes such as the Square Kilometre Array. Such instruments will collect raw data on scales too large to store. Therefore, the use of on the fly modelling techniques, harnessing the power of machine learning, is crucial. I will show that it is possible and beneficial to use machine learning algorithms to tackle scientific questions in extragalactic astronomy in this way.

This thesis starts by investigating the use of machine learning algorithms for rapidly discriminating between disturbed and orderly rotating gas discs in galaxies. Specifically, cold dense molecular gas discs are embedded onto a latent manifold using convolutional autoencoders (CAE) which boast powerful automated feature embedding capabilities. Using hydrodynamical simulations to create mock observational data, the CAE is trained on millions of naturally augmented moment one maps before testing on observational HI data from the Local Volume HI Survey (Koribalski et al., 2018), as well CO observational data from various surveys using ALMA. Using a simple binary classifier on the embeddings, it can be shown that disturbed and orderly rotating discs are separately classified with high accuracy even in the presence of injected noise. Such models may be useful as fast filtering tools for identifying mergers or relaxed discs for further kinematic modelling.

Bearing in mind that transfer learning for next generation survey datasets holds great risk, a new approach to kinematically characterising gas in galaxies is studied next. Using self-supervised physics-aware neural networks, the need for a throw-away training set is removed entirely, and replaced with a model which can learn physical parameterisations of galaxy rotation curves at rapid speed. With the

introduction of monte carlo dropout, it is also possible to recover modelling errors for kinematic parameters, which will be useful in gauging the validity of learned parameters. These models are tested on simulated data as well as observational CO data from the WISDOM survey and HI data from THINGS (Walter et al., 2008). Learned rotation curves match well with those derived from more analytically motivated modelling tools (e.g. `BBarolo` Di Teodoro & Fraternali 2015), but compute parameterisations in a fraction of the time.

Finally I study the use of the aforementioned self-supervised physics-aware neural networks, to recover the Hα Tully-Fisher relation (TFR) from largest IFU dataset to date. To do so, moment maps from both SAMI and MaNGA IFU surveys are used to derive the rotational velocities of low redshift galaxies. These are then fit against mass to derive both the forward and reverse TFR. The fits are in agreement with those found in the wider literature except that my fits have shallower gradients because a correction for asymmetric drift is applied in this work, but not in the comparison fits from the literature. Here, I identify and quantify trends between position along (and perpendicular to) the TFR and galaxy properties, namely: age and mass-to-light ratio. A clear relation is also discussed between velocity turnover radius, r_{turn}/r_e, and stellar mass. The application of models originally designed for use with millimetre and radio interferometric data, shows the benefits of using self-supervised physics-aware approaches to circumvent the problems often associated with transfer learning. Such methods will be useful when applied to next generation IFU survey data releases, with instruments such as HECTOR.

In summary, in this thesis, I explore the different machine learning approaches to kinematically characterise galaxies in a time-efficient manner. I conclude with some remaining questions and avenues for future research.

Publications

First Author:

James M Dawson, Timothy A. Davis, Edward L. Gomez, Justus Schock, Nikki Zabel, and Thomas G. Williams (2019), MNRAS, 491, 2506: "*Using machine learning to study the kinematics of cold gas in galaxies*" –Chapter II

James M Dawson, Timothy A Davis, Edward L Gomez, and Justus Schock, 2021, MNRAS, 503, 1: "*A Self-supervised, Physics-aware, Bayesian Neural Network architecture for modelling galaxy kinematics in preparation for the SKA*" –Chapter III

James M Dawson, Timothy A Davis, and Edward L Gomez, 2021, in prep: "*The stellar mass Tully-Fisher relation with SAMI & MaNGA using self-supervised, physics-aware, Bayesian neural networks*" –Chapter IV

Co-Author:

Timothy A. Davis, Jenny E. Greene, Chung-Pei Ma, John P. Blakeslee, **James M Dawson**, Viraj Pandya, Melanie Veale and Nikki Zabel, 2019, MNRAS, 486, "*The MASSIVE survey – XI. What drives the molecular gas properties of early-type galaxies*"

Timothy A. Davis, Nikki Zabel, and **James M Dawson**, June 2020, ascl:2006.003: "*KinMS: Three-dimensional kinematic modelling of arbitrary gas distributions*"

Contents

List of Tables

List of Figures

Chapter 1
Introduction

> Light is the only connection we have with the Universe beyond our solar system,
> and the only connection our ancestors had with anything beyond Earth.
> Follow the light and we can journey from the confines of our planet to other worlds that orbit the Sun
> without ever dreaming of spacecraft.
>
> Professor Brian Cox, *"Wonders of the Universe"*

1.1 Extragalactic Astronomy

The Milky Way Galaxy, within which our solar system is but one of many, was first observed with a telescope in 1610 by none other than Galileo Galilei. It was then that he discovered the Milky Way is made up of a vast number of stars and is not simply a pale, cloudy band on the night sky. It wasn't until 1925, when Edwin Hubble pointed his telescope at the Andromeda galaxy, that astronomers could prove the existence of stellar systems definitively beyond our own Galaxy. Upon discovering Cepheid variable stars in M31, Hubble was able to show that M31 must be at least 300 kpc away and therefore of comparable size to the Milky Way. Thus, Hubble concluded that it must be stellar system external to the Milky Way (Hubble, 1929). At the time, galaxies simply were regarded as 'nebulae' and astronomers were divided

on whether they were relatively small objects within the Galaxy or large bodies residing outside the Milky Way. Hubble's conclusive evidence ended this so-called 'Great Debate' (also known as the Shapley–Curtis Debate, Shapley & Curtis 1921), and birthed the beginning of extragalactic astronomy. It is both interesting and ironic to note that the method used by Hubble to end the Great Debate was the same as that used by Harlow Shapley (who was at the time, measuring the distances of globular clusters inside the Galaxy) who was opposed to the idea of galaxies being external to the Milky Way.

Today, extragalactic astronomy is the study of the formation and evolution of galaxies from the *Big Bang* right down to redshift zero. Thanks to advances in instrumentation over time, astronomers are now able to study galaxies close to the edge of the observable Universe. All sky extragalactic surveys are no longer hypothetical and allow for large statistical samples of multiwavelength galaxy observations. Thanks to this multiwavelength approach, extragalactic astronomy concerns itself with everything from active galactic nuclei (AGN) and black hole accretion on light-minute scales, to the presence of dark matter structures spanning across the entire universe. These days, extragalactic astronomy is even concerned with studying the effects of General Relativity, such as gravitational lensing and gravitational waves, which are virtually impossible to observe on galactic scales.

1.1.1 GALAXY MORPHOLOGY

Our own galaxy, the Milky Way, is known to have spiral substructure, as do many other galaxies around us. Yet spiral galaxies do not make up the entire population of galaxies in the Universe. For example, our closest extragalactic neighbours, the Magellanic Clouds, are more irregular, and devoid of strong spiral-like substructure, and if we look a little further out into the Universe we come across elliptical galaxies. These are galaxies apparently devoid of substructure and are characterised by their smooth, regular, form. They are (as the name suggests) ellipsoidal and random internal motions play a more dominant role in determining their shape.

Spiral galaxies, on the other hand, are characterised by substructures of gas, dust, and bright star forming regions. As young stars tend to be more blue than older stars, spirals also tend to be bluer than ellipticals, which must contain more old, low-mass (and thus red) stars (see Figure 1.1). Upon first inspection, spiral galaxies appear as flat disks, but do often have a central 'bulge' which exhibits similar velocity dispersion characteristics as elliptical galaxies.

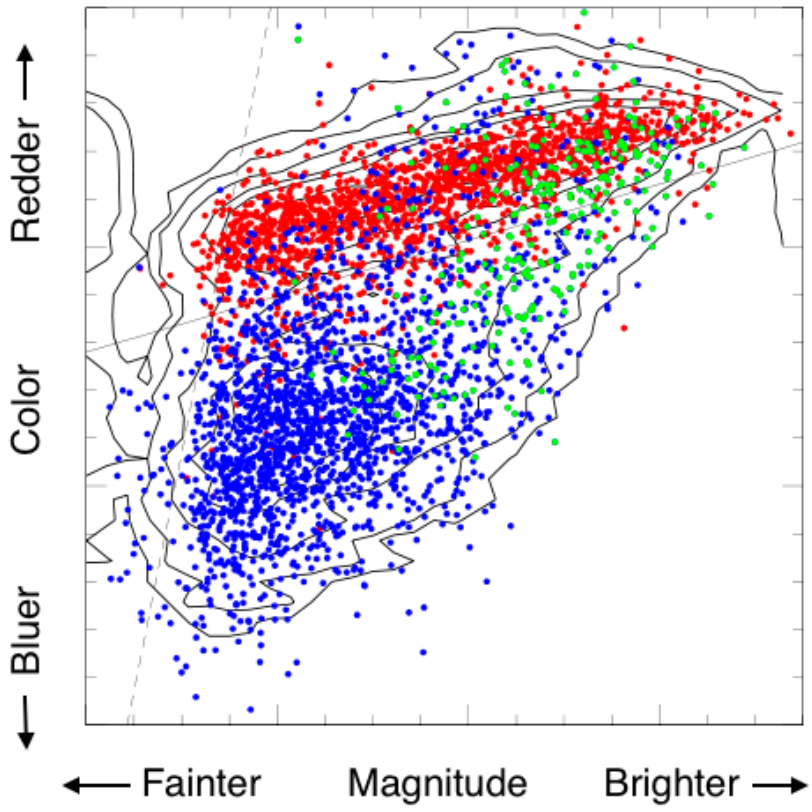


Figure 1.1. A colour-magnitude diagram of the Sloan Great Wall, a collection of superclusters spanning a redshift range of z = 0.04 - 0.12. Axes show $g-i$ colour versus i-band magnitude. Spiral galaxies (blue points) tend to be bluer and fainter than the ellipticals (red points) which lie in a relatively tight "red sequence". Green points are described as 'bulge galaxies' (Sa-Sb in the Hubble classification scheme). The image is adapted from Figure 3 of Gavazzi et al. 2010 by Ben Cook for Astrobites. In this adaption, the physical axes labels have been removed for clarity; in the publication itself the axes range from 0 to 1.5 in colour, and -16 to -23.5 in magnitude.

Already we have listed several 'species' of galaxy shape, including spiral, elliptical, and irregular (as exhibited by the Small Magellanic Cloud) which we call morphological type. John Reynolds made the first attempt at a more granular galaxy morphological classification in the early 1900s. However, this was before the end of the 'Great Debate' and it was Hubble (again) whose classification system (Hubble, 1936) is still used to this day. This is known as the 'Hubble tuning fork' and is shown in Figure 1.2.

The 'handle' of the fork shows the elliptical galaxies which range from circular to more flattened appearances, when projected on the sky. The two 'prongs' of the tuning fork represent the spiral galaxies separated by whether they exhibit a central bar emanating from the central bulge or not. Roughly half of all galaxies in the

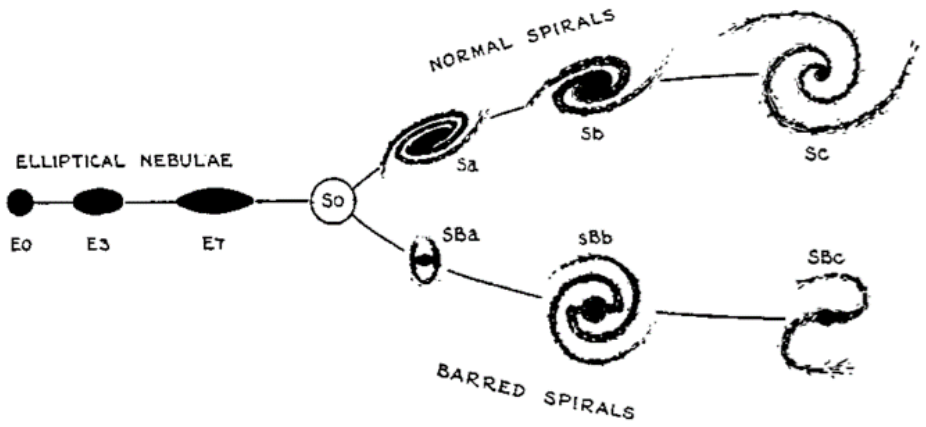


Figure 1.2. Hubble's 'tuning fork' diagram of galaxy morphological classification. Galaxies are subdivided along the tuning fork into ellipticals ('early-types'), spirals, and barred spirals (both 'late types'). Originally, astronomers presumed that galaxies evolved from left to right in the diagram, however we now know this to be incorrect. Image from Alladin & Hasan (2007).

Universe appear to be barred (Mihalas & Routly, 1968; Eskridge & Frogel, 1999). When traversing the two prongs, the classifications change depending on how tightly wound the spiral arms are. Since its inception, many have built upon the original tuning fork, adding more physically motivated classifications. Yet, the tuning fork still remains a starting point for morphological classification to this day.

It was initially believed that galaxies evolved from ellipticals into spirals (i.e. left to right on the tuning fork). Thus, they are also called 'early-type' and 'late-type' galaxies respectively. We now have mounting evidence that suggests galaxies actually evolve in the reverse way, from spirals to ellipticals. In the next section, I will describe how and why galaxies evolve in this manner.

1.1.2 Galaxy Evolution

The hierarchical Universe

The currently accepted cosmic history of galaxies in the Universe is based on the Λ–CDM model. This model assumes that the matter in the Universe is dominated by cold dark matter (CDM) and that dark energy (being represented by the cosmological constant Λ) drives the expansion of the Universe at late times. Primordial fluctuations in the early Universe (Turner, 1999) are thought to have initiated over-densities, growing by attracting dark matter and gas, leading to the formation

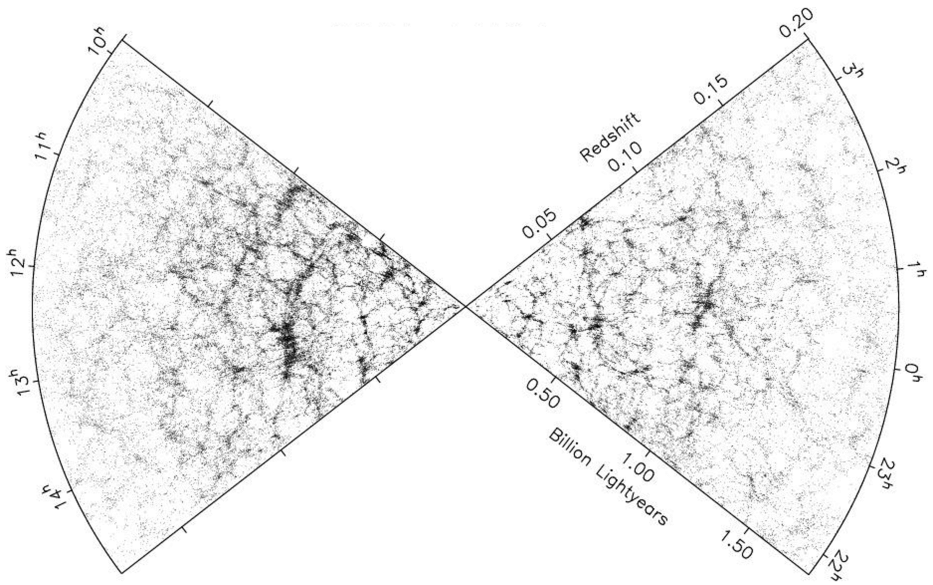


Figure 1.3. The galaxy distribution uncovered by the 2dF galaxy redshift survey. 221414 galaxies in the final 2dFGRS catalogue are shown as black points. Clearly visible are the filaments and clusters of galaxies surrounding low density voids. Image courtesy of (Colless et al., 2003).

of dark matter *halos* (Springel et al., 2005). The gas collected in these halos is then thought to have cooled, condensing out into the first stars and galaxies. This process of accumulating matter and consolidating smaller cosmological objects into larger ones is known as the 'hierarchical model' of the Universe (Kauffmann et al., 1999).

The earliest galaxies known to exist exhibit irregular, clumpy morphologies. According to our developing theoretical framework of high redshift galaxy formation, cold gas is thought to inflow into high redshift galaxies along filaments of the larger cosmic web, constantly replenishing gas and opposing losses of star formation (from both environmental and secular processes). High gas fractions and high density in the early Universe lead to perpetual states of violent disc instabilities (characteristically higher velocity dispersions and more massive perturbations), operating on short orbital time-scales as opposed to the slow secular evolution of some galaxies observed in nearby galaxies (Dekel et al., 2009; Cacciato et al., 2012).

The hierarchical model goes beyond the creation of the first galaxies and stars; galaxies can merge to become more massive galaxies. Similarly, it is thought that galaxies tend to stick together and form groups which share a common dark matter halo. When enough groups combine, they can eventually form galaxy clusters, which can combine to form *super*clusters. This extension of the 'hierarchical model' is known as 'hierarchical clustering'. Even superclusters are connected, filaments of

galaxies joining superclusters in the large-scale structure of the Universe run adjacent to 'voids' –parts of the Universe where galaxy densities are comparatively low. Together, the voids and filaments form the 'cosmic web', shown in Figure 1.3.

The evolution of galaxies over cosmic timescales

The evolution of spiral galaxies into ellipticals coupled with the hierarchical model of the Universe point to a Universe which is constantly evolving. Slow, steady evolution of galaxies can either be a result of long-term interactions with surrounding environments, or induced by the internal actions of e.g. spiral arms or bars. This secular evolution plays an important role in the appearance and properties of galaxies but cannot account for the changes in morphology seen between high-z galaxies and those found in the local Universe. In the local Universe, galaxies merge and evolve until their star formation halts with elliptical-like morphologies. If this were to continue, we will end up with a mostly empty Universe, sparsely populated by old dead galaxies. However, galaxies did not spring into existence with peak star formation rates. Instead, the star formation rate of the Universe increases until $z \approx 2$, where it peaks and continues its decline today (Madau & Dickinson, 2014). This behaviour is reflected in the number of quiescent galaxies through cosmic time, found to increase rapidly as we approach the present (Moustakas et al., 2013). In fact, Kelvin et al. (2014) found that $\sim 70\%$ of the stellar mass in local Universe is found within elliptical and lenticular galaxies.

Below, I describe in more detail some of the processes believed to be important in the ongoing evolution of galaxies.

Galaxy–galaxy mergers

The first process thought to be important in galaxy evolution is an immediately obvious one.

Due to the hierarchical nature of the ΛCDM model Universe, galaxy–galaxy mergers are thought to play a key role in the creation of ellipticals and thus contribute to galaxy evolution as a whole (Beckman et al., 2008).

As higher intrinsic velocity dispersions are prevalent in ellipticals, dissimilar to their progenitors, it follows that the evolution of spiral galaxies into ellipticals can (in some cases) be dynamical, rather than a simple evolution of their stellar populations (Moore et al., 1996; Martinez-Valpuesta et al., 2006). In fact, contrary to first appearances, when looking closely at ellipticals we begin to uncover a menagerie

of interesting features ranging from dust lanes and shells (Mancillas et al., 2019), to tails (Ren et al., 2020) and even double nuclei (Lawrence et al., 2020). All of these features indicate a dynamical past for some galaxies and to some degree provide evidence of past merging events and ultimately, a dynamically evolutionary path of galaxies through time.

It is not fully understood just how significant mergers are in galaxy evolution in comparison to other evolutionary mechanisms. In addition, merger events can vary characteristically depending on many factors. For example, differences in mass and gas content of merging galaxies play a pivotal role in the state of the final merger product. If one or both of the galaxies is a gas-rich spiral, then the merging may result in an initial 'star burst', where the star formation rate is increased massively due to the compression of gas over a short period of time. Regardless of the merging orientation, proximity, or constituent galaxy properties, it may at first seem unlikely that post-merger galaxies would not leave some evidence of their merging histories whether in their material abundances or kinematics. Yet relaxing dynamics, and ageing populations of stars mean that visibility times for most of this evidence is only around a gigayear at most (Lotz et al., 2008).

Environment

In the local Universe, with more than 40% of galaxies residing in groups or clusters (Zabludoff & Mulchaey, 1998; Robotham et al., 2011), it follows that environment plays an important role in the evolution of galaxies. Overdensities of galaxies can have an impact on the prevalence of star formation quenching when compared to that in under-dense regions (also called 'the field').

A correlation does exist between galaxy type and environment (Oemler, 1974; Dressler, 1980). Specifically, in dense environments there are more 'passive' galaxies than in sparser environments (see Figure 1.4). That is, more galaxies which either currently exhibit a state of minimal star formation, or have undergone 'passive' evolution, using up their fuel and leaving a population of old stars (mostly found for high mass galaxies and those with a particularly large bulge Fang et al. 2013; Bluck et al. 2014; Bremer et al. 2018). Additionally, the denser the region the greater the fraction of ellipticals. The number of dwarf galaxies also appears to show a dependence on density (Mistani et al., 2015). From a larger time-scale perspective, galaxy clusters at higher redshifts have been shown to have more blue galaxies compared to low-redshift clusters. This is the 'Butcher-Oemler effect' (Butcher & Oemler, 1978). It implies that most of the star formation occurs in clusters at higher redshifts and

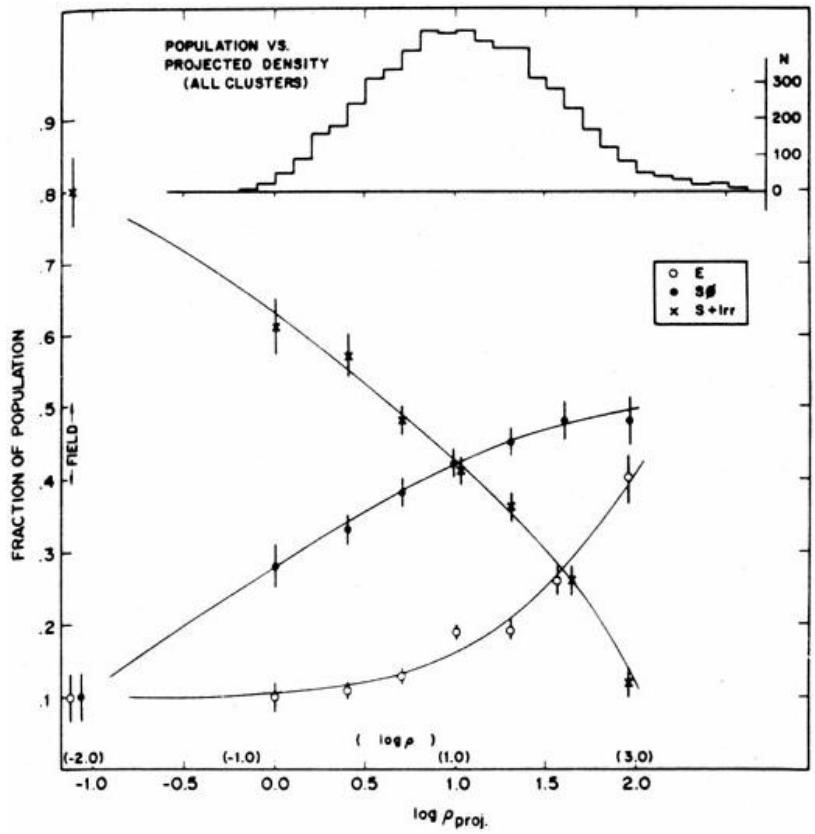


Figure 1.4. From Dressler (1980): the morphology-density relation. This figure shows the relationship between galaxy type abundance as a function of the projected density of galaxies per Mpc^{-2}. As the projected density of galaxies increases, the fraction of spiral and irregular galaxies decreases, while the fraction of ellipticals increases. A histogram for the total population of galaxies in this sample is shown in the upper region of the figure. This work was one of the first highlighting the clear correlation between morphology and galaxy density.

so SFR quenching is relatively recent. Each of these relationships between environmental density and galaxy morphology must arise due to either 'nature or nurture'. i.e. either different formation mechanisms or different evolutionary paths for galaxies in different environments.

Interestingly, the relative velocities of galaxies in groups clusters are much higher than in the field. Contrary to premature logic (in which one might relate overdensities to higher merger rates), despite the overdense nature of groups and clusters, mergers (one of the main ways in which spiral galaxies evolve into ellipticals in the field) are in fact rare here. The relative velocities of galaxies in clusters are simply too high to allow them to become gravitationally bound to one another. Instead, when two galaxies meet in clusters, they usually fly by tidally disrupting either one or both members' morphology and content.

A lack of galaxy mergers is not to say that dynamical evolution of galaxies does not occur in groups and clusters. Some of the more extreme events, such as ram pressure stripping (Gunn & Gott, 1972) and strangulation (Vaughan et al., 2020), occur when galaxies fall into the high density intracluster medium (ICM) present in clusters.

Feedback

Galaxies can be thought of as closed systems for most of their lives, regulating their star formation and evolution in a feedback loop. Feedback processes which impede star formation mainly come from stellar winds, supernovae, and active galactic nuclei (AGN). When star formation rates are high, stellar winds and supernovae are powerful enough to directly heat and propel cold gas out of galaxies into their halos. Both stellar winds and supernovae influence the efficiency of star formation in galaxies too. Stellar winds can disperse self-gravitating dense clouds of gas and can be responsible for supplying half the total gas mass returned to the ISM (Leitherer et al., 1992). Supernovae (one of the first mechanisms thought to regulate star formation) shock the surrounding gas, increasing its temperature and significantly slowing down star formation. Some supernovae are even capable of generating galactic-scale winds capable of removing material from galaxies. Some of this gas will escape and enrich the IGM, while the rest will eventually fall back onto the galaxy. When material is accreted onto super-massive black holes at the centres of galaxies, energy and radiation are released into the surrounding ISM. These AGN were first discovered in 1943 and can play a significant role in galaxy evolution. There are two types of known AGN feedback modes: radiative and kinetic (Qiu et al., 2019). In radiative feedback, cold gas is displaced from (typically) disc galaxies. In kinetic feedback, powerful jets can be seen extending out from what are typically bulge-dominated galaxies into their hot halos. The expanding cocoons driven by these jets shock heat the halo, stopping this material from cooling onto the galaxy and going on to form new stars. Importantly, the energy produced by accretion of material onto the black hole, exceeds the energy needed to bind material to the galaxy. This, when coupled with also heating the surrounding ISM, points to AGN having a direct impact on the material available for star formation.

1.1.3 The ISM and its phases

When describing galaxy evolution, above, almost all of the important processes concern the gas in galaxies, which provide the future fuel for star formation. This

gas is part of the material between stars, yet still bound to galaxies. We call this collection of gas, the interstellar medium.

The interstellar medium of galaxies exhibit a variety of different states, or 'phases', of matter. This matter can be ionised, atomic, or molecular gas, as well as dust (the sub-micron, solid-particulate kind, not the kind you find covering your long-lost lecture notes!) and cosmic rays. These phases have different densities and temperatures ranging from $10^{-4} \rightarrow 10^{6}$ particles per cm^3 and $10 \rightarrow 10^{7}$ K respectively, and are thus visible at different wavelengths/frequencies. Studying the different phases of the ISM is driven by different science cases and goals. In the following section I will outline the phases of the ISM most relevant to the work described in this thesis. Figure 1.5, shows the approximate structure of the galactic ISM phases and can be used as a reference throughout the following sections.

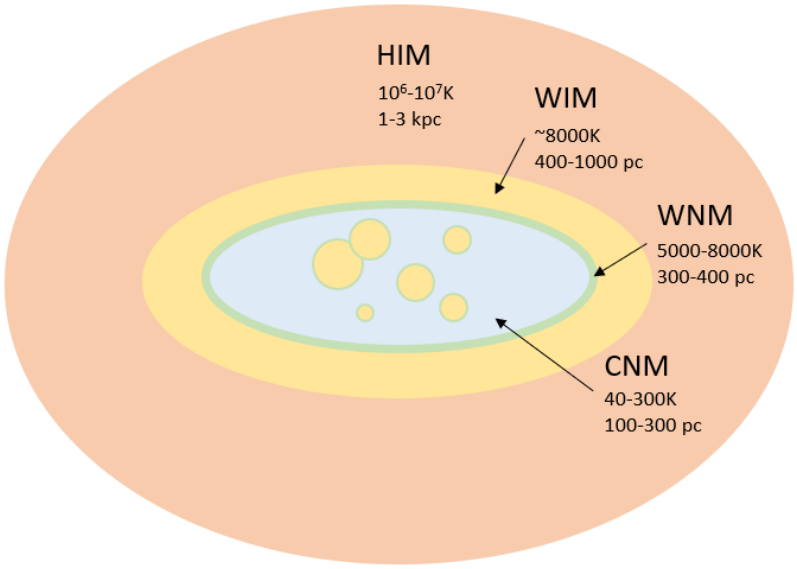
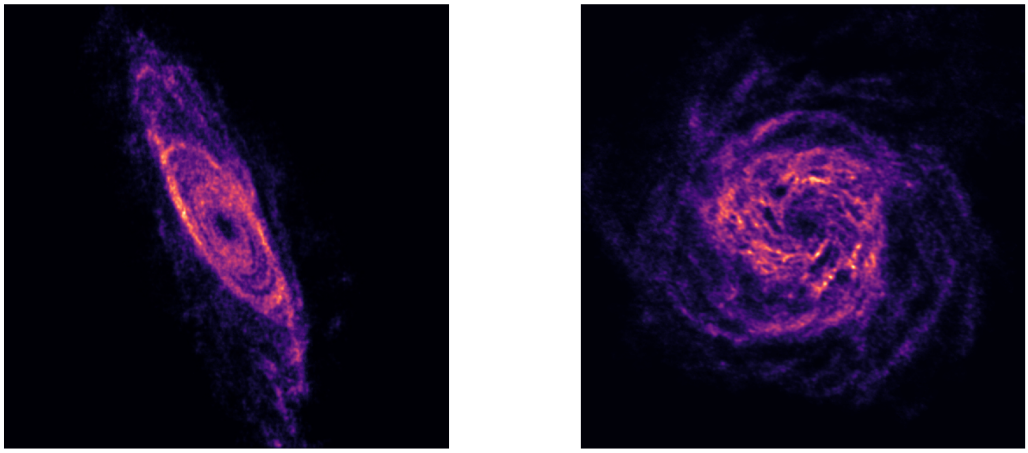


Figure 1.5. A pictorial representation of the ISM phases of a galaxy, as seen edge on. Phases are in descending temperature and ascending scale height: the hot ionised medium (HIM), warm ionised medium (WIM), warm neutral medium (WNM), and cold neutral medium (CNM). Temperatures and scale heights shown are approximate bounds on the ISM phases in equilibrium. Circular ionised regions in the CNM represent HII regions. Note: scale heights are provided for each phase in this pictorial representation, but the image itself is not to scale.

Atomic Hydrogen (HI)

Hydrogen is the most abundant element in the Universe. Atomic hydrogen (HI, belonging to both the warm and cold neutral medium) constitutes one of the



(a) NGC 2841 **(b)** NGC 628

Figure 1.6. VLA observations of the 21 cm HI emission line, from The HI Nearby Galaxy Survey (THINGS, Walter et al. 2008). Featured are NGC 2841, and NGC 628. Both are archetypal spiral galaxies rich in HI.

most useful ISM states to observe when investigating galaxy kinematics and environmental interactions. This is in part due to the fact that for field galaxies, HI tends to reside in a disk extending far beyond all other ISM phases. Coupled with its relatively low density (less than a few hundred per cm^3), HI is easily influenced by the surrounding environment and thus lends itself as a useful tracer of dynamical events and interactions. Example observations of HI in galaxies are shown in Figure 1.6.

In 1944 Henk van de Hulst predicted that neutral hydrogen would emit in the radio waveband at 21.1 cm (1.4 GHz) due to a hyperfine transition (Oort, 1997). The proton and electron in a hydrogen atom have charges which, along with their spin motions, creates a dipolar magnetic field -a 'magnetic moment'. Besides the quantum numbers representing the allowed electron orbits in a hydrogen atom, the electron has a 'spin' quantum number. If the electron spin is aligned with the proton spin then the energy differs from if they are aligned in the opposite directions. If the electron 'flips' from an aligned state (where the magnetic moments add together) to antiparallel, energy is released as a 21 cm photon. Although the transition is very unlikely (a given electron is likely to flip only once every 3.5×10^{14}s), there is so much neutral hydrogen in galaxies that it is easily and widely detected. For example, the neutral hydrogen mass of our the Milky Way Galaxy is $\sim 6 \times 10^9 M_\odot$(Licquia & Newman, 2015), which corresponds to 10^{66} hydrogen atoms and therefore $\sim 3 \times 10^{51}$ transitions

a second. In addition, exchanges of electrons between colliding hydrogen atoms in the ISM can lead to adopting electrons leading to antiparallel spin alignments and thus leads to even further transitions. Despite the huge amounts of HI in galaxies, it still generally remains optically thin. As such, it is possible to use the intensity of the 21 cm line to infer the column density and hence the total hydrogen masses of galaxies. However, it is worth noting, it has been shown that self-absorption can become severe for highly inclined galaxies, in which case a correction factor is sufficient for converting the observed 21 cm line flux to the correct HI mass (Dickey, 1990).

Molecular gas

Molecular gas clouds in the ISM can have temperatures as low as $\sim$ 10K, and densities as high as $\sim 10^{12}$ molecules/cm^3. Stars typically form in 'giant molecular clouds' (GMCs) and as cold molecular gas is an intermediate stage between the HI phase and stars, it is extremely likely that molecular gas is the direct fuel for star formation in all but the most metal poor regions of the universe (Glover & Clark, 2016).

Molecules radiate energy via rotational and vibrational transitions of the whole molecule. Given that hydrogen is the most abundant material in the ISM, it follows that molecular hydrogen (H_2) should be the most abundant molecule. However, H_2 has no dipole moment and therefore observational tracers of its presence are hard to come by. Vibrational transitions of H_2 do exist, however these only occur at temperatures > 1000 K and are therefore inappropriate as tracers of the cold dense gas of the ISM. Instead, astronomers observe the rarer but more easily detected carbon monoxide (CO) molecule as a tracer of where H_2 resides. It is worth noting that this method of tracing the H_2 content induces the need for a conversion factor between CO and H_2 abundances. The validity of this conversion factor (known as X_{CO}) remains a controversial topic to this day (Bolatto et al., 2013) and is (thankfully) not used in the work presented in this thesis. Rotational transitions of CO (more specifically the molecular isotope ^{12}CO are particularly strong, given high enough densities, even at temperatures between $10 \rightarrow 20$ K. The most commonly observed rotational lines of CO are the CO(1-0) and CO(2-1) lines, at $\sim$ 115 GHz (1.3 mm) and $\sim$ 230 GHz (2.6 mm) respectively.

Thanks to the improvement of observational capabilities in the last decade with instruments like ALMA (see §1.2), it is now possible to extract exquisite kinematics of molecular gas in galaxies beyond the Milky Way. in this thesis I make use of such data in order to test novel machine learning models on high resolution data

throughout.

Ionised Hydrogen

Besides the atomic and molecular hydrogen found in the ISM, hydrogen is not only found in the outer extremities of galaxies but also in the more central regions of the disc and typically in the so-called 'HII regions', alongside other ions such as oxygen, nitrogen, and sulphur. Typically ionised gas is found in a "skin" around the colder atomic and molecular material, where the clouds meet the ambient interstellar radiation field. In HII regions, on the other hand, are regions around massive hot O and B type stars. In these regions, luminous young stars emit ionising ultraviolet photons which, under recombination, result in line spectra. In the case of hydrogen, the capturing of electrons into the n=2 orbital or higher state, result in further non-ionising photons being released as they drop to the lower energy levels. A significant number of these produce the characteristic Balmer lines, notably: Hα where $\lambda = 656.3\,\text{nm}$. Other important lines in ionised regions include O[II] and O[III] at $\lambda = 372.9$ and $\lambda = 495.9\,\text{nm}$ respectively, N[II] at $\lambda = 373.9\,\text{nm}$, and S[II] found at both $\lambda = 671.7$ and $\lambda = 673.3\,\text{nm}$.

1.1.4 Galaxy kinematics

At it's simplest, studying the kinematics of galaxies means observing the motions of a galaxy's internal components (typically the gas and stars)in order to answer basic questions such as how galactic discs form and evolve. In practice, this usually means inferring the distribution of dynamical mass within galaxies, which in turn leads to understanding the shape of their gravitational potentials.

One of the large benefits of observing emission lines is that they encode information on the kinematics of the gas. In an ideal world, emission lines would look like delta functions. However, in reality they often have Gaussian like profiles with widths caused by different broadening effects. Intrinsic broadening arises due to the Heisenberg uncertainty principle which states that we cannot know the velocities of gas particles with perfect certainty. Further broadening can arise due to the actual motions of gas, this includes turbulent broadening as well as thermal broadening (or thermal Doppler broadening, Draine 2011).

The dominant source of motions for gas in external galaxies is actually due to gravitational interactions of gas clouds with their surroundings- often dominated by rotation (and sometimes by non-circular motions). For rotation dominated gas, their

rotational velocities are dictated by their internal mass distributions and can have velocities of up to $\pm 600\,\mathrm{km\,s^{-1}}$ (Davis et al., 2015).

Observations probing the interstellar gas of galaxies allow us to measure the rotational properties of galaxies far beyond the limits of their relatively centrally concentrated stellar components. Measuring the dynamics of gas at the farthest reaches of galaxies allows us to put constraints on gravitational interactions between galaxies but it also allows us to begin probing the dynamics of the dark matter halos within which galaxies reside.

Deep HI observations show that even beyond the stellar disc, the rotational velocity of galaxies tends to fail to fall in the expected Keplerian fashion. Flat rotation curves out to the limits of observations tell us that the masses enclosed within the farthest reaches of observed discs must be much greater than the constituent matter seen by astronomical instruments (Bosma, 1978). This provided the main impetus behind the serious theoretical consideration of dark matter in the 1980s and feeds into the components of the hierarchical ΛCDM Universe.

Deviations from purely circular motions of gas in galaxies, on the other hand, can arise from different mechanisms, including chaotic non-circular motions induced by star formation, systematic non-circular motions related to secular processes such as spiral arms and bars removing gas from circular orbits (e.g. Trachternach et al. 2008; Galloway et al. 2015), or more dynamical events such as mergers, tidal interaction, and gas accretion (e.g. Bournaud et al. 2004; Polletta et al. 2011; Ho et al. 2019).

1.1.5 BIG DATA IN EXTRAGALACTIC ASTRONOMY

Astronomy is currently undergoing a rapid acceleration in data collection with next generation instruments poised to image enormous portions of the sky. The advantages of collecting data at faster rates and in more detail are obvious but remain key. Higher instrument sensitivities will lead to the detection of previously unseen sources, faster data collection rates means a higher chance of detecting transient events, larger datasets means greater homogeneity in large-scale studies... the list goes on.

The study of HI requires the use of interferometers which collect data of enormous scale per-source already, and next generation interferometers are only going to get more powerful. The Square Kilometre Array (SKA, which will be the world's largest radio telescope) will generate more than an Exabyte of data every day, dwarfing internet usage of the entire globe, squarely seating astronomy in the *petabyte era* (Dewdney et al., 2009). Even the current (and operational) SKA pathfinder telescope,

MeerKAT (Jonas & MeerKAT Team, 2016), has an input data rate of hundreds of gigabytes per second and ASKAP (the Australian SKA pathfinder telescope, Duffy et al. 2012) with its advanced phased array feeds generates tens of terrabytes per second.

It's not only radio astronomy that is making the push into the *big data era*. For example the Legacy Survey of Space and Time (LSST), is a ten year optical survey in which the entire night sky will be images every few nights, generating 500 petabytes of data over the survey lifetime (Jurić et al., 2017). Across all of astronomy, yearly advances in electronics bring instruments that vastly increase data collection size, and with them astronomers have had to update data reduction and analysis pipelines to keep delivering valuable scientific output. Yet, typical sky surveys can archive hundreds of terrabytes of data with billions of detected sources and hundreds of attributes per source. The vast size and dimensionality of these datasets calls for a new way of working. With these new instruments, gone are the days where astronomers analyse sources by hand. In their place automated pipelines, statistics, and machine learning are emerging or are already securely rooted as part of the astronomical scientific method accompanied by high performance computing hardware to support them.

In this thesis I will present my work applying machine learning techniques to try and maximise the scientific return from future emission line surveys carried out using e.g. the SKA. In order to do this, I use gas kinematic data products to identify and study circular and non-circular motions in the ISM of galaxies. In the next section I will discuss the facilities that provide the data used throughout the thesis, before going on to introduce the machine learning techniques used, in section 1.3.

1.2 Telescopes & Instruments

For the work presented in this thesis, data from several astronomical telescopes and instruments are used. As such, it is useful to discuss these telescopes and instruments, in order to better understand the properties of their resultant datasets and the link between operational wavelength and the science that they produce.

1.2.1 Interferometery

A single radio telescope dish acts like a bucket, collecting radio-waves and reflecting them onto a receiver at its focus. The radio receiver translates the incoming

Figure 1.7. IRAM located in the Sierra Nevada, Spain, close to the Pico Veleta peak, with me posing nearby while very chilly. Image credit: me.

radio waves into a voltage that can be measured and recorded. Depending on the onboard instrument, the measurements can be used to create a radio intensity map, or spectral line profile of a source. Single-dish long-wavelength telescopes have been in operation around the globe for many decades. Notably among them are (in ascending size), IRAM (30 m, see Figure 1.7), Jodrell Bank (76 m), Arecibo (300 m), and FAST (500 m). Some of these telescopes are shown in Figure 1.8 around the globe.

The resolving power of a telescope is dependent on the wavelength being observed and the diameter of the telescope due to the physics of diffraction. Specifically,

$$\theta_{min} = 1.22\frac{\lambda}{D} \approx \frac{\lambda}{D}, \tag{1.1}$$

where θ_{min} is the angular resolution, λ is the observed wavelength, and D is the diameter of the telescope.

However, millimetre to radio waves have much longer wavelengths than visible light. So if an optical telescope and a radio telescope had the same diameter, the radio telescope would have much weaker resolving power. For example, for a 5 m optical telescope (such as the Hale optical telescope), in order to achieve the same

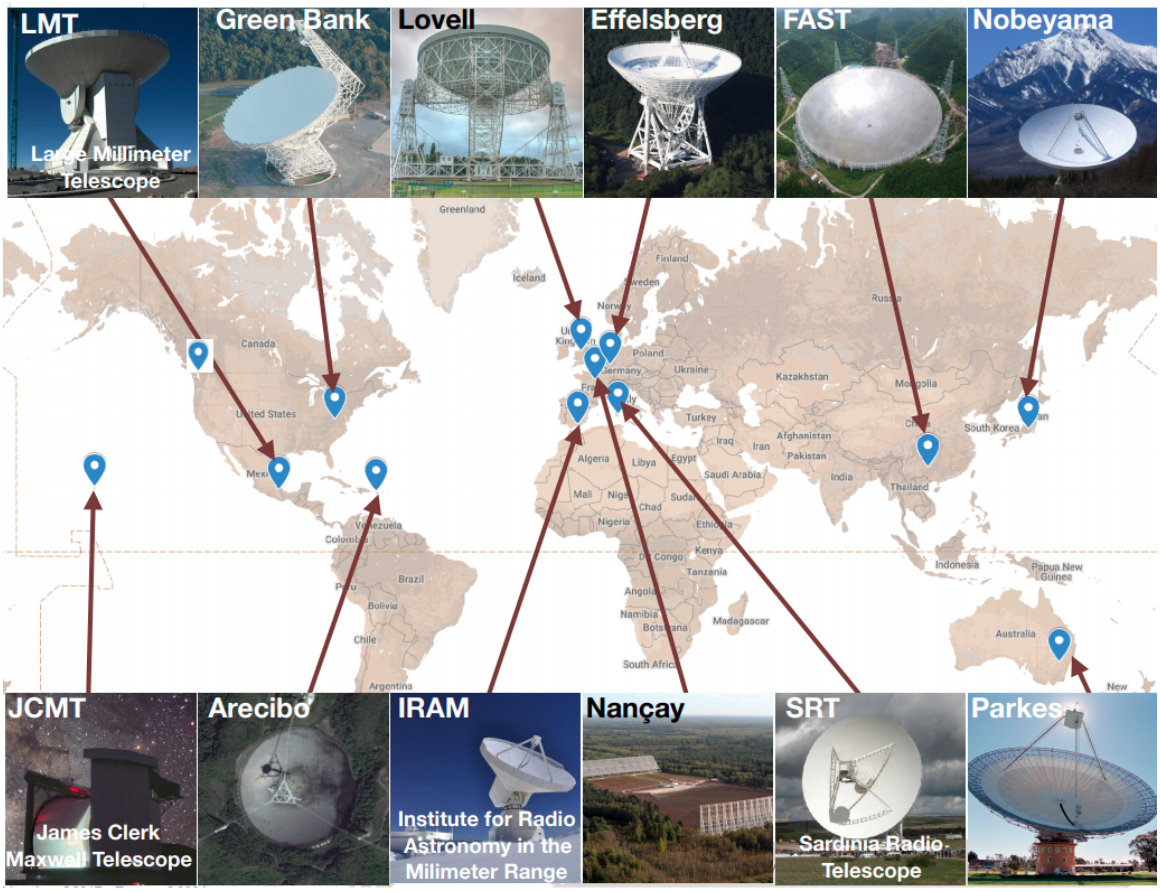


Figure 1.8. Some of the single-dish, long-wavelength, telescopes from around the world. Observed wavelengths for these features telescopes range from $\sim 1\,\mathrm{mm}$ up to just over $\sim 1\,\mathrm{m}$. This image was taken from the 2017 ATCA radio astronomy summer school and was created by Jennifer West (University of Toronto).

theoretical resolving power, a radio telescope would have to have a diameter on the order of around 500 km. A telescope like this is of course, practically impossible to construct on Earth. Therefore astronomers who wish to resolve small sources (such as external galaxies) use *interferometry* instead.

Radio interferometry first emerged in 1946, when Martin Ryle, Joseph Lade Pawsey, and Ruby Payne-Scott observed radio emission from space using sea-cliff interferometry. It is worth noting that they weren't the first to image the sky in the radio domain, this milestone belongs to Karl Jansky, who identified the Milky Way's radio emission in the 1930s (Brittain, 1984) and Grote Reber, who made detailed maps of the radio sky by the 1940s (Reber, 1940). World war II forced technical developments in the radio and radar work which led to the discovery of the Sun being a strong radio emitter. After the war, radio astronomy continued in groups in Britain, the Netherlands, and Australia and eventually became a key part of astronomy in all

areas of the globe.

Imagine two radio telescopes placed a distance **b** apart (see Figure 1.9), where **b** is also called the *baseline*. If an observed source is directly overhead (equidistant from the two receivers), the light waves will arrive at the telescopes in phase and at the same time. When the signals from the two telescope are combined, they constructively combine to give a strong signal. If the source has moved across the sky, parallel to an imaginary line connecting the two telescopes, then the path length of the light waves to the receivers will differ. If the path lengths differ by $\lambda/2$, they will arrive 180° out of phase and destructively interfere. So when observing the night sky, a two dish interferometer would observe a series of peaks and troughs –an interference pattern exactly like that as seen in Young's famous double slit experiment.

A key principle of interferometry is that the Fourier transform of the signal from a distant, incoherent source is equal to its complex visibility. This is called the Van Cittert-Zernike theorem (van Cittert, 1934):

$$T(x,y) = \int\int V(u,v)e^{-2\pi i(ux+vy)}dudv, \tag{1.2}$$

where T is is the 2-dimensional mutual coherence function between two points in the observation plane, V is the intensity of the source, and u and v are the x and y distances respectively, between the two observation points on the observation plane in units of wavelength.

Thus, an interferometer measures the Fourier transform of the light distribution of a source on the sky. The 'visibility function' $V(u,v)$, is measured as a collection of discrete points, where each baseline (and therefore pair of dishes) provides such a point. In order to build up the desired Fourier transform, many baselines are needed. Building up the Fourier transform corresponds to increasing the 'uv plane' sampling and coverage, where increasing the sampling requires increasing the number of baselines in an array and increasing the coverage depends on the array configuration (the arrangement of the dishes themselves). The greater the coverage, the better the quality of the resulting images. Because the Earth rotates with time, a point on the Earth's surface will appear to move over time from the point of view of the source. It is therefore possible to use the Earth's rotation as a way to fill the uv plane.

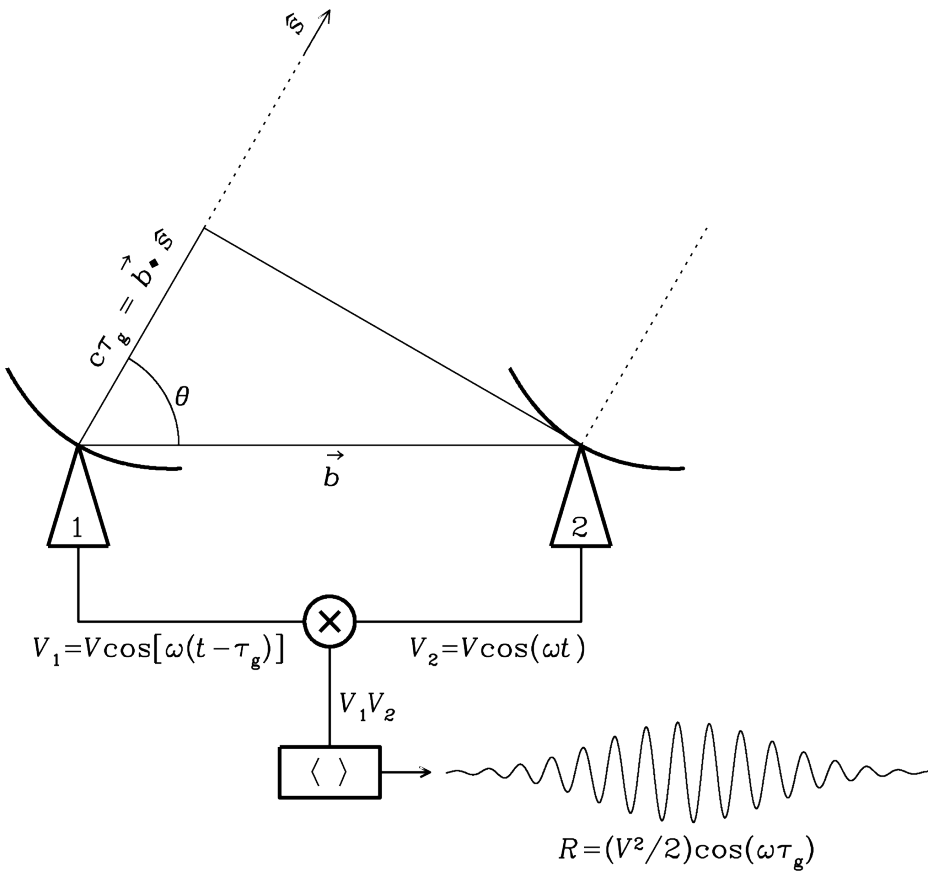


Figure 1.9. The geometry of a two dish interferometer, with antennas spaced at a distance **b** apart and aimed at a source in the direction **s**. Source signals are combined at a correlator and changed into a machine readable signal. Image taken from the NRAO course on interferometers: Section 3.7.1 at `https://www.cv.nrao.edu/~sransom/web/Ch3.html`.

If we consider a single baseline pair of antennas separated by a distance B (see Figure 1.9), each dish measures a voltage V of a point source as

$$V_1 = Ecos(\omega t) \tag{1.3}$$

$$V_2 = Ecos[\omega(t - \tau_g)], \tag{1.4}$$

where V_1 and V_2 are the voltages measured by each antenna, E is the electromagnetic field strength of the source, $\omega = 2\pi v$, where v corresponds to the observing frequency, t the time of signal arrival, and τ_g the geometric time delay between source detection at both receivers. τ_g depends on the position of the source relative to the array, as well as the baseline distance and is described by $\tau_g = \frac{\mathbf{b}\cdot\mathbf{s}}{c}$, where $\mathbf{b}$ is the baseline distance, $\mathbf{s}$ is a unit vector in the direction of the source from a receiver, and c is the

speed of light. Both signals are then combined in a 'correlator' which cross-correlates the two signals from the same E field to obtain a *response*

$$R = V_1 \otimes V_2 = \frac{1}{2}E^2[cos(\omega\tau_g) + cos(2\omega t - \omega\tau_g)]. \tag{1.5}$$

For extended astronomical sources (i.e. not point sources), Equation 1.5 changes to

$$R = \int\int I_v \cos\left(\omega\frac{\mathbf{b}\cdot\mathbf{s}}{c}\right) d\Omega, \tag{1.6}$$

where I_v is the source intensity at frequency v, and $d\Omega$ is the solid angle observed by the telescope.

It is worth noting that the interferometer described above assumes that the observed interference pattern is described by a cosine function. Being an even function, this means that the response R goes to zero for observed odd brightness distributions. In order to rectify this behaviour, we often use a second correlation using a sine function so that we minimise the loss of observed emission. Therefore, we can finally define the complex visibility, from passing through a 'complex correlator' as:

$$V = R_c - iR_s = Ae^{-i\phi} = \int\int I_v(s)e^{-2\pi iv\frac{\mathbf{b}\cdot\mathbf{s}}{c}} d\Omega, \tag{1.7}$$

where $A = \sqrt{R_c^2 + R_s^2}$ (the cosine and sine responses combined) and $\phi = \arctan\left(\frac{R_s}{R_c}\right)$. It is easy to recognise the resulting equation as a two-dimensional Fourier transform.

A modified version of Equation 1.1

$$\theta_{min} \approx \frac{\lambda}{\mathbf{b}_{max}}, \tag{1.8}$$

where $\mathbf{b}_{max}$ is the maximum baseline distance in an array, gives us not only the maximum resolving power of an interferometer, but also a *minimum* resolving power (given instead by $\mathbf{b}_{min}$, typically referred to as the *lowest angular scale*). Since two dishes can never be infinitely close to one another, in interferometer is always "blind" to extended structures beyond this maximum size. The emission of the source in these regions is 'resolved out'. Therefore, it has become popular in recent years to combine single dish with interferometric observations and, in some cases, utilise different array configurations/sizes to fill in the gaps.

Below I will discuss the two interferometers used in this thesis.

1.2.2 ALMA

Figure 1.10. The Atacama Large Millimeter/submillimeter Array (ALMA), located 5000m above sea level in the Chilean Andes. To date it is the most expensive ground-based telescope ever built. Image credit: ESO.

The Atacama Large Millimeter/submillimeter Array (ALMA, see Figure 1.10) is located on the Chajnantor plateau, roughly 500 m above sea level in the Chilean Andes. ALMA is an array of 66 12 m dishes spread out over distances of up to 16 km. In order to overcome the problem of 'resolving out' source emission (see § 1.2.1), ALMA also includes the 'Atacama Compact Array' (ACA or Morita Array). The ACA is comprised of twelve 7 m and four 12 m antennas at the heart of the larger array, the latter often being referred to as the 'Total Power (TP) Array'. Observations with both ACA and TP arrays are often carried out in addition to observations using the 12 m array.

ALMA was conceptualised in 1997 when the National Radio Astronomy Observatory (NRAO) and European Southern Observatory (ESO) agreed to combine the predecessors: the Millimeter Array (MMA, in the United States) and the Large Southern Array (LSA, in Europe), into one project with superior sensitivity and resolution. Later in the project, Japanese, Taiwanese, and Chilean partners joined, contributing to the construction of the ACA and additional receiver bands. At a cost

of $1.4 billion US dollars, ALMA is the most expensive ground-based telescope in operation (ESO 2017, until construction of the SKA completes at an estimated cost of 1.9 billion Euros).

With its unprecedented resolution and sensitivity in the millimetre regime, ALMA is sensitive to the coldest parts of the Universe i.e. dense molecular gas and dust. The science goals of ALMA include studying protoplanetary discs, star formation, spectral line emission in galaxies, and even the chemistry and structure of solar system bodies (Wootten & Thompson, 2009). In the work presented in this thesis, ALMA observations of CO in nearby galaxies helps in evaluating the use of machine learning models across a variety of instruments with variable spectral and spatial resolutions[†].

1.2.3 VLA

The Karl G. Jansky Very Large Array (VLA, Kellermann et al. 2020, see Figure 1.11) is another interferometer; this time, located on the plains of San Agustin in central New Mexico, at an altitude of 2124 m above sea level. The VLA is made up of twenty-eight 25 m dishes deployed in a Y-shaped array. This is in contrast to ALMA which can distribute its antennas freely. The antennas of the VLA are on tracks though, allowing the radius and density of the array to be changed in accordance with desired angular resolution and sensitivity requirements. The lengths of each of these tracks are 21 km (13 miles). As a centimetre-wavelength radio telescope, one of the primary missions of the VLA is to observe neutral hydrogen in galaxies beyond the Milky Way. A number of surveys, observing 21 cm line emission in galaxies have been conducted and proposed, including the Local Volume HI Survey (LVHIS) and THe HI Nearby Galaxy Survey (THINGS), both of which feature in Chapters 2 and 3 respectively.

The VLA is a component of the National Radio Astronomy Observatory (NRAO, NRAO 2022), which is itself a funded research center of the United States National Science Foundation. Construction of the array was completed in 1980, and at the time was the largest radio telescope in the world. The array underwent a major upgrade in 2011, enhancing the instrument's sensitivity, observable frequency range, and effective angular resolution. In the future, the Next Generation VLA

[†]It is worth noting that the term 'spatial resolution' is often rejected in interferometry. Instead astronomers commonly refer to terms related to the restoring beam calculated for the observation in question. The beam effectively acts as a quantified PSF, and more appropriately represents the resolving power of an interferometer at a given time.

Figure 1.11. The Karl G. Jansky Very Large Array, located around 80 km (50 miles) west of Socorro in New Mexico. The array is distinctive for its Y-shape configuration of antennas. Image credit: NRAO.

(ngVLA) is set to supersede the VLA. ngVLA will comprise 244 dishes of 18 m diameter, with baselines of up to 1000 km across the United States and extended baselines to continental scales of up to 8860 km. An additional 19 dishes of 6 m diameter will be included to make up a short baseline array (SBA) for the same purposes as the ACA. The ngVLA will complement SKA by observing at higher (with some overlap) frequencies, but with much higher spatial resolution (ngVLA, 2022).

As mentioned previously, data collected from the VLA is used in the work presented throughout this thesis. These are strictly HI 21 cm line observations, and play a critical role in evaluating the performance of machine learning models on data at the same frequency range the SKA will cover.

1.2.4 IFU INSTRUMENTS

Integral field spectroscopy (IFS) is the process of obtaining spatially resolved spectra in astronomy and other fields of research (including, but not limited to, medical imaging and earth observation). In astronomy we usually refer to integral field units (IFUs), which are combined with a spectrograph to make a multiplexed optical instrument capable of spectrographic and imaging capabilities.

IFS is an important area of astronomy (Gunn et al., 2006), owing to the need to study the optical spectra of extended objects as a function of spatial location, or of clusters of many sources distributed in a small field. In the past, such observations were carried out using long-slit spectrographs, where spectra are dispersed perpendicular to the slit. The spatially resolved component of the image is in the dimension along the slit, and a total image can be obtained by stepping the position of the slit along the extended object on the sky. This process is slow and fraught with uncertainties due to e.g. changing sky conditions during the observations. Hence the field has moved towards the use of integral field spectrographs, which can obtain spectra in a 2-dimensional field with a single pointing.

Both MaNGA and SAMI instruments, outlined in §1.2.5 and §1.2.6 respectively, use hexabundles of fibres as their IFUs. This is a grouping of optical fibres, which form the spectrograph's entrance. For each projected fibre on the sky source, a spectrum is obtained.

1.2.5 MANGA

The Mapping Nearby Galaxies at APO (MaNGA) survey, is one of the surveys carried out as part of the larger Sloan Digital Sky Survey (SDSS) project, aimed at measuring spectra across the face of $\sim 10\,000$ galaxies in the local Universe (Gunn et al., 2006). The MaNGA instrument comprises 29 fibre IFUs hexabundles, with fiber numbers ranging from 19 to 127, delivering a field of view ranging from 12″to 32″respectively. With 17 dedicated object IFUs, the MaNGA instrument can observe 17 sources at any one time, with 12 additional 'mini-bundles' used for calibration purposes. The IFUs are used to spectroscopically map sources in the wavelength range of $360 \rightarrow 1040$nm, at a resolution of ~ 2000 ($\lambda/\delta\lambda$). The MaNGA instrument is one of several used by the SDSS 2.5-m wide-angle optical telescope situated at Apache Point Observatory, in New Mexico (see Figure 1.12).

At a rest frame of 656 nm, the Hα emission line, along with many other optical lines (such as H$_\beta$, O[II], O[III], N[II] and S[II]), are captured by MaNGA and therefore the survey offers thousands of spatially and spectroscopically resolved velocity maps

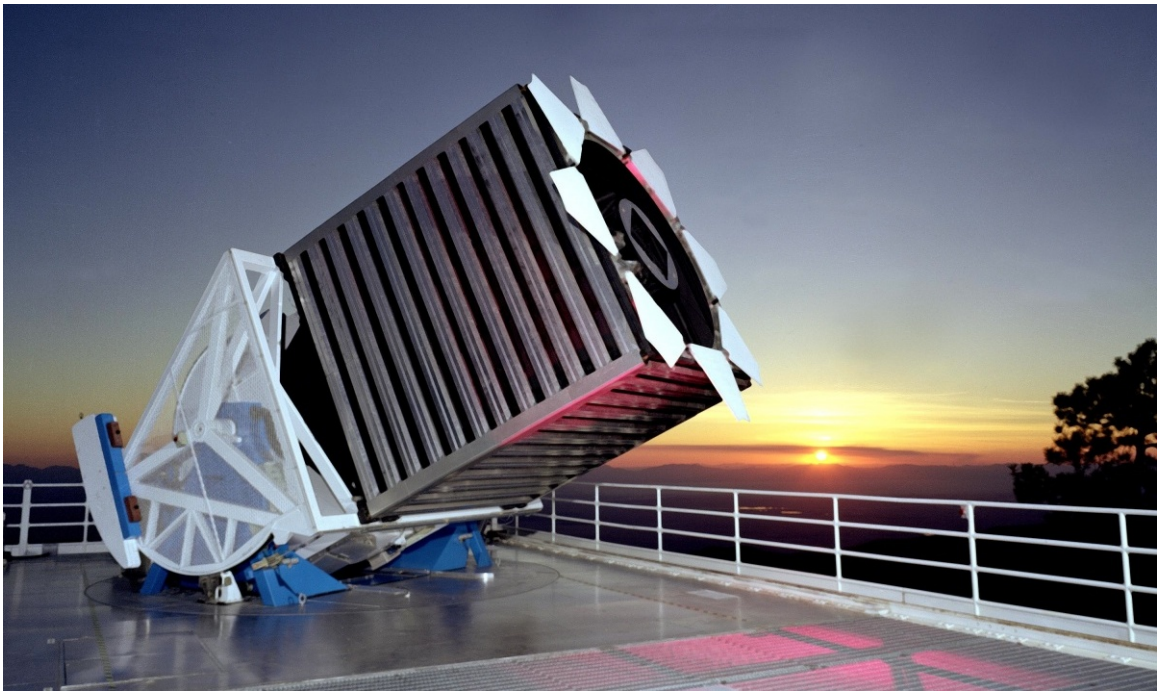


Figure 1.12. The SDSS 2.5-m wide-angle optical telescope situated at Apache Point Observatory, in New Mexico. The MaNGA IFU spectrograph is one of several instruments used by the telescope as part of SDSS. Image credit: SDSS.

to use. In the work presented in this thesis, we use MaNGA data in order to demonstrate the generalisation power of our machine learning models, which were originally targeted at interferometric data products. We use the data in combination with our models to recover known physical relationships with the largest sample sizes available and at the fastest speed to date (see §4).

1.2.6 SAMI

The Sydney-AAO Multi-object Integral-field unit (SAMI) is the IFU instrument that feeds the AAOmega spectrograph. SAMI is situated at the Anglo-Australian Telescope (AAT, see Figure 1.13) at the Australian Astronomical Observatory (AAO) in suburban Sydney, Australia. It is comprised of thirteen 16-fibre hexabundle IFUs. Each hexabundle has a 15″ field of view on the sky, and can be deployed by plates within a 1° field of view. Unlike MaNGA (see §1.2.5), SAMI has a set of three resolution gratings (labelled low, medium, and high), providing resolutions of R $\sim$ 1000, R $\sim$ 5000, and R $\sim$ 10000 ($\lambda/\delta\lambda$) across the instrument's wavelength range of 330 $\rightarrow$ 900nm.

Within the observable wavelength range, Hα line emission is captured at a greater resolution with SAMI than with MaNGA (and lines in the blue, such as

Figure 1.13. The Anglo-Australian Telescope (AAT) in suburban Sydney. Attached is the SAMI IFU instrument, used observe nearby galaxies as part of the SAMI survey. Image credit: AAO.

H_β/O[III] are even higher resolution still). For the work presented in this thesis, the overlap of targets in the SAMI and MaNGA surveys is large enough to allow for a comparison of novel machine learning model performance on the same sources. The overlap is also low enough (as a fraction of the survey size) to also boost the total IFU dataset size (used in the work presented in §4) allowing a direct comparison of model performance when using datasets of differing spectroscopic resolution.

1.3 Machine Learning

The field of machine learning pertains to the use of algorithms that can improve through experience and exposure to data; computers discovering how to perform tasks without explicit programming to do so. For simple tasks it's possible to tell a program how to execute all the steps needed to solve a problem. For more complex tasks it can be more difficult for humans to cover the necessary logic and programmatic steps to solve problems. In practice it can be more effective to help machines develop their own algorithms. Employing various methods of teaching computers to develop algorithms encompasses the field of machine learning.

The term *machine learning* has been popular since the 1950s, born out of an era in which pattern recognition was of major interest. Modern day machine learning applications primarily focus on classification and future prediction and are often mentioned along with *artificial intelligence* and *data mining.*

Machine learning grew out of the quest for artificial intelligence (AI), i.e. intelligence demonstrated by a machine as opposed to humans or animals. Machine learning only really began to flourish in the 1990s, when the field changed from its initial goal of achieving artificial intelligence to tackling practical, solvable, problems. Focus shifted away from approaches inherited in pure AI, to models and methods found in statistics, pattern recognition, and probability theory. Whether machine learning is still considered a branch of AI or a completely separate field is widely debated even today.

While machine learning focuses on making predictions based on information learned from data, data mining purely focuses on the discovery of unknown properties of data. In the modern world, data mining uses many machine learning methods but usually with different goals in mind. Conversely, machine learning also employs data mining methods as 'unsupervised learning' (these days, referred to as *self-supervised learning*), as a preprocessing step for classification or for complex data exploration. While machine learning inherits from data mining, and vice versa, they remain distinct due to their performance evaluation styles. A task aimed at reproducing known knowledge and predicting consequential future observations is well suited to supervised learning. Meanwhile, for typical data mining tasks, supervised methods cannot be used due to the unavailability of training data and instead, unsupervised learning methods are more favourable.

1.3.1 Machine learning approaches

Supervised learning

A supervised learning task is that of learning a function which maps an input vector to an output based on training input-output pairs. A supervised learning model learns feature representations of input vectors X necessary for mapping inputs into desired output values Y. The inferred function from training ideally can be used to correctly determine the class, label, or value for unseen input instances. This requires the model to generalise in a reasonable manner. Supervised learning problems can be grouped into regression and classification problems.

Many supervised learning algorithms exist, each with their own strengths and weaknesses. Factors to consider when choosing an appropriate algorithm for a learning task depend on the complexity of the target function and both the size and complexity of the training data. An obvious drawback of supervised learning is that it requires prior knowledge of the output space limits or number of target classes. Consequently, there is often a trade-off between predictive accuracy and induced bias when using supervised learning algorithms.

In a classification problem, output variables are categorical (e.g. object colour). Whereas, in a regression problem, the output variables Y have real continuous values (e.g. object height). An example supervised classification model might be a support vector machine (SVM, Cortes & Vapnik 1995) and an example supervised regression model might be a linear regression model. However, there do exist algorithms capable of performing both, such as random forest (Breiman, 2001) and neural networks (see §1.3.2).

Unsupervised (self-supervised) learning

In contrast to supervised learning, unsupervised learning algorithms are not provided labelled targets and as such must *self-discover* patterns in training set data. The removal of explicit target variables allows some unsupervised algorithms to find unknown patterns and relationships within datasets, with reduced workload in preparing the training data. However, disadvantages include overly-complex mappings of input data (making results difficult to interpret), increased computational load during training, and the unknown effects of bias due to erroneous or irrelevant training data.

Common families of unsupervised learning include: clustering, neural networks, anomaly detection, and latent variable models. All of these families have a degree of overlap which tie into the overarching goal of unsupervised learning – uncovering patterns and internal relationships within datasets with minimal oversight.

The *Kohonen self-organising map*, or SOM (Kohonen, 1990), is one such example of

unsupervised learning techniques, which first emerged in 1990 and is often seen as an entry level algorithm into the world of unsupervised learning. An exemplar use-case of Kohonen SOMs in astronomy is their application to variable star classification with the Next Generation Transit Survey (Armstrong et al., 2017). It primarily serves as a data visualisation tool at the time of writing but has been shown to work well with supervised learning techniques, boosting classification accuracy in machine learning models. Naturally, the combined use of supervised and unsupervised learning has been dubbed semi-supervised learning and shows great potential for improving the power of classification problems. Due to the entry level nature of the SOM and its past use in astronomy (e.g. Armstrong et al. 2016), is is worth dissecting the algorithm here to give a better understanding of the differences between more familiar supervised learning problems and unsupervised learning tasks.

The premise. Imagine we have a collection of feature vectors. These could be vectors of length 3, where each dimension represents a different feature such as colour, or shape. The motivation behind using a Kohonen SOM, is to train a 2D grid to group input vectors which share similar features, and consequently displace feature vectors which are very dissimilar away from one another. Each point on the grid therefore represents a vector with the same number of dimensions as the input feature vectors used to train the map.

The algorithm. To begin, a Kohonen layer is initialised, this is simply a randomised 2-dimensional grid of weights $\mathbf{w_{ij}}$. Each weight has a topological position in the Kohonen layer and is a vector of the same number of dimensions as each input vector. An input vector of features $\mathbf{x_n}$ is chosen at random from the input array list of length N and presented to the Kohonen layer. The best matching unit (BMU) in the Kohonen layer is found using a user specified distance metric. A popular metric is the Euclidean distance (see equation 1.9).

$$Dist = \sqrt{\sum_{n=1}^{n=N} (x_n - w_{ij})^2} \tag{1.9}$$

The weights w_{ij} of the Kohonen layer within a defined proximity of the BMU are updated at a time $t + 1$, as in equation 1.10.

$$w_{ij}(t+1) = w_{ij}(t) + \alpha(t)\beta(t)[x(t) - w_{ij}(t)] \tag{1.10}$$

$\alpha(t)$ is known as the learning rate and usually takes values: $0 < \alpha(t) < 1$.

$\beta(t)$ is known as the *neighbourhood function*, which controls the influence the BMU has on neighbouring weights, depending on their proximity to the BMU. Usually $\beta(t)$ is chosen to be some radially decaying function such as a Gaussian or decaying exponential. Together. $\alpha(t)$ and $\beta(t)$ make up what is known as the *adaption gain*, controlling the total decay of influence over a node's weight adjustment at each time step. This process is then repeated for $\mathbf{T}$ iterations until a user specified convergence condition is fulfilled.

These equations above show the underlying mathematical processes used in the SOM algorithm. The pseudo-code for algorithm itself is outlined in Algorithm 1.

Algorithm 1: Kohonen SOM

Data: Training inputs $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$
Input: Kohonen map with randomly initialised weights $\mathbf{m_{ij}}$
Input: Number of training epochs $\mathbf{T}$
Input: Learning rate α
Input: Neighbourhood function β
Input: Convergence criterion η
Result: Trained SOM $\mathbf{M}_{opt}\mathbf{(x)}$

```
   /* Iterate over all epochs                                    */
1  for t ∈ {1, 2, ..., T} do
       /* Do until number of trainsamples is reached             */
2      for n ∈ {1, 2, ..., N} do
           /* Choose current trainsample                         */
3          x_n ← choose random elements(X)
           /* Calculate best matching unit for input             */
4          b_ij ← BMU(x_n)
           /* Update neighbouring weights                        */
5          m_ij(t + 1) = m_ij(t) + α(t)β_ij[x_n,t − b_ij(t)]
           /* Current model has converged                        */
6          if η = True then
7              M_opt ← M
8          end
9      end
10 end
```

The result of training an SOM is a clustering of locally-similar weights and, once completed, test inputs are shown on an SOM by highlighting the test inputs' BMUs on the converged Kohonen layer. In Figure 1.14 we can see an exemplar SOM where the training inputs are normalised RGB vectors. It should be noted that although Figure 1.14 shows a 2-dimensional Kohonen layer, most applications of SOMs use a 1-dimensional adaption for speed while 2D Kohonen layers are used for visualisation purposes.

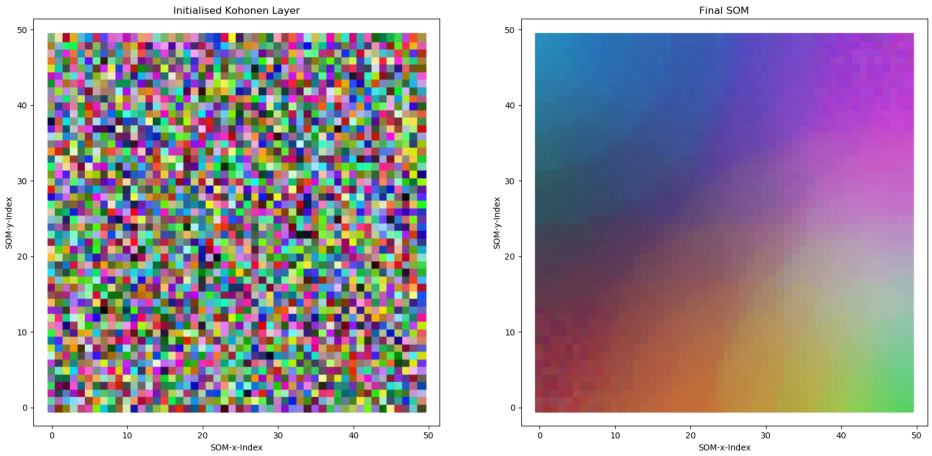


Figure 1.14. An example use of a Kohonen layer before (left) after (right) training on RGB pixels.

Deep learning

Deep learning is part of machine learning based on a special use case of artificial neural networks (see §1.3.2). The adjective 'deep' in deep learning refers to the use of more than one layer in the neural network. Whereas a single layered, linear perceptron (defined in §1.3.2) cannot be a universal classifier (i.e. a model capable of performing classification tasks perfectly regardless of the complexity of the task), a network with non-linear activation functions and a hidden layer of unbounded width can. Deep learning is a modern variation concerned with an unbounded number of hidden layers of bounded size. The limited layer sizes offer practicality and implementation while moving towards the power of a universal classifier. In deep learning, the layers are also permitted to deviate from connectionist models.

It is widely accepted that the deep learning revolution occurred in 2012, when fast implementations of convolutional neural networks (see §1.3.2) progressed the field of pattern recognition. Around this time, convolutional neural networks began

to supersede human performance in pattern recognition tasks (Krizhevsky et al., 2017) and deep learning would go on to play significant roles in advancing computer vision, image analysis, speech recognition, natural language processing, and even game design.

1.3.2 ARTIFICIAL NEURAL NETWORKS

An artificial neural network (ANN) is a collection of connected nodes, which are loosely derived from the neurons found in a biological brain. Each connection between nodes models the transmission of information like synapses in the brain. Although often compared to the hypothesised function and architecture of the brain, the true similarity between neural networks and the biological brain remains heavily debated and unsolved.

Perceptron

The perceptron (Rosenblatt, 1958) is a single layer artificial neural network (ANN). As such it can be used exclusively for linear regression or classification problems. A schematic for the perceptron is shown in Fig. 1.15.

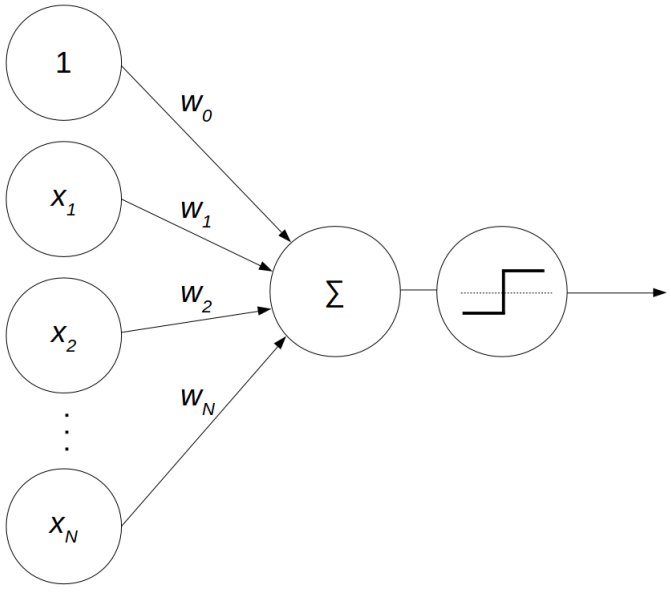


Figure 1.15. Diagram of a perceptron featuring: inputs, weights, summation function, and activation function.

When training a perceptron, inputs $(x_1, x_2 ... x_N)$, are presented to the model. These input vectors are multiplied by weights $(w_1, w_2 ... w_N)$ and summed with a bias

weight w_0. The summed value is then passed to an activation function which maps the result to some desire range such (e.g. $0 \to 1$ or $-1 \to 1$). A likelihood function is used to quantify the error in predicted output, based on X-Y input-output pairs, and minimised by updating the weights through the gradient descent method. In practice perceptrons are rarely used and therefore, for an outline of the training algorithm for a perceptron, one can simply use a single-hidden-layer adaption of Algorithm **??**.

In the early days (1950s) when the perceptron was first created, computers were deemed too unsophisticated to handle the long runtime required by larger neural networks. Coupled with the perceptron's inability to solve the *exclusive-or problem* (XOR), neural network development did not become popular again until advances in computation power and the backpropogation algorithm (Rumelhart et al., 1988) in the 1970s. And it wasn't until the development of advanced graphics processing units in the turn of the century that we began to witness the emergence of trainable deep learning algorithms, as perceptrons grew into deep neural networks.

Convolutional neural networks

Convolutional neural networks (CNNs) are a special class of deep neural network used primarily with multi-channel input matrices, or images, rather than 1D vectors. Information is derived from raw pixels, negating the need for a separate feature extraction stage; this results in latent models with increased empirical accuracy. Today, they are used for a range of problems from image classification, remote sensing, and self-driving cars.

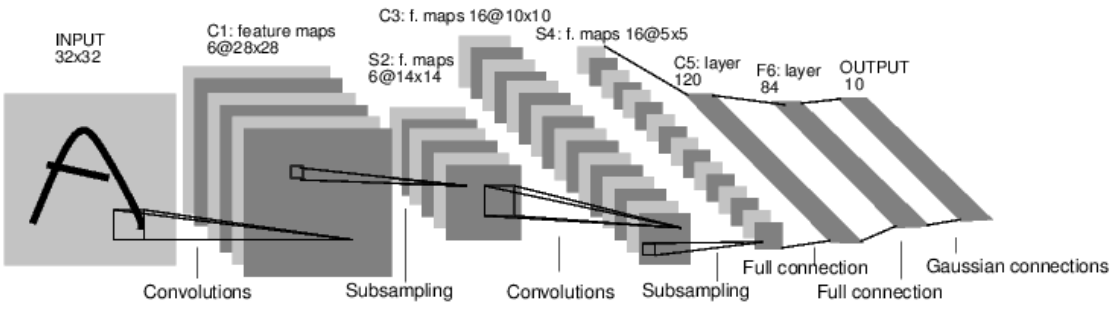


Figure 1.16. Architecture of the CNN *LeNet-5* featuring convolutional layers, max-pooling, and fully connected layers (Lecun et al., 1998).

A conventional CNN will have a series of layers often including: input, convolutions, max-pooling, activations, fully connected layers, and output (see Fig. 1.16) as well as regularisation techniques to improve generalisation.

In deep learning, many convolutions are applied where consecutive convolutions draw higher level features. These networks are only trainable (through back propagation) thanks to the use of advanced graphics processing units (GPUs, Steinkraus et al. 2005). Arguably, the most notable CNN to utilise GPU capabilities in the history of machine learning is AlexNet (Krizhevsky et al., 2017). This network achieved a 15.3% top-5 error rate (a performance metric measuring the number of times the true target variable appears in the top-5 most likely predicted variables by a network for a given input) when trained on the ImageNet LSVRC-2010 dataset[†] which was a huge step forward from previous approaches to the classification problem. The contest involved training on 1.2 million images with 1000 different classes. Since then, there have been various network architectures outperforming *AlexNet* including: *GoogleNet*, *VGGNet*, *ZFNet* and *ResNet*. *ResNet* boasts the current best top-5 error rate of 3.57% (He et al., 2015) which surpasses the Human top-5 error rate, reported to be 5.1% (Russakovsky et al., 2014; Bhandare et al., 2016).

A CNN will undergo training in similar fashion to a standard neural network (consisting of multiple hidden layers to differentiate it from the perceptron), with the slight difference of backpropogating errors through the convolution kernels.

A number of mathematical processes, data handling techniques, and general machine learning training terms are often associated with using CNNs. Here they are outlined in order to better understand some of the nuances, goals, and difficulties associated with training a CNN model.

Overfitting is a term used when a modelling function is too closely aligned to a limited set of datapoints as to cause the model to only be useful in reference to its initial data set. Another way to describe an overfitting model, is one which has more parameters (and therefore higher complexity) than can be justified by the data. In essence, overfitting is to have unknowingly modelled the variation (i.e. noise) in the dataset as if it represents the underlying model structure. In extreme cases, models can memorise data in its entirety, if the number of model parameters is the same or greater than the number of observations. Such models will typically fail to make new predictions and thus have low *generalisation* power. A number of approaches exist to minimise overfitting; in the case of neural networks, which can often have millions of parameters, these approaches are often labelled as *regularisation* procedures which are described further on in this section.

[†]The ImageNet Large Scale Visual Recognition Challenge (ILSVRC) is a contest in which algorithms are evaluated on object detection and image classification at large scale. A key motivation for the challenge is to allow researchers to compare progress in detection across a wider variety of objects – taking advantage of the quite expensive labelling effort.

Max pooling is often used after or between convolutional layers. It acts to reduce the spatial sizes of layer outputs and therefore decreases the number of parameters of need to be optimised. This has the added benefit of reducing overfitting. A max pooling layer will record the maximum values in a matrix grid-wise striding. The stride parameter controls whether the pooling is *general* or *overlapping*, where *general* pooling will have a filter size equal to the stride length whereas *overlapping* pooling will have stride lengths shorter than the filter size. Pooling with a stride length larger than the filter size is considered too destructive and is rarely employed. There are other types of pooling, namely *average* and *L2-norm* pooling, however none have been shown to work as effectively as max pooling.

Activation functions map the output of CNN layers to lie between a desired range, depending on the function. Activation functions generally come in two types: linear and non-linear. Non-linear functions are popular in CNN architectures thanks to their ability to allow easier generalisation of models. The most popular activation functions are: sigmoid, tanh, softmax, and the Rectified Linear Unit (ReLU). ReLU stands as the most used activation function at the time of writing thanks to its success in CNNs (see Fig. 1.17) and maps outputs to lie between $0 \rightarrow \infty$.

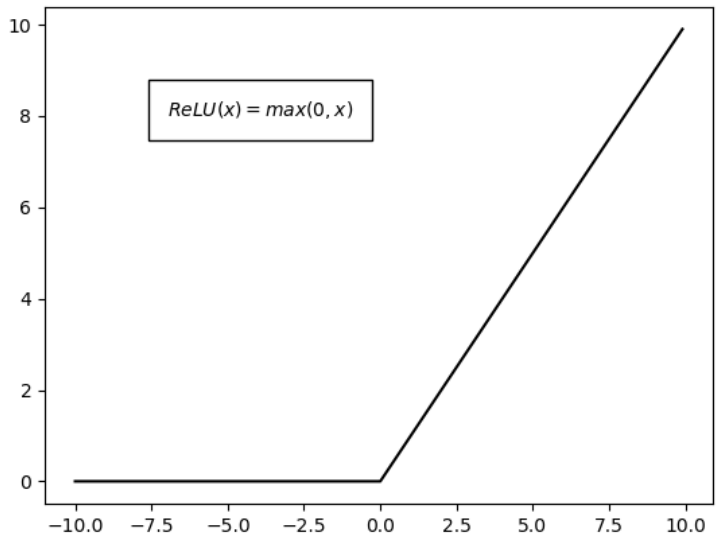


Figure 1.17. ReLU activation function. For a given unit-less input (x-axis), the mapped unit-less result (y-axis) is always real and positive.

Despite its status as the most used activation funtcion, ReLU functions suffer

from aggressive zeroing of negative values, depending on the input dataset. ReLU is still considered a state of the art activation function, but as such, is gradually being replaced by *leaky ReLU* whose zeroed negative values become a variable gradient function.

Regularisation is the process of reducing the generalising error of an algorithm while not affecting the training error rate. The problem of overfitting to training data is all too common in machine learning, where a dataset is considerably smaller than the number of trainable parameters in an algorithm. A deep learning image classifier can have millions of such parameters, therefore several tricks are commonly employed to regularise networks, forcing them to have greater generalisation power.

Data Augmentation. A very popular technique for regularisation is to alter the dataset in such a way that a network will see the same input differently through each augmentation and therefore has the effect of pseudo-dataset-enlargement. Augmentation can include convolving the image, applying noise, rotating or transforming the image, and more.

Dropout regularisation. A brute force method for encouraging generalisation, dropout layers will randomly assign node weights to have a value of zero with some predefined probability. This minimises the co-adaption of nodes and one can think of dropout regularisation as a way of generating many thinned networks which share parameters, closer resembling an ensemble method (Srivastava et al., 2014). Several machine learning models, including CNNs, suffer from misclassifying *adversarial examples*. Minor perturbations to input images, with no discernible difference to the human eye, are perceived differently by models posing a large trust issue when classifying unseen data. This resulted in the rise of *adversarial training*, where networks are exposed to *adversarial examples* and corrected accordingly. Such training is said to outperform dropout regularisation (Goodfellow et al., 2014) in some cases.

Early stopping. If one can find the moment at which a network undergoing training begins to overfit, it is possible to mark this point as the network's best generalising epoch. This is often found by evaluating a validation set (which the network never trains on) and looking for the turning point of the validation accuracy. The turning point is indicative of the network fitting to noise in the dataset. Halting the network training procedure at this point results in the best generalising hyper-parameter optimisation without employing any other regularisation techniques.

Deep learning in practice

There are many pretrained deep learning architectures, written using different Application Programming Interfaces (APIs), and trained on different data available for anyone to access and manipulate for free on the internet. As will become evident in the following chapters, the work presented in this thesis makes use of neural networks developed *from scratch* and written using the PyTorch API. In this short section I will outline the reasons for this and the benefits from doing so.

The choice of deep learning API isn't a hugely important one when creating and using neural networks, but some are easier to get started with than others. For a complete novice, it can be argued that `TensorFlow` (and by extension `Keras`, it's baby relative) is the most beginner-friendly API tool for creating and training deep neural networks. This is due to `TensorFlow` being one of the first scalable APIs for neural network creation, meaning it has a long history of community support and development. Conversely, `PyTorch` is relatively nascent in comparison, so why would one elect to use `PyTorch` over its more developed predecessor? I made the conscious choice to use `PyTorch` throughout this thesis because it is far more flexible in neural network creation and more efficient in memory usage. The former point is thanks to `PyTorch`'s use of dynamic computational graphs, i.e. the internal mathematics of the network (as inputs are passed through them) are calculated in the moment –from layer to layer. This makes pinpointing errors in your network fast and also allows for breaking up and building branched networks with relative ease. Despite the longer learning curve of getting up and running with `PyTorch`, once complete it quickly becomes clear just how much more *pythonic* and flexible it is.

With regard to `PyTorch`'s community forum and documentation, it used to be (at the beginning of the PhD certainly) that the level of detail in `PyTorch`'s official documentation and the activity on the community forums was in a word... lacking. However, since then `PyTorch` has exploded in popularity thanks to the flexibility mentioned above, and now has much more comprehensive documentation and an active community user base with licensed administrators on hand to answer questions quickly.

For the majority of this thesis, I go on to talk about the use of GPU accelerated neural network creation. An important factor of deep learning APIs is that they are designed to make use of the computational speed-up from performing tensor operations on GPUs. All APIs designed for deep learning can operate in CPU or GPU mode, with the former being strongly discouraged if significant image task results are

needed within human lifetimes. Casting data to GPUs and training neural networks is easy in all modern deep learning APIs, yet `PyTorch` does so in a way which is much less memory intensive than other APIs. This makes prototyping new networks more energy efficient, and scalable for future use on high performance computing stacks.

When data volumes increase in the future and tasks requiring automated, high-dimensional feature extraction and/or inference must cope with vastly increased data collection rates, then it makes sense to use the most efficient tools for the jobs at hand. Throughout my PhD, automated image classification tasks using deep convolutional neural networks fit the bill for solving the problems described in the following chapters, and the `PyTorch` API allowed for that progress to be made.

1.3.3 Applications of machine learning in astronomy

With astronomy pushing into the *big data era*, machine learning is now beginning to permeate into most fields within astronomy. Gaia (Gaia Collaboration et al., 2021) is charting a 3-dimensional map of the Galaxy and local group, providing accurate positional information for billions of stars, the perfect stomping ground for applying machine learning. In exoplanet studies, machine learning has been used to recover hidden transit signals in surveys of millions of stars since the Kepler (and K2) mission, through to the current TESS mission and even the future ARIEL mission (Nikolaou et al., 2020). Even the detection of gravitational waves (which only truly became a proven practical field in the last decade) is now aided by machine learning (Shen et al., 2017; George & Huerta, 2018; Gabbard et al., 2018; Zevin et al., 2017).

It could be argued that extragalactic astronomy entered the *big data era* around 2007 with sky surveys such as the Sloan Digital Sky Survey (SDSS), which provided the astronomical community with multi-colour images of $\sim$1/3 of the sky as well as high resolution spectra of millions of Galactic and extra-galactic sources. This lead to the Galaxy Zoo project (Schawinski et al., 2007) –a citizen science project. Hundreds of thousands of public participants took part in labelling galaxies for the project, classifying them visually by morphology. This was a great success, and did reduce the near impossible task of an individual classifying the sources alone. However, it also highlighted the need for revolutionising data analysis techniques, even back in 2007-2009. Since then, the Galaxy Zoo project has used those early public classifications as information to evaluate machine learning models attempting the same job (Walmsley et al., 2019). This marks the beginning of machine learning being applied to such large, observational, extragalactic datasets. In time, citizen science will not be able to support the data influx from the multiple next generation telescopes and

thus the evolution of data handling/filtering to machine learning approaches instead, seems rather logical.

However, the use of interferometers, for studying galaxies beyond the Milky Way, is still a recent and evolving technology in itself. No extragalactic surveys, of substantial size to compare to those expected from instruments like SKA, exist, and those large enough for machine learning studies are very recent. Traditionally interferometers have been point and shoot instruments, limited to observing single sources and producing raw data products of enormous size. This may explain why the application of machine learning to interferometric, extragalactic data is a nascent area of research, limited to only a handful of published works (e.g. Scaife & Porter 2021; Wu et al. 2019; Alger et al. 2018; Lukic et al. 2018; Tang et al. 2019).

1.4 Thesis outline

This thesis presents an in-depth study of applying machine learning to solve *big data era* problems in extragalactic astronomy. In particular, the work focuses on exploring an area of extragalactic astronomy previously untouched by machine learning techniques –fast kinematic analysis of galaxies observed using interferometers. This includes the use of mock observational data synthesised from hydrodynamical simulations, observations of cold molecular gas, as well as HI. The latter is the main driver of the thesis, as the overarching motivation behind the work presented here is to better prepare for next-generation interferometric surveys using instruments such as the SKA.

The key questions addressed in this thesis are "*Is it possible to use machine learning to filter observations of cold molecular gas in galaxies into ordered versus disordered classifications?*", which is then built upon by asking "*can machine learning be used to predict kinematic properties while reducing problems associated with self-contained training datasets?*". The latter then goes on to inspire the final question, "*can these models be used to derive known physical relationships on larger scales than previously achieved using non machine learning approaches?*". While this final question takes us away from interferometric data, and into the realm of IFU surveys, it plays off the aims of answering the second question. Specifically, in solving the 'transfer learning problem' (using pretrained models on unseen datasets, often with very different characteristics e.g. noise) it opens another question: "*can these models be used with data products from non-interferometric surveys too?*" These questions are addressed by the following chapters:

Chapter 2: "Using machine learning to study the kinematics of cold gas in galaxies" (an adaption of Dawson et al. 2019)

Chapter 3: "A self-supervised, physics-aware, Bayesian neural network architecture for modelling galaxy emission-line kinematics" (an adaption of Dawson et al. 2021a)

Chapter 4: "The stellar mass Tully-Fisher relation with SAMI & MaNGA using self-supervised, physics-aware, Bayesian neural networks" (an adaption of Dawson et al. in prep)

Chapter 5: A summary of how this thesis fits into the bigger scientific picture of today's knowledge of machine learning in extra-galactic astronomy, as well as related, ongoing, and proposed work.

Chapter 2
Using machine learning to study the kinematics of cold gas in galaxies

> "As with most of life's problems, this one can be solved by a box of pure radiation."
>
> Andy Weir, *"The Martian"*

In this chapter I introduce the first steps in applying deep learning to interferometric dataproducts, in an attempt to begin addressing their suitability as techniques in the future of extragalactic astronomy. Next generation interferometers, such as the Square Kilometre Array, are set to obtain vast quantities of information about the kinematics of cold gas in galaxies. Given the volume of data produced by such facilities astronomers will need fast, reliable, tools to informatively filter and classify incoming data in real time. In this chapter, I use machine learning techniques with a hydrodynamical simulation training set to predict the kinematic behaviour of cold gas in galaxies and test these models on both simulated and real interferometric data. Using the power of a convolutional autoencoder I embed kinematic features, unattainable by the human eye or standard tools, into a three-dimensional space and discriminate between disturbed and regularly rotating cold gas structures. My simple binary classifier predicts the circularity of noiseless, simulated, galaxies with a recall of 85% and performs as expected on observational CO and HI velocity maps, with a heuristic accuracy of 95%. The model output exhibits predictable behaviour when varying the level of noise added to the input data and I am able to explain the roles of all dimensions of my mapped space. My models also allow fast predictions of input galaxies' position angles with a 1σ uncertainty range of $\pm 17°$ to $\pm 23°$ (for

galaxies with inclinations of 82.5° to 32.5°, respectively), which may be useful for initial parameterisation in kinematic modelling samplers. Machine learning models, such as the one outlined in this chapter, may be adapted for SKA science usage in the near future.

2.1 INTRODUCTION

The age of *Big Data* is now upon us; with the **S**quare **K**ilometre **A**rray (SKA) and **L**arge **S**ynoptic **S**urvey **T**elescope (LSST) both set to see first light in the mid-2020's.

A key area for *big data* in the next decades will be the studying of the kinematics of cold gas in galaxies beyond our own. This field will rely on interferometers, such as the SKA, thanks to their ability to reveal the morphology and kinematics of the cold gas at high spatial and spectral resolution. Current instruments like the **A**tacama **L**arge **M**illimeter/submillimeter **A**rray (ALMA) have revolutionised the study of gas in galaxies with their sensitive, high resolution, observations of gas kinematics. However, this field lacks the benefits afforded by fast survey instruments, having long been in an era of *point and shoot* astronomy. As such, large datasets capable of containing global statistics in this research domain have yet to emerge and studies are plagued by slow analytical methods with high user-involvement.

At the time of writing, large-scale radio interferometric surveys such as WALLABY (Duffy et al. 2012) and APERTIF (Oosterloo et al. 2010) are set to begin and will motivate the creation of tools that are scalable to survey requirements. However, these tools will be insufficient for screening objects come the advent of next-generation instruments which are set to receive enormous quantities of data, so large in fact that storing raw data becomes impossible.

In recent times, disc instabilities, feedback, and major/minor mergers have become favoured mechanisms for morphological evolution of galaxies (e.g. Parry et al. 2009; Bournaud et al. 2011; Sales et al. 2012), the effects of which are visible in their gas kinematics. Therefore, gas kinematics could be used to rapidly identify interesting structures and events suitable for understanding drivers of galaxy evolution (e.g. Diaz et al. 2019). If the kinematics of galaxies can accurately yield information on feedback processes and major/minor merger rates, then astronomers using next generation instruments could develop a better understanding of which mechanisms dominate changes in star formation properties and morphology of galaxies. In order to do this we must develop fast, robust, kinematic classifiers.

Recently, machine learning (ML) has been used successfully in astronomy for

a range of tasks including gravitational wave detection (e.g. Shen et al. 2017; George & Huerta 2018; Gabbard et al. 2018; Zevin et al. 2017), exoplanet detection (e.g. Shallue & Vanderburg 2018), analysing photometric light curve image sequences (e.g. Carrasco-Davis et al. 2018), and used extensively in studies of galaxies (e.g. Domínguez Sánchez et al. 2018a,b; Dieleman et al. 2015; Ackermann et al. 2018, Bekki 2019).

While using ML requires large data acquisition, training time, resources and the possibility of results that are difficult to interpret, the advantages of using ML techniques over standard tools include (but are not limited to) increased test speed, higher empirical accuracy, and the removal of user-bias. These are all ideal qualities which suit tool-kits for tackling hyper-large datasets. However, the use of ML on longer wavelength millimetre and radio galaxy sources has been absent, with the exception of a few test cases (e.g. Alger et al. 2018; Ma et al. 2018; Andrianomena et al. 2019), with the use of such tests to study the gas kinematics of galaxies being non-existent. It is therefore possible that, in the age of big data, studying gas kinematics with ML could stand as a tool for improving interferometric survey pipelines and encouraging research into this field before the advent of the SKA.

Cold gas in galaxies that is unperturbed by environmental or internal effects will relax in a few dynamical times. In this state, the gas forms a flat disc, rotating in circular orbits about some centre of potential, to conserve angular momentum. Any disturbance to the gas causes a deviation from this relaxed state and can be observed in the galaxy's kinematics. Ideally therefore, one would like to be able to determine the amount of kinetic energy of the gas invested in circular rotation (the so called *circularity* of the gas; Sales et al. 2012). Unfortunately this cannot be done empirically from observations because an exact calculation of circularity requires full six-dimensional information pertaining to the three-dimensional positions and velocities of a galaxy's constituent components. Instead, in the past, astronomers have used approaches such as radial and Fourier fitting routines (e.g. Spekkens & Sellwood 2007, Krajnović et al. 2006, Bloom et al. 2017a) or 2D power spectrum analyses (e.g. Grand et al. 2015) to determine the kinematic regularity of gas velocity fields.

In this work I use a ML model, called a convolutional autoencoder, and a hydrodynamical simulation training set to predict the circularity of the cold interstellar medium in galaxies. I test my resulting model on both simulated test data and real interferometric observations. I use the power of convolutional neural networks to identify features unattainable by the human eye or standard tools and discriminate between levels of kinematic disorder of galaxies. With this in mind, I create a binary

classifier to predict whether the cold gas in galaxies exhibit dispersion dominated or disk dominated rotation in order to maximise the recall of rare galaxies with disturbed cold gas.

In §2.1.1 I provide the necessary background information for understanding what ML models I use throughout this chapter. In §2.2.1 I describe the measuring of kinematic regularity of gas in galaxies and how it motivates the use of ML in my work. In §2.2 I outline my preparation of simulated galaxies into a learnable training set as well as the ML methods used to predict corresponding gas kinematics. In §2.3 the results of the training process are presented and discussed with a variety observational test cases. Finally, in §2.4 I explain my conclusions and propose further avenues of research.

2.1.1 Background to convolutional autoencoders

Convolutional neural networks (CNNs), originally named *neocognitrons* during their infancy (Fukushima 1980), are a special class of neural network (NN) used primarily for classifying multi-channel input matrices, or images. Information is derived from raw pixels, negating the need for a user-involved feature extraction stage; the result being a hyperparametric model with high empirical accuracy. Today, they are used for a range of problems from medical imaging to driverless cars.

A conventional CNN can have any number of layers (and costly operations) including convolutions, max-pooling, activations, fully connected layers, and outputs and often utilise regularisation techniques to reduce overfitting. For a more in depth background to the internal operations of CNNs I refer the reader to Krizhevsky et al. (2012). These networks are only trainable (through back propagation) thanks to the use of modern graphics processing units (GPUs; Steinkraus et al. 2005). It is because of access to technology such as GPUs that I am able to explore the use of ML in a preparatory fashion for instrument science with the SKA in this chapter.

A CNN will train on data by minimising the loss between sampled input images and target variables. Should training require sampling from a very large dataset, training on batches of inputs (also called *mini-batches*) can help speed up training times by averaging the loss between input and target over a larger sample of inputs. Should the network stagnate in minimising the loss, reducing the learning rate can help the network explore a minimum over the parameter space of learnable weights and thus increase the training accuracy. Both of the aforementioned changes to the standard CNN training procedure are used in my models throughout this chapter.

An autoencoder is a model composed of two subnets, an *encoder* and a *decoder*. Unlike a standard CNN, during training, an autoencoder learns to reduce the difference between input and output vectors rather than the difference between output vector and target label (whether this be a continuous or categorical set of target classes). In an *undercomplete autoencoder* the encoder subnet extracts features and reduces input images to a constrained number of nodes. This so-called *bottleneck* forces the network to embed useful information about the input images into a non-linear manifold from which the decoder subnet reconstructs the input images and is scored against the input image using a loss function. With this in mind, the autoencoder works similar to a powerful nonlinear generalisation of principle component analysis (PCA), but rather than attempting to find a lower dimensional hyperplane, the model finds a continuous nonlinear latent surface on which the data best lies.

Autoencoders have been used, recently, in extra-galactic astronomy for deblending sources (Reiman & Göhre 2019) and image generation of active galactic nuclei (AGN; Ma et al. 2018).

A convolutional autoencoder (CAE) is very similar to a standard autoencoder but the encoder is replaced with a CNN feature extraction subnet and the decoder is replaced with a transposed convolution subnet. This allows images to be passed to the CAE rather than 1D vectors and can help interpret extracted features through direct 2D visualisation of the convolution filters. For an intuitive explanation of transposed convolutions I direct the reader to Dumoulin & Visin (2016) but for this chapter I simply describe a transpose convolution as a reverse, one-to-many, convolution.

2.2 Methodology

2.2.1 Circularity parameter

As described previously, in order to find and classify kinematic disturbances one would like to measure the circularity of a galaxy's gas disc. For an object composed of point sources (e.g. molecular clouds, stars, etc.), with known positions, masses, and velocities, the circularity measure

$$\kappa = \frac{\mathrm{K_{rot}}}{\mathrm{K}} \text{ where } \mathrm{K_{rot}} = \sum_{\mathrm{i}=1}^{\mathrm{N}} \frac{1}{2} \mathrm{m_i} \left(\frac{\mathrm{j_{z,i}}}{\mathrm{R_i}} \right)^2 \text{ and } \mathrm{K} = \sum_{\mathrm{i}=1}^{\mathrm{N}} \frac{1}{2} \mathrm{m_i v_i^2}, \tag{2.1}$$

analyses the fraction of kinetic energy invested in circular, ordered, rotation (Sales et al. 2012). Here, $\mathrm{K_{rot}}$ is a measure of the rotational kinetic energy about some axis and K is the total kinetic energy of the object. m, j, R, and v represent the mass,

specific angular momentum, radius from the centre of rotation, and velocity of each point in an object respectively. Objects with perfectly circular, disk like, rotation have $\kappa = 1$, while objects with either entirely random motion or no motion at all have $\kappa = 0$.

As κ can only be calculated empirically from simulated galaxies, combining ML techniques with simulations will allow us to explore their abilities to learn features that can be used to recover κ in observations faster, and more robustly, than by human eye. In fact, κ has been used in previous studies to infer the origin of galaxy stellar morphologies (Sales et al. 2012) and, more recently, to investigate the kinematics of gas in post starburst galaxies (Davis et al. 2019).

2.2.2 EAGLE

The *Evolution and Assembly of GaLaxies and their Environments* (EAGLE) project[1] is a collection of cosmological hydrodynamical simulations which follow the evolution of galaxies and black holes in a closed volume Λ cold dark matter (ΛCDM) universe. The simulations boast subgrid models which account for physical processes below a known resolution limit (Schaye et al. 2015; Crain et al. 2015; The EAGLE team 2017). These simulations are able to reproduce high levels of agreement with a range of galaxy properties which take place below their resolution limits (see e.g. Schaye et al. 2015). Each simulation was conducted using smooth particle hydrodynamics, meaning users can directly work with the simulated data in the form of particles, whose properties are stored in output files and a database that can be queried.

In this chapter I make use of these simulations, in conjunction with kinematic modelling tools, to generate a learnable training set. I then probe the use of this training set for transfer learning with the primary goal being to recover kinematic features from generated velocity maps. Using simulations has certain advantages over collecting real data including accessibility, larger sample sizes, and the ability to calculate empirical truths from the data. However, there are drawbacks, including: unproven model assumptions, imperfect physics, and trade-off between resolution and sample size due to computational constraints.

The scripts for reading in data, from the EAGLE project database, were adapted versions of the EAGLE team's pre-written scripts[2]. The original simulations are saved into twenty-nine snapshots for redshifts z = 0-20 and for this work

[1] `http://icc.dur.ac.uk/Eagle/`

[2] `https://github.com/jchelly/read_eagle`

I utilise snapshot 28 for `RefL0025N0376` and `RefL050N0752` and snapshots 28, 27, 26, and 25 for `RefL0100N1504` (i.e. redshifts z = 0-0.27). When selecting galaxies from these snapshots, I set lower limits on the total gas mass ($> 1 \times 10^9$ $M_\odot$) and stellar mass ($> 5 \times 10^9$ $M_\odot$) within an aperture size of 30 kpc around each galaxy's centre of potential (i.e. the position of the most bound particle considering all mass components), in order to exclude dwarf galaxies. In order to select particles which are representative of cold, dense, molecular gas capable of star formation, I only accepted particles with a SFR > 0 for pre-processing (as described in §2.2.3). There are many ways to select cold gas in the EAGLE simulations (Lagos et al. 2015) but I use this method for its simplicity as my primary goal is to create a model that is capable of learning low-level kinematic features so as to generalise well in transfer learning tests. The upper radial limit for particle selection of 30 kpc, from the centre of potential, is in keeping with the scales over which interferometers, such as ALMA, typically observe low-redshift galaxies. It is important that I replicate these scales in order to test my model performance with real data as described in §2.3.3. One should note that for future survey instruments, such as the SKA, an alternative scaling via consideration of noise thresholds would be more appropriate. However, as I am particularly interested in the performance of my models with ALMA observations, I instead impose a radial limit for this work. At this stage I also set a lower limit on the number of particles within the 30 kpc aperture to > 200. This was to ensure I had enough particles to calculate statistically valid kinematic properties of the galaxies and reduce scaling issues caused by clipping pixels with low brightness when generating velocity maps. With these selection criteria, I work with a set of 14,846 simulated galaxies.

2.2.3 Data preparation

Each galaxy was rotated so that their total angular momentum vector was aligned with the positive z-axis using the centre of potential (as defined in the EAGLE Database, see The EAGLE team 2017) as the origin. I then made use of the `Python` based kinematic simulator `KinMS`[3] (KINematic Molecular Simulation) from Davis et al. (2013) to turn EAGLE data into mock interferometric observations. `KinMS` has flexibility in outputting astronomical data cubes (with position, position, and frequency dimensions) and moment maps from various physical parameterisations and has been used for CO molecular gas modelling in previous work (e.g. Davis et al. 2013) and for observational predictions from EAGLE (Davis et al. 2019). Using

[3] `https://github.com/TimothyADavis/KinMSpy`

`KinMS` I generate simulated interferometric observations of galaxies directly from their 3D particle distributions.

Thanks to the controllable nature of the EAGLE data, I have the ability to generate millions of images from just a handful of simulations by using combinations of rotations and displacements of thousands of simulated galaxies per snapshot. This flexibility also has the added benefit of naturally introducing data augmentation for boosting the generalising power of an ML algorithm. For any given distance projection, galaxies were given 8 random integer rotations in position-angle ($0^\circ \leq \theta_{\rm pos} < 360^\circ$) and inclination ($5^\circ \leq \phi_{\rm inc} \leq 85^\circ$). Each galaxy is displaced such that they fill a $64'' \times 64''$ mock velocity map image in order to closely reflect the field of view (FOV) when observing CO(1-0) line emission with ALMA. I define the displacement of each simulated galaxy in terms of physical size and desired angular extent. Each galaxy's radius is given as the 98th percentile particle distance from its center of potential in kpc. I use this measurement, rather than the true maximum particle radius, to reduce the chance of selecting sparsely populated particles for calculating displacement distances, as they can artificially scale down galaxies.

The EAGLE galaxies were passed to `KinMS` to create cubes of stacked velocity maps, with fixed mock beam sizes of bmaj $= 3''$, ready for labelling. Each cube measured $64 \times 64 \times 8$ where 64×64 corresponds to the image dimensions (in pixels) and 8 corresponds to snapshots during position-angle and inclination rotations. The median physical scale covered by each pixel across all image cubes in a representative sample of my training set is 0.87 kpc. It should be noted that I set all non-numerical values or infinities to a constant value, as passing such values to an ML algorithm will break its training. I adopt 0 $\mathrm{km\,s^{-1}}$ as my constant (similarly to Diaz et al. 2019) to minimise the the background influencing feature extraction. My training set has a range in blank fraction (i.e. the fraction of pixels in images with blank values set to 0 $\mathrm{km\,s^{-1}}$) of 0.14 to 0.98, with a median blank fraction of 0.52. Figure 2.1 shows simulated ALMA observations of galaxies when using `KinMS` in conjunction with particle data from the EAGLE simulation `RefL0025N0376`.

2.2.4 SIMULATING NOISE

Often it is useful to observe the performance of ML models when adding noise to the input data, in order to test their robustness and their behavioural predictability. In one of my tests, I seeded the mock-EAGLE-interferometric-datacubes with

Gaussian distributed noise of mean $\mu = 0$ and standard deviation

$$\sigma = \frac{1}{\mathrm{S/N}} \left(\frac{1}{\mathrm{N}} \sum_{\mathrm{c=0}}^{\mathrm{c=N}} \mathrm{I_{max,c}} \right), \tag{2.2}$$

i.e. some fraction, $\frac{1}{\mathrm{S/N}}$, of the mean maximum intensity, $\mathrm{I_{max}}$, of each cube-channel, c, containing line emission. The resulting noisy data cubes are then masked using *smooth masking*, a method that is representative of how one would treat a real data cube (Dame 2011). An intensity weighted moment one map is then generated in `KinMS` from the masked cube as

$$\mathrm{M_1} = \frac{\int (\mathrm{v}) \mathrm{I_v} \mathrm{dv}}{\int \mathrm{I_v} \mathrm{dv}} = \frac{\sum (\mathrm{v}) \mathrm{I_v}}{\sum \mathrm{I_v}}, \tag{2.3}$$

where $\mathrm{I_v}$ is the observed intensity in a channel with known velocity v, before being normalised into the range of -1 to 1.

Noise presents a problem when normalising images into the preferred range. Rescaling, using velocities beyond the range of real values in a velocity map (i.e. scaling based on noise), will artificially scale down the true values and thus galaxies will appear to exhibit velocities characteristic of lower inclinations. I clip all noisy moment 1 maps at a fixed 96th percentile level, before normalising, in order to combat this effect. Note that this choice of clipping at the 96th percentile level is arbitrarily based on a handful of test cases and represents no specific parameter optimisation. Although simple, this likely reflects the conditions of a next generation survey in which clipping *on the fly* will be done using a predetermined method globally rather than optimising on a case by case basis.

2.2.5 Labelling the training set

Each galaxy, and therefore every cube, is assigned a label in the continuous range of 0 to 1 corresponding to the level of ordered rotation, κ, of that galaxy.

In Figure 2.1, the difference between levels of κ is clear in both structure and velocity characteristics, with low κ galaxies exhibiting less regular structures and more disturbed velocity fields than high κ galaxies.

Figure 2.2 shows the distribution of κ in my training set. It is clear that my training set is heavily imbalanced with a bias towards the presence of high κ galaxies. Additionally, as κ approaches one, the possible variation in velocity fields decreases as there are limited ways in which one can create orderly rotating disk-like structures. However, my dataset contains a surplus of galaxies as κ approaches one. Therefore,

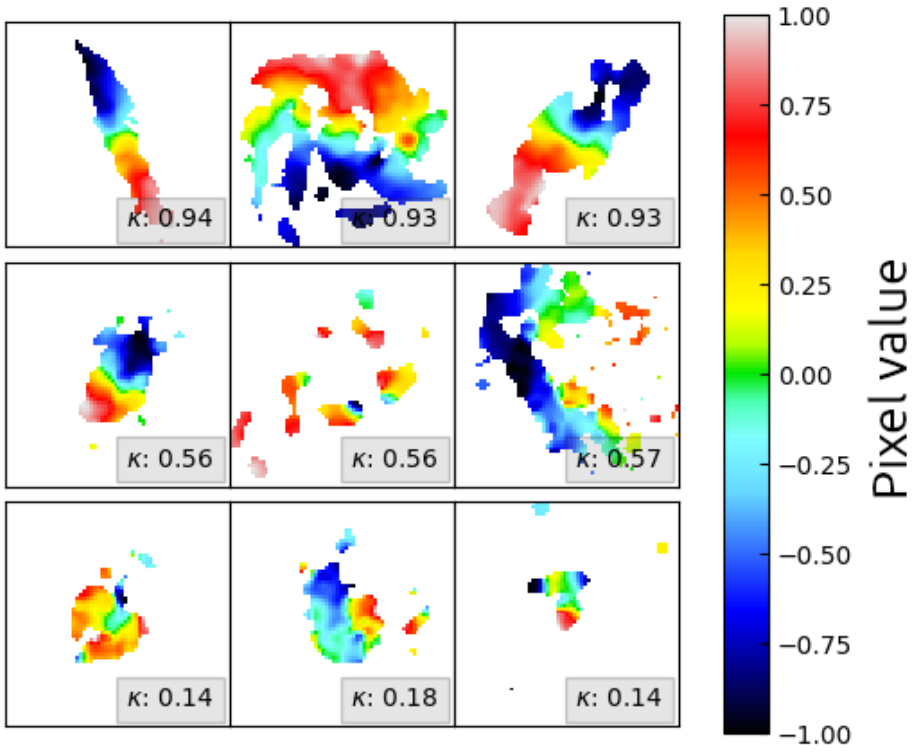


Figure 2.1. Random exemplar velocity maps for the noiseless EAGLE dataset. Rows of increasing order, starting from the bottom of the figure, show galaxies of increasing κ. The κ for each galaxy is shown in the bottom right of the frame in a grey box. Each galaxy has randomly selected position angle and inclination and the colourbar indicates the line of sight velocities, which have been normalised into the range -1 to 1 and subsequently denoted as *pixel values*. The images have dimensions of 64×64 pixels in keeping with the size of input images to my models in this chapter, as described in §2.2.6. One can easily see the changes in velocity field from $\kappa \sim 1$ to $\kappa \sim 0$ as galaxies appear less disk-like with more random velocities.

if one were to randomly sample from my dataset, for training an ML model, then the model would undoubtedly overfit to high κ images. This is a common problem in ML particularly with outlier detection models whose objectives are to highlight the existence of rare occurrences. In §2.2.6 I describe my solution for this problem with the use of weighted sampling throughout training to balance the number of galaxies with underrepresented κ values seen at each training epoch.

2.2.6 Model training: Rotationally invariant case

In this section I describe the creation and training of a convolutional autoencoder to embed κ into latent space and build a binary classifier to separate galaxies with κ above and below 0.5. Note that 0.5 is an arbitrarily chosen threshold for my

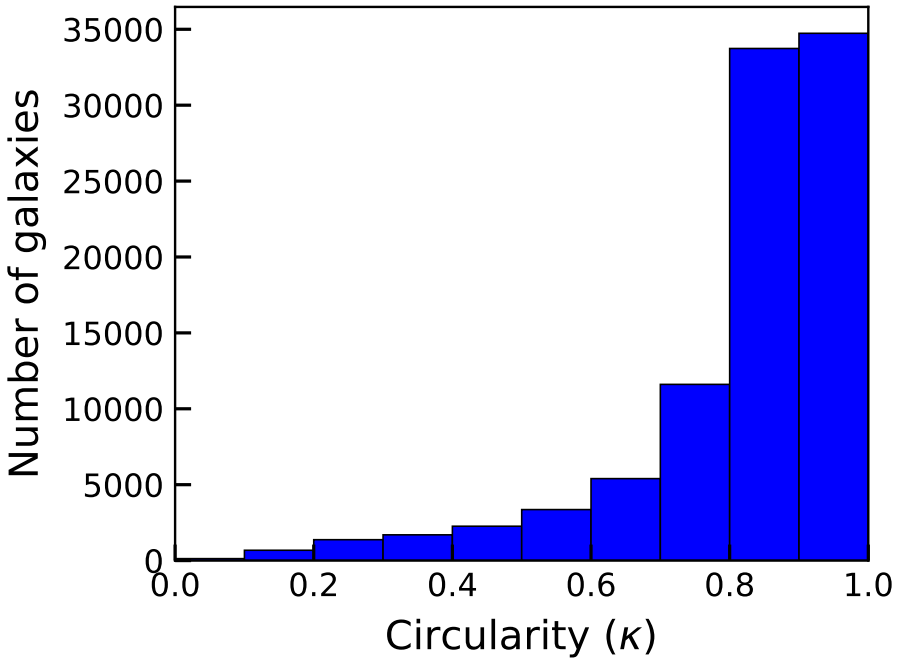


Figure 2.2. A histogram of κ labelled galaxies in the noiseless EAGLE training set. Galaxies have been binned in steps of $\delta\kappa = 0.1$ for visualisation purposes but remain continuous throughout training and testing. The distribution of κ is heavily imbalanced, showing that more galaxies exhibit a κ closer to 1 than 0.

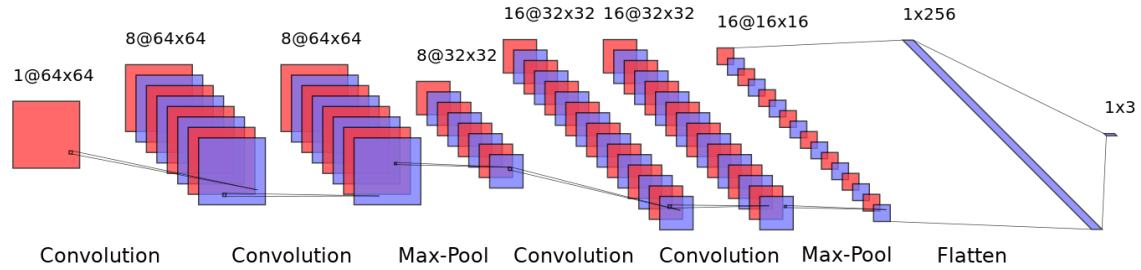
classification boundary but is motivated by the notion of separating ordered from disturbed gas structures in galaxies.

In order to construct my ML model, I make use of `PyTorch`[4] 0.4.1, an open source ML library capable of GPU accelerated tensor computation and automatic differentiation (Paszke et al. 2017). Being grounded in `Python`, `PyTorch` is designed to be linear and intuitive for researchers with a C99 API backend for competitive computation speeds. I use `PyTorch` due to its flexible and user friendly nature for native `Python` users.

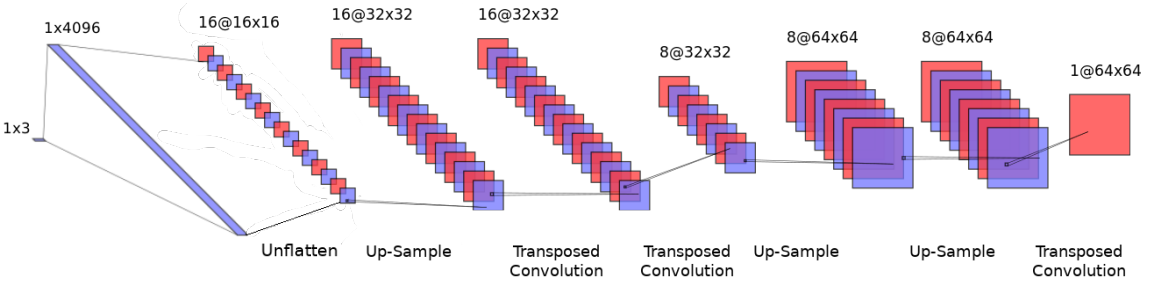
A visual illustration of the CAE architecture is shown in Figure 2.3 and described in Table A.1 in more detail.

The model follows no hard structural rules and is an adaption of standard CNN models. The decoder structure is simply a reflection of the encoder for simplicity. This means my CAE is unlikely to have the most optimised architecture and I propose this as a possible avenue for improving on the work presented in this chapter. The code developed for this chapter is available on GitHub[5] as well as an ongoing development

[4] `http://pytorch.org/`
[5] `https://github.com/SpaceMeerkat/CAE/releases/tag/v1.0.0`



(a) Encoder subnet



(b) Decoder subnet

Figure 2.3. Illustration of the CAE architecture used in this chapter. The encoder subnet (top) makes use of a series of convolutions and max-pooling operations to embed input image information into 3 latent dimensions. The decoder subnet (bottom) recovers the input image using transposed convolutions and up-sampling layers. The output of the encoder is passed to the decoder during training but throughout testing only the encoder is used map velocity maps into latent space.

version[6].

The CAE is trained for 300 epochs (with a batch size of 32) where one epoch comprises a throughput of 6400 images sampled from the training set. I do this to reduce the memory load throughout training given such a large training set. Images are selected for each mini-batch using a weighted sampler which aims to balance the number of images in each κ bin of width $\delta\kappa = 0.1$. Inputs are sampled with replacement allowing multiple sampling of objects to prevent under-filled bins. The model uses a mean squared error (MSE) loss,

$$\mathcal{L} = \frac{1}{\mathrm{N}} \sum_{\mathrm{i}=0}^{\mathrm{N}} \left(f\left(\mathrm{x_i}\right) - \mathrm{y_i} \right)^2 , \tag{2.4}$$

for evaluating the error between input and output images and weights are updated through gradient descent. N, $f(\mathrm{x})$, and y denote the batch size, model output for an input x, and target respectively. I use an adaptive Adam learning rate optimiser

[6] `https://github.com/SpaceMeerkat/CAE`

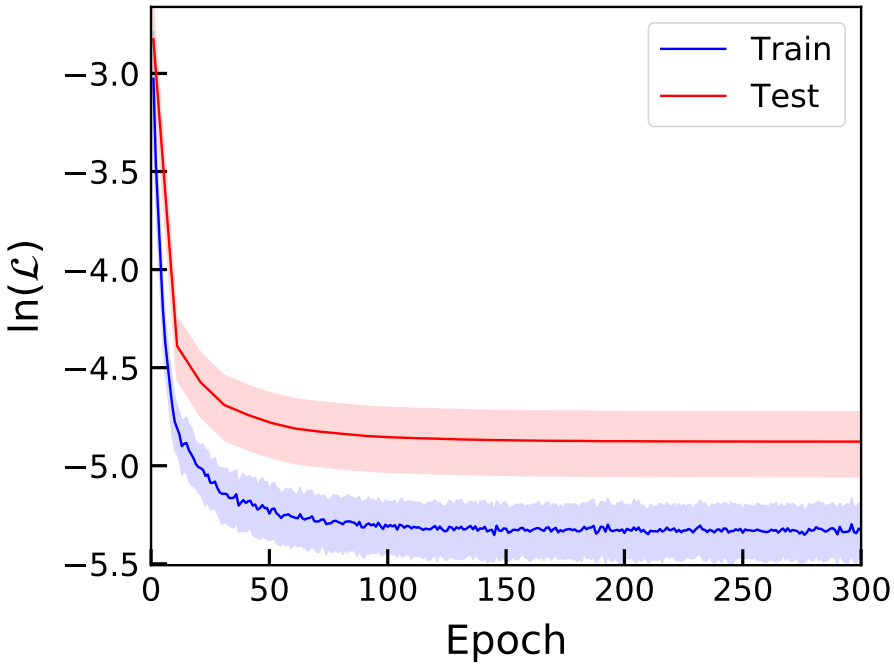


Figure 2.4. Training the CAE on noiseless EAGLE velocity maps. Solid lines show the natural log mean MSE loss and solid colour regions show 1σ spread at any given epoch. In order to reduce computational time, the test accuracy is evaluated every 10th epoch. I see smooth convergence of my CAE throughout training with no turn over of the test accuracy indicating that my model did not overfit to the training data.

(Kingma & Ba 2014), starting with a learning rate of 0.001 which halves every 30 epochs; this helps to reduce stagnation in the model accuracy from oversized weight updates. In Figure 2.4 I see that the model has converged well before the 300^{th} epoch and observe no turn-over of the test MSE loss, which would indicate overfitting.

The CAE learned to encode input images to 3 dimensional latent vectors (chosen for the ease of visualising resulting latent vectors, rather than physical motivation). Further testing showed that any higher compression, to lower dimensions, resulted in poor performance for the analyses described in §2.3 and compression to higher dimensions impaired my ability to directly observe correlations between features and latent positions with no improvement to the model's performance. I use `scikit-learn`'s[7] principal component analysis (PCA) function on these vectors to rotate the latent space so that it aligns with one dominant latent axis, in this case the z axis. As seen in Figure 2.5, the 3 dimensional latent space contains structural symmetries which are not needed when attempting to recover κ (but are still astrophysically useful; see §2.3.5). Because of this, the data is folded around the z and x

[7]https://scikit-learn.org

axes consecutively to leave a 2-dimensional latent space devoid of structural symmetries with dimensions |z| and $\sqrt{x^2+y^2}$ from from which I could build my classifier (see §2.3.3).

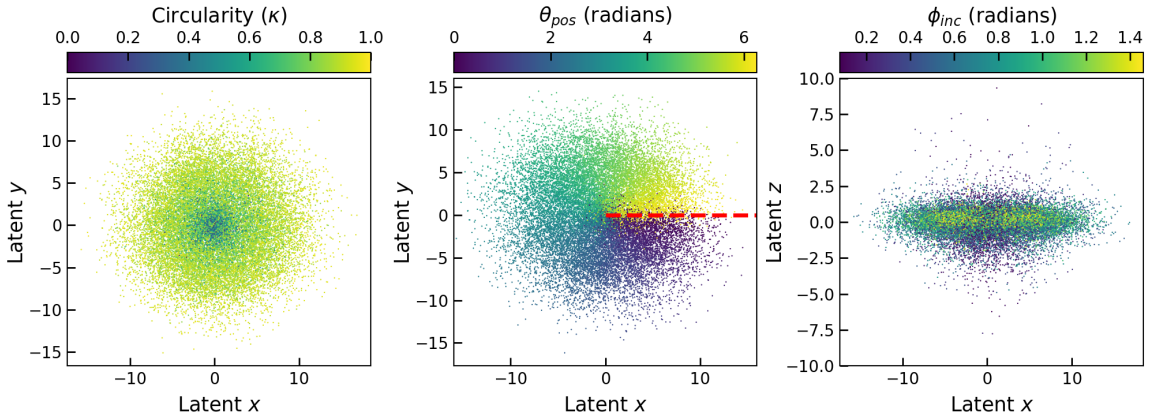


Figure 2.5. Noiseless eagle test data in 3D latent space. All subplots show the same latent structure but coloured differently by: true κ (left), true position angle ($\theta_{\rm pos}$, middle), and true inclination ($\phi_{\rm inc}$, right). It is clear from the left subplot that low κ galaxies lie close to the $z = 0$ region. $\theta_{\rm pos}$ is very neatly encoded in the clockwise angle around the latent z-axis. The red dashed line indicates the positive latent x axis from which θ_{pos} is measured. $\phi_{\rm inc}$ appears to be encoded in a much more complex fashion than κ and $\theta_{\rm pos}$.

Having tested multiple classifiers on the 2D latent space (such as high order polynomial and regional boundary approaches), I find that a simple vertical boundary line is best at separating the galaxies whose κ are greater than or less than 0.5. This is highlighted in Figure 2.6, where I see the spread on latent positions taken up by different κ galaxies makes a regression to recover κ too difficult. The exclusion of the z-axis in using a decision boundary is due to κ seemingly having no correlation with the latent encoding along this axis upon visual inspection and further testing. A discussion of the possible physical characteristics of galaxies encoded in the latent z-axis is explored later in this chapter.

In order to optimise the boundary line location, I measure the true positive (TP), true negative (TN), false positive (FP) and false negative (FN) scores when progressively increasing the boundary line's x location. The intersection of TP and TN lines (and therefore the FP and FN lines) in Figure 2.7 indicates the optimal position for my boundary, which is at $\sqrt{x^2+y^2} = 2.961 \pm 0.002$, when equalising the importance of correctly identifying both orderly and disorderly rotating classes. The smoothness of the lines in Figure 2.7 show how the two κ populations are well structured. If the two populations were clumpy and overlapping, one would observe

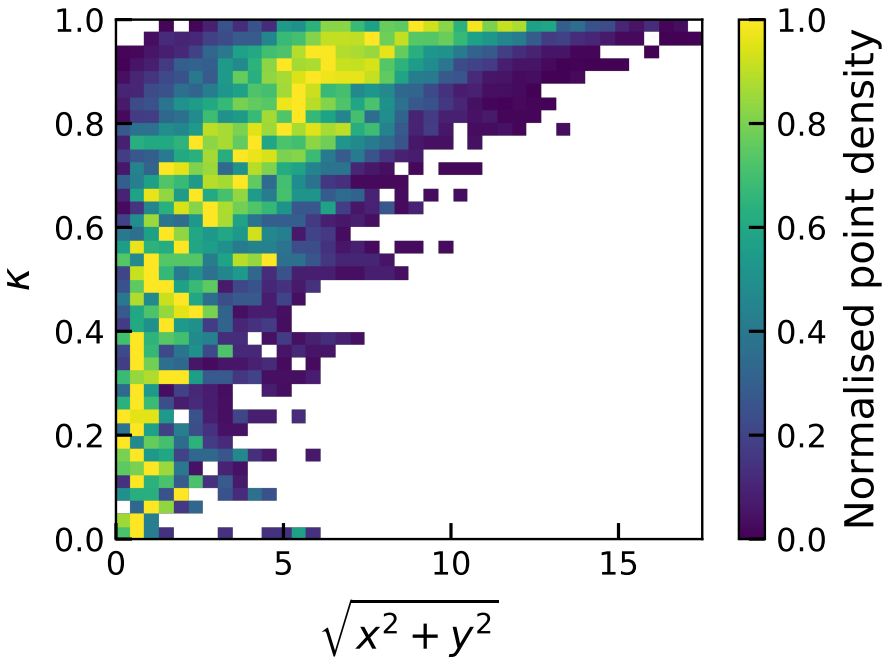


Figure 2.6. 2D histogram of κ against latent position for noiseless EAGLE test data. Pixels are coloured by point density normalised such that the point density in each row lies in the range 0 to 1. I see a very clear relationship between κ and latent position but also a high spread of latent positions occupied by high κ galaxies, making a regression task to recover κ from my encoding difficult.

unstable lines as the ratio of positive and negatively labelled galaxies constantly shifts in an unpredictable manner.

2.3 Results and discussion

2.3.1 Test case I: Noiseless EAGLE data

The number of high and low κ labelled images, in both the training and test sets, for the noiseless EAGLE dataset are shown in Table 2.1.

Table 2.1 Proportions of high and low κ labelled images in both training and test sets for the noiseless EAGLE dataset.

	Number of images		
Dataset	$\kappa > 0.5$	$\kappa < 0.5$	Total
Training	88840 (94%)	6144 (6%)	94984
Test	22224 (93%)	1560 (7%)	23784

Figure 2.8a shows the classification accuracy on the noiseless EAGLE training set. The TP and TN accuracy scores are unsurprisingly identical given the method

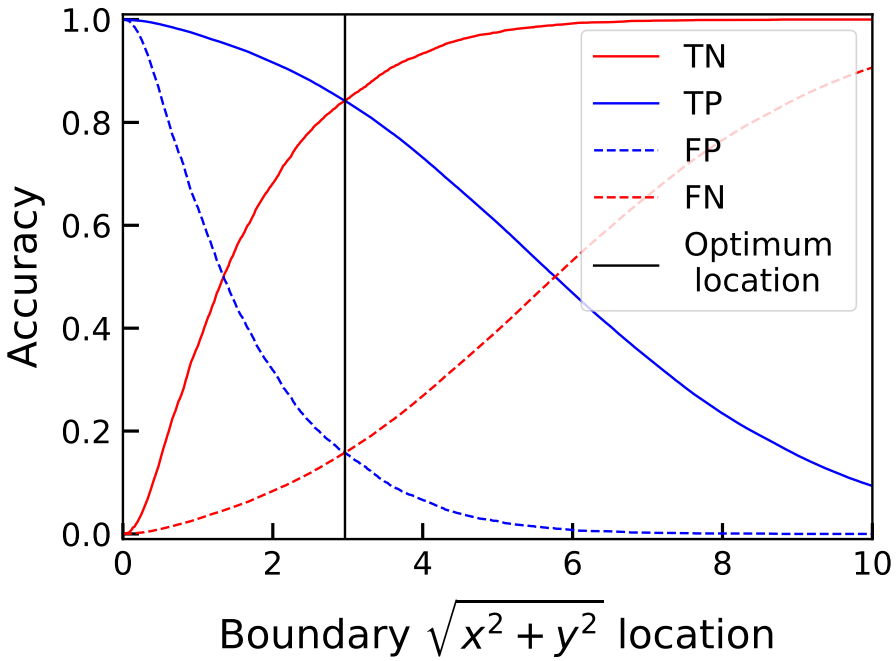


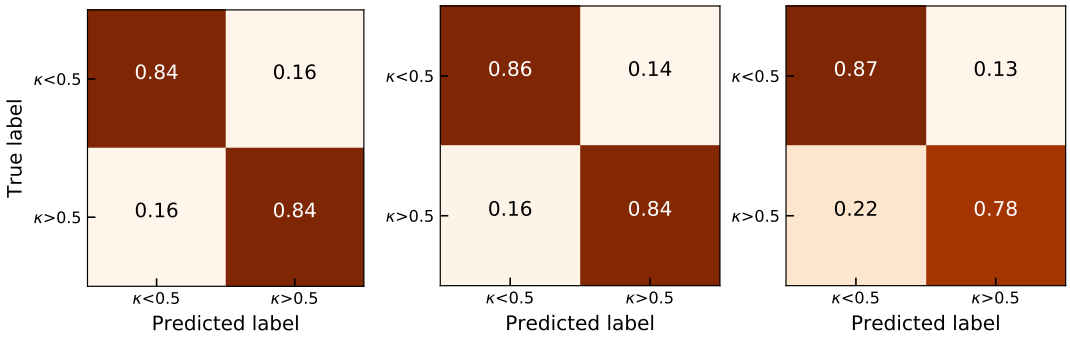
Figure 2.7. The observed change in all four components of a confusion matrix when changing the boundary line x-location. The optimal position for a binary classification is chosen as the intersection of TP and TN lines, which is identical to the location at the intersection of FP and FN lines. I observe smooth changes to the TN, TP, FP, and FN lines as the boundary line location changes, showing that both target populations are well clustered.

used to find the optimal boundary in §2.2.6 was designed to achieve this (see intersection points in Figure 2.7). The classifier has a mean training recall of 84% for both classes.

Figure 2.8b shows the confusion matrix when testing the noiseless EAGLE test set using my boundary classifier. I see that the model performs slightly better than when tested on the training set, suggesting that the model did not overfit to the training data and is still able to encode information on κ for unseen images.

2.3.2 TEST CASE II: NOISY EAGLE DATA

Figure 2.8c shows the results of classifying noisy EAGLE test data with S/N = 10 and masking at 3 times the RMS level (see §2.2.4 for details). Note that this is a simple test case and places no major significance on the particular level of S/N used. It is also a test meant purely for further understanding and verifying the expected behaviour of the model under different conditions, in a transfer learning style task. In reality, for maximising the predictive performance of the model in the presence



(a) Noiseless train **(b)** Noiseless test **(c)** Noisy test

Figure 2.8. Normalised confusion matrix showing the performance of the classifier when testing the 2.8a noiseless EAGLE training set, 2.8b noiseless EAGLE test set (seeded with Gaussian noise with $\frac{1}{\mathrm{S/N}} = \frac{1}{10}$ and masking at 3 times the RMS level of line free regions), and 2.8c noisy EAGLE test set. The mean recall scores are 84%, 85%, and 82.5% respectively.

of noise, one should ideally be training on the noisy data itself. The introduction of noise has a clear and logical, yet arguably minor, impact on the classifier's accuracy. The combination of adding noise followed by using an arbitrary clipping level causes test objects to gravitate towards the low κ region in latent space. This should come as no surprise as κ correlates with ordered motion; therefore, any left over noise from the clipping procedure, which itself appears as disorderly motions and structures in velocity maps, anti-correlates with κ causing a systematic shift towards the low κ region in latent space.

One could reduce this shifting to low κ regions in several ways. (1) Removing low S/N galaxies from the classification sample. (2) For my test cases I used a single absolute percentile level for smooth clipping noise; using levels optimised for cases on a one-by-one basis will prevent over-clipping. (3) If one were to directly sample the noise properties from a specific instrument, seeding the simulated training data with this noise before retraining an CAE would cause a systematic shift in the boundary line, mitigating a loss in accuracy. It should also be noted that I have not tested the lower limit of S/N for which it is appropriate to use my classifier but instead I focus on demonstrating the effects of applying noise clipping globally across my test set under the influence of modest noise.

2.3.3 Test case III: ALMA data

I tested 30 velocity maps of galaxies observed with ALMA to evaluate the performance of the classifier on real observations. Given that I used `KinMS` to tailor

the simulated velocity maps to closely resemble observations with ALMA I expect similar behaviour as seen when testing the simulated data. For my test sample I use an aggregated selection of 15 velocity maps from the mm-**W**ave **I**nterferometric **S**urvey of **D**ark **O**bject **M**asses (WISDOM) CO(1-0, 2-1, and 3-2) and 15 CO(1-0) velocity maps from the **A**LMA **Fo**rnax **C**luster **S**urvey (AlFoCS, Zabel et al. 2019). I classify each galaxy, by eye, as either disturbed or regularly rotating (see Table A.2) in order to heuristically evaluate the classifier's performance.

Figure 2.9 shows the positions of all ALMA galaxies (round markers) in my folded latent space, once passed through the CAE. Of the 30 galaxies, 6 (20%) are classified as $\kappa < 0.5$; this higher fraction, when compared to the fraction of low κ galaxies in the simulated test set, is likely due to the high number of dwarf galaxies, with irregular H_2 gas, targeted in AlFoCS.

I find one false positive classification (as determined by eye, yet distinctly belonging to the negative class) close to the classification boundary and one false positive classification far from the classification boundary. The false negative classification of NGC1351A can be explained by its disconnected structure and edge-on orientation (see Zabel et al. 2019; Figure B1). Since low κ objects appear disconnected and widely distributed among their velocity map fields of view, it is understandable why NGC1351A has been misclassified as a disturbed object. It should be noted that upon inspection the false positive classification of FCC282 can be attributed to the appearance of marginal rotation in the galaxy. I see evidence of patchy high κ galaxies residing closer to the classification boundary than non-patchy examples. This may indicate a relationship between patchiness and latent positioning. The classifier performs with an accuracy of 90% when compared to the predictions by human eye. Of the 30 galaxies observed with ALMA, 6 (20%) are classified as low κ and of the 23 (77%) galaxies identified by eye as likely to be high κ galaxies, only one was misclassified as low κ.

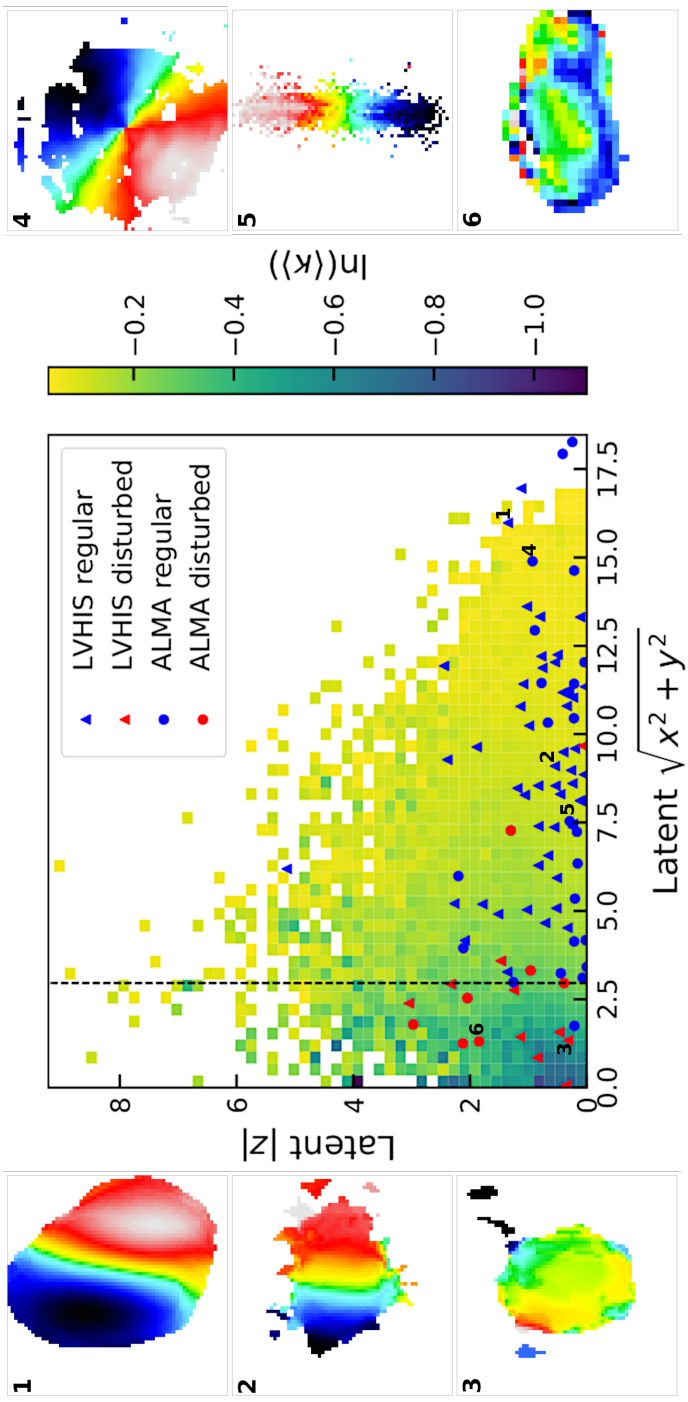


Figure 2.9. Folded latent space positions of noiseless EAGLE galaxies (coloured 2D histogram), ALMA galaxies (circular markers), and LVHIS galaxies (triangular markers); the suspected low κ galaxies are coloured red and suspected high κ galaxies coloured blue. The EAGLE points are coloured by average value of κ in each 2D bin and the grey dashed line shows the classifier boundary between $\kappa < 0.5$ and $\kappa > 0.5$ objects. 8 of the 10 suspected low κ LVHIS galaxies are classified as $\kappa < 0.5$ and 5 of the 7 suspected low κ ALMA galaxies are classified as $\kappa < 0.5$. In both cases I see only 1 misclassification far from the boundary line and low κ region. A selection of 3 LVHIS and 3 ALMA galaxies are shown to the left and right of the central plot respectively, their locations in latent space indicated by black numbers. Images are scaled to 64×64 pixels with values normalised between -1 to 1, because of this images are shown as a close representation of CAE inputs before latent encoding. For illustration purposes the backgrounds have been set to `nan` whereas the CAE would instead see these regions as having a value of 0.

2.3.4 TEST CASE IV: LVHIS DATA

In order to test the robust nature of the classifier, I used it to classify velocity maps of HI velocity fields from the the **L**ocal **V**olume **HI S**urvey (LVHIS; Koribalski et al. 2018). This is an important test as it determines the applicability of the classifier to HI line emission observations, the same emission that the SKA will observe. As described in §2.2.3, the EAGLE training set was designed to reflect observations with ALMA, making this transfer learning test a good opportunity to evaluate the model's ability to generalise to unseen data containing different systematic characteristics.

Rather than moment masking the data cubes, like in §2.3.2, each cube is clipped at some fraction of the RMS (calculated in line free channels) to mimic the noise removal processes used in generating velocity maps in the LVHIS database. All galaxies whose positions could not be found using the `Python` package `astroquery`[8] (searching the SIMBAD Astronomical Database[9]), or whose HI structures were clearly misaligned with the true galaxy centres, were omitted from further testing. This was to prevent misclassification based on pointing error which correlates with features of disorderly rotation to the CAE and would artificially increase the FN rate. This left 61 galaxies (see Table A.3) from which velocity maps were made and passed through the CAE. Finally, where images were not 64×64 pixels, I used PyTorch's `torch.nn.functional.interpolation` function (in bilinear mode) to rescale them up or down to the required dimensions prior to clipping.

The latent positions of all HI galaxies are shown in Figure 2.9 (triangular markers). Of the 61 galaxies, 8 (13%) are classified as low κ. By eye, I identified 10 galaxies in the LVHIS which are likely to be definitively classified as $\kappa < 0.5$ (see Table A.3). Of these 10 candidates, 8 were correctly identified as $\kappa < 0.5$, 1 is observed as very close to the classification boundary and 1 is unquestionably misclassified.

2.3.5 RECOVERING POSITION ANGLE

Scientists who wish to model the kinematics of galaxies often require initial estimates for parameters such as position angle, inclination, mass components, radial velocity profiles etc. Given that position angle is clearly encoded in my latent xy plane (see Figure 2.5), it is possible to return predicted position angles with associated errors. This could prove useful for fast initial estimates of $\theta_{\rm pos}$ for scientists requiring them for kinematic modelling. I define the predicted position angle, $\theta_{\rm latent}$, as the clockwise angle between the positive latent x-axis and the position of data points

[8] `https://astroquery.readthedocs.io/en/latest/`
[9] http://simbad.u-strasbg.fr/simbad/

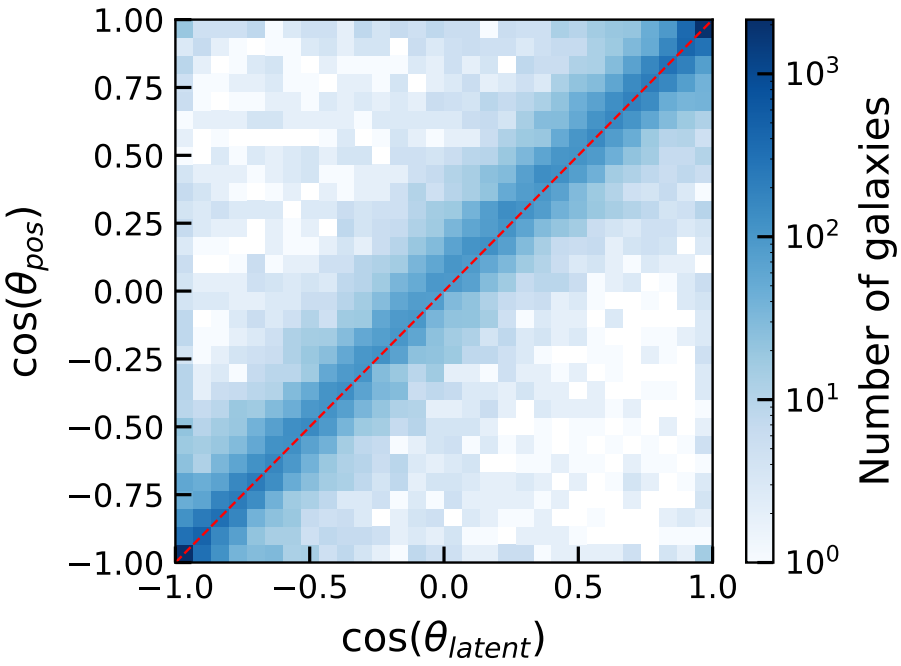


Figure 2.10. 2D histogram showing the predicted position angles for the noiseless EAGLE test set against their true position angles. The red dashed line shows the 1:1 line along which all data would lie for a perfect predictor of position angle.

in the latent xy plane. I removed the systematic angular offset, $\delta\theta$, between the positive latent x-axis and the true position angle $(\theta_{\rm pos}) = 0°$ line by rotating the latent positions by the median offset, found to be $\delta\theta \sim 36.6°$, and subtracting an additional 180°. In the now rotated frame, $\theta_{\rm latent}$ is defined as $\tan^{-1}\left(\frac{y}{x}\right)$, where x and y are the latent x and y positions of each galaxy (see Figure 2.10). I calculated errors on the resulting predictions of $\theta_{\rm latent}$ by taking the standard deviation of residuals between $\theta_{\rm latent}$ and $\theta_{\rm pos}$.

I repeated this procedure for the noisy EAGLE data, with S/N = 10, the results of which are also shown in Figure 2.11 with red error bars. I can see that the recovery of $\theta_{\rm pos}$ is still well constrained at higher inclinations with only a slight increase in the error most notably at lower inclinations (see Figure 2.11. I see that at higher inclinations the error in predicted $\theta_{\rm pos}$ is better constrained than for lower inclinations. This should come as no surprise as the ellipticity of galaxies and the characteristic shape of their isovels are gradually lost as a galaxy approaches lower inclinations thus making it more difficult to calculate $\theta_{\rm pos}$. During further testing I also observe reduced errors on position angles when limiting to higher κ test galaxies.

It should be noted that my method for recovering $\theta_{\rm pos}$ is not the only one. Other kinematic fitting routines exist for this purpose including `fit_kinematic_pa`

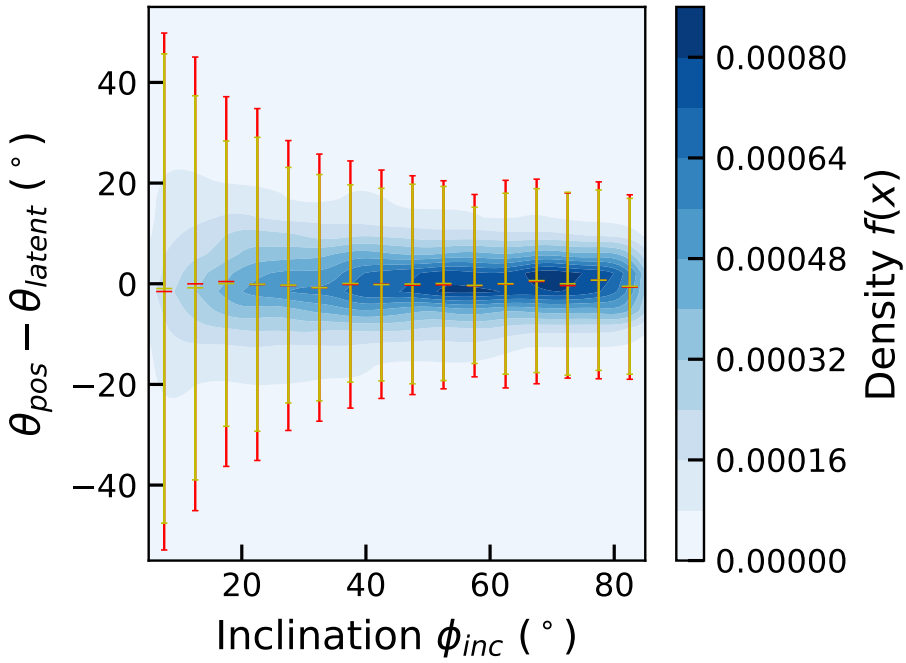


Figure 2.11. Kernel density estimation of error in $\theta_{\rm pos}$ against inclination for noiseless EAGLE test data (yellow error bars) and noisy EAGLE test data (red error bars). Coloured contours show the 2D probability density, central horizontal line markers show the mean error in $\theta_{\rm pos}$ in bins of width $\delta\theta_{\rm pos} = 5°$. The error bars show the standard error in each bin.

(Krajnović et al. 2006) and the radon transform method (Stark et al. 2018). These methods likely have higher accuracy than seen here, as my network was not optimised for the recovery of θ_{pos}. Bench marking an ML model against existing ones, as a dedicated standalone mechanism for recovering $\theta_{\rm pos}$, is an avenue for future research.

Given that there is such a strong overlap in z-positions occupied by different galaxy inclinations, I were unable to recover the inclinations of galaxies in the simple manner as for $\theta_{\rm pos}$. However, from visualising the distribution of inclinations against latent-z position, I am confident that inclination plays a part in latent positioning of galaxies. Because of this I am confident in my understanding of all 3 latent dimensions that the CAE has learned.

2.4 CONCLUSIONS

I have shown that it is possible to use ML to encode high dimensional (64×64 pixels) velocity maps into 3 dimensions, while retaining information on galaxy kinematics, using convolutional autoencoders. I have successfully recovered the level of ordered rotation of galaxies using a simple binary classifier, from a multitude of

test sets including simulated EAGLE velocity maps, ALMA velocity maps, and HI survey velocity maps. When testing real observational data, I see a clustering of low κ galaxies towards the origin and around the classification boundary, in line with my understanding of my folded two dimensional latent space. My tests on simulated data show a mean recall of 85% when attempting to recover the circularity parameter as well as 90% and 97% heuristic accuracy when recovering the circularity parameter for galaxies observed with ALMA and as part of LVHIS respectively. I have managed to mitigate the problems associated with a heavily imbalanced training set by using both weighted sampling during training and balancing the true positive and true negative accuracy scores when constructing my classifier. In addition to recovering information on the ordered rotation characteristics of galaxies, I have also been successful in providing estimates on position angle from the full 3D latent positions of velocity maps with associated errors. These will be useful for initial guesses at $\theta_{\rm pos}$ for kinematic modelling routines in related work.

I were able to show my classifier's positive performance when testing LVHIS data. This outcome is important for two reasons: (1) it shows the robustness of the classifier when making the transition from simulated to real data of different origins and (2) it shows that using machine learning to study the kinematics of HI sources is likely possible and therefore applicable to SKA science.

Recovering inclinations, $\phi_{\rm inc}$, of galaxies was not possible using my CAE due to the high overlap in latent z positions for the entire range of $\phi_{\rm inc}$. However, the spread of z positions occupied by galaxies at mid-range inclinations was considerably less than at lower inclinations, indicating that while $\phi_{\rm inc}$ is not recoverable, I am confident that it is partly responsible for the positions of galaxies in the latent z-axis. Therefore, I have a rational understanding of what information all three latent dimensions are encoding from the input images. This makes my model predictable and logical in how it behaves when seeing input data. This understanding is often missing in CNN style networks, and especially in deep learning models.

The main caveat with this work pertains to the use of percentages in my maximum likelihood function when calculating the optimal boundary line for the binary classifier. This makes my classifier independent of the underlying distribution of high and low κ galaxies in an attempt to maximise the recall of both classes. The means my classifier will work well in situations where both classes are more equally distributed (such as galaxy clusters). However, one should take care when testing heavily imbalanced datasets where, although the dataset has been drastically thinned of high κ galaxies, it is likely that the user will still need to examine the low κ classification set for contaminants.

As demonstrated by Diaz et al. (2019), using a combination of morphology and kinematics for classification purposes improves performance over using only one attribute. Therefore, a logical improvement on my work would be using a branched network or an ensemble of networks which use both moment zero and moment one maps to make predictions on kinematic properties. My models rely on using maps of galaxies which are centred on their centres of potential (i.e. the position of the most bound particle); therefore, my classifier is sensitive to the choice of centre of potential proxy. This is undeniably an issue for on-the-fly surveys where the centre of potential of a target is estimated rather than empirically calculable. Therefore, including information such as intensity maps may allow re-centring based on observed characteristics rather than archived pointings for improving the classifiers performance. I see this as the most lucrative avenue for improving my models in the future.

Performing operations on a velocity map, as I have done in this work, means I am working several levels of abstraction away from the raw datacubes that future instruments, such as the SKA, will create. Therefore improvements could be made on my methods to analyse the effects of encoding datacubes into latent space rather than velocity maps. CNNs have long been capable of performing operations on multi-channel images, making this avenue of research possible and useful in reducing the need for heavy processing of raw datacubes before processing with ML algorithms as I have done in the work.

Chapter 3
A self-supervised, physics-aware, Bayesian neural network architecture for modelling galaxy emission-line kinematics

> "You only have one life.
> You have to spend it doing something that matters."
>
> Mike Massimino, *"Spaceman: An Astronaut's Unlikely Journey to Unlock the Secrets of the Universe"*

In this chapter, I present my work in building upon the knowledge gained in Chapter 2. With this in mind, the techniques discussed in this Chapter are more suited to next generation survey styles as will be explained. In the upcoming decades large facilities, such as the SKA, will provide resolved observations of the kinematics of millions of galaxies. In order to assist in the timely exploitation of these vast datasets I explore the use of a self-supervised, physics aware neural network capable of Bayesian kinematic modelling of galaxies. I demonstrate the network's ability to model the kinematics of cold gas in galaxies with an emphasis on recovering physical parameters and accompanying modelling errors. The model is able to recover rotation curves, inclinations and disc scale lengths for both CO and HI data which match well with those found in the literature. The model is also able to provide modelling errors over learned

parameters thanks to the application of quasi-Bayesian Monte-Carlo dropout. This work shows the promising use of machine learning, and in particular self-supervised neural networks, in the context of kinematically modelling galaxies. This work represents the first steps in applying such models for kinematic fitting and I propose that variants of our model would seem especially suitable for enabling emission-line science from upcoming surveys with e.g. the SKA, allowing fast exploitation of these large datasets.

3.1 INTRODUCTION

In studying galaxy evolution, astronomers often use the atomic Hydrogen (HI) 21-cm line to trace the outermost regions of galactic discs (e.g. Warren et al. 2004; Begum et al. 2005; Sancisi et al. 2008; Heald et al. 2011; Koribalski et al. 2018). This region can mark the continuous boundary between galaxies and their surrounding environments, including the dark matter halos within which galaxies are thought to reside. The rotation curves of extended HI discs can be used to begin probing the properties of dark matter halos as well as allow the detailed modelling of galaxies' mass distributions when coupled with ancillary observations (e.g. van Albada et al. 1985; de Blok et al. 2008). In the local Universe, HI discs are useful in determining the gaseous content of a galaxy as well as allowing astronomers to probe kinematic properties ranging from substructures such as bars, warps, counter-rotating discs, and spiral arms (e.g. Józsa et al. 2007; Spekkens & Sellwood 2007; Kamphuis et al. 2015; Di Teodoro & Fraternali 2015). Molecular gas observations (typically of the CO molecule) can provide a complimentary view of these regions at high resolution, revealing the interplay between these gas phases. HI is typically more extended than molecular gas, however, allowing it to trace environmental properties such as extended tidal features and the existence of dwarf companions (Hibbard et al. 2001; Sancisi et al. 2008; Heald et al. 2011; Serra et al. 2013; Bosma 2016; Koribalski et al. 2018).

The evolution of HI gives astronomers insight into the method by which galaxies accrete material from surrounding environments and how the mass of galaxies builds and evolves through star formation. The next generation of HI survey instruments (e.g. the Square Kilometre Array, Dewdney et al. 2009, Australian Square Kilometre Array Pathfinder, Johnston et al. 2007, 2008, the South African Meer-Karoo Array Telescope, Jonas & MeerKAT Team 2016, the Chinese Five-hundred metre Aperture Spherical Telescope, Li & Pan 2016) are poised to collect observations spanning a large look-back time, advancing our HI driven science as well as

pushing this field of astronomy firmly into the *Big Data* era.

Currently it is estimated that the **S**quare **K**ilometre **A**rray (SKA) will collect data on the order of hundreds of petabytes per year. Given that amount of data is not only too much to fully exploit by hand but also too large to store, astronomers should be looking to develop real-time models that can perform efficient science on incoming data. In an ideal world, physical information would be extracted from incoming data automatically, leaving the work of unravelling the prevailing science to astronomers. However, with such large data volumes and time-intensive techniques how are astronomers to begin moving in a direction in which we can fully exploit the data quality promised by the SKA?

In previous work I sought to begin addressing this challenge via the application of machine learning (see §2 and Dawson et al. 2019), and in particular neural networks, to extract kinematic properties of cold gas in galaxies. Models and tools exist to do this kind of work already. With the upcoming data releases from surveys such as the Widefield ASKAP L-Band Legacy All-Sky Blind Survey (WALLABY), it comes as no surprise that kinematic modelling tools (e.g. `3D-BAROLO`[1] Di Teodoro & Fraternali 2015, `2DBAT`[2] Oh et al. 2017), `FAT`[3] Kamphuis et al. 2015, and `KinMS`[4] Davis et al. 2013; Davis et al. 2020) have been in use and ongoing development for some time. Yet these models typically require several minutes or more to provide a full kinematic model of a single object, and longer if errors are required, which may prove problematic for kinematic analyses at SKA survey speeds (terrabytes per second, where data storage is not feasible).

In the past decade machine learning (ML) has become a popular solution to many *Big Data* challenges in galaxy evolution studies (e.g. Dieleman et al. 2015; Domínguez Sánchez et al. 2018a,b; Ackermann et al. 2018, Bekki 2019), but remains an under-utilised resource among the galaxy kinematics community. Computer vision, which often utilises ML techniques, has been successfully applied to kinematic characterisation (e.g. Stark et al. 2018). Yet, there is a distinct absence of directly exploiting ML (with the notable exception of a few recent works, e.g. Shen & Bekki 2020). Recently our group has made attempts to exploit the use of ML in this field, featuring the use of convolutional autoencoders to identify disturbed cold gas in galaxies using data from both simulations and observations (see Dawson et al. 2019). We still have a long way to go in fully exploring the application of ML to galaxy kinematic characterisation but it appears to be a promising avenue of research and one which I

[1] `https://editeodoro.github.io/BBarolo/`
[2] `https://github.com/seheonoh/2dbat`
[3] `https://github.com/PeterKamphuis/FAT`
[4] `https://github.com/TimothyADavis/KinMSpy`

explore further in this work.

While conventional ML models are capable of high empirical accuracy and low testing time (e.g. Breiman 2001; Krizhevsky et al. 2012), they are often highlighted for their slow training times (Lim et al. 2000) in comparison to testing times and data inflow rates and, in some cases, reluctance to generalise to unseen datasets (Dinh et al., 2017; Kawaguchi et al., 2017). These qualities are unsuitable for survey tasks proposed for the SKA and therefore we are required to look at alternative methods that incorporate the benefits of ML, without the drawbacks associated with standard ML practice.

Such an approach may exist in the form of self-supervised learning (Liu et al., 2020), whereby models train themselves without the need for an isolated training set. This has huge benefits in that one does not require long training times on a throw-away-dataset, essentially eliminating data wastage. As with all machine learning approaches, self-supervised learning does have its disadvantages including requiring fixed analytical functions to perform training, as well as results which change depending on when one wishes to evaluate test data throughout the model training procedure. Few pilot tests of these networks exist in astronomy (and even fewer utilising physics-aware capabilities, e.g. Aragon-Calvo & Carvajal 2020) and none exist in the modelling of galaxy kinematics. In this chapter I present the current results from our first attempts at creating a self-supervised neural network with the primary goal of inferring the kinematic properties of gas discs in galaxies and an emphasis on extracting (simplistic) characteristics of their rotation curves.

The chapter is divided into 3 main sections. §3.2 gives an in depth description of the model architecture used throughout this work, with emphasis lying on the decoder subnet described in §3.2.4. §3.3 presents the results from testing the network using synthetic and real interferometric observations, and §3.4 summarises the main outcomes of the work presented in this chapter as well as proposed avenues for future work.

3.2 THE MODEL

3.2.1 INPUT DATA

A typical interferometric observation returns visibilities in a complex plane from which one can obtain a 3D datacube consisting of 2D spatial flux observations separated into discrete channels which correspond to observed frequency. It is this

channelisation that allows astronomers to measure the line of sight velocities and hence the kinematic properties of galaxies' gas reservoirs. In practice, one can collapse these datacubes further to create 2D maps that reflect the mean properties of the gas in galaxies. A moment zero (integrated intensity) map is simply a summation along a cube's frequency/velocity dimension:

$$\text{Moment zero} = \int \mathrm{I_v dv} = \Delta \mathrm{v} \sum \mathrm{I_v}, \tag{3.1}$$

and a moment one (velocity) map is an intensity weighted averaging of the line of sight velocities

$$\text{Moment one} = \frac{\int \mathrm{v I_v dv}}{\int \mathrm{I_v dv}} = \frac{\sum \mathrm{v I_v}}{\sum \mathrm{I_v}}. \tag{3.2}$$

Working directly with the datacubes, or in fact the complex visibilities, would be optimal for any fast pipeline kinematic modelling tool. However, I have chosen to work with moment maps in this work as a first step and to avoid the problems associated with channelised inputs as explained further in §3.4. It should be noted that, because of our choice to use moment maps, the models described in this work are also suitable to analyse optical IFU maps, as they will be handled similarly by the model described in this work and have been shown to encode kinematic information which can be extracted using both analytical and ML approaches (e.g. Stark et al. 2018; Hansen et al. 2020). This will be explored further in future work (Dawson et al., in prep).

It should be noted that in this work, I am not making attempts to mitigate the effects of "beam smearing" (Swaters, 1999; Blais-Ouellette et al., 1999). During the recovery of datacubes from complex visibilities, the raw observational datacubes are convolved with a *restoring beam* which effectively encodes the complex visibility plane coverage and is, in some ways, analogous to resolution. It is this convolution step which gives rise to "beam smearing", the effects of which are discussed further in §3.3.1 along with implications for interpreting the model results discussed in this work. Counteracting "beam smearing" will need to be tackled in future work to maximise the effectiveness of models of this type.

3.2.2 Model aim

An autoencoder (Rumelhart et al., 1986) is a model composed of two subnets, an *encoder* and a *decoder*. In an *undercomplete autoencoder* the encoder subnet extracts features and reduces input images to a constrained number of nodes. This

so-called *bottleneck* forces the network to embed useful information about the input images into a nonlinear manifold from which the decoder subnet reconstructs the input images and is scored against the input image using a loss function.

The aim of the model used in this work is to extract semantically meaningful information from observational data. Typical approaches using a convolutional autoencoder (CAE, Masci et al. 2011) (such as that presented in §2) are powerful for extracting arbitrary (hyperparametric) features that define dataset characteristics. During training, a CAE learns to minimise the difference between input and output tensors rather than the difference between an output and target label (whether this be a continuous or categorical set of target classes). A CAE works similarly to a powerful nonlinear generalisation of principle component analysis (PCA, Plaut 2018) whereby it finds a continuous nonlinear latent surface on which input data best lies. In this work, however, I would like to extract semantically meaningful parameters of observed systems. In order to achieve this I have combined a convolutional autoencoder with a set of analytical, gradient trackable functions (i.e. the differential of the operation with respect to network weights and biases is a tractable process) which approximate the functional forms of observed kinematics of galaxies.

The model, known as a *semantic autoencoder* (SAE, Kodirov et al. 2017), is a modified CAE created using `PyTorch`[5] 0.4.1, an open source ML library capable of GPU accelerated tensor computation and automatic differentiation (Paszke et al., 2017). The model has a neural network architecture suited to self-supervised learning, with additional Bayesian capabilities. Figure 3.1 shows a simplified pictorial representation of the model architecture.

The encoder subnets extract lower dimensional feature representations from input images (here the integrated intensity and mean velocity maps as described in §3.2.1) using a combination of convolutional and linearly connected layers; the decoder then reconstructs the input images from the learned feature representations. In a standard convolutional autoencoder, the decoder would make use of transposed convolution operations, however in this network the decoder is composed of analytical functions written using native `PyTorch`. This imposes a constraint on the CAE by forcing the network to generate a semantic encoding of the input images. As highlighted by Aragon-Calvo & Carvajal (2020), the decoder function can take any possible form, no matter how representative of the true underlying functions being modelled. In this way, we can be assured that the encoders are learning semantically meaningful properties of the input images and are no longer tied to traditional

[5] `http://pytorch.org/`

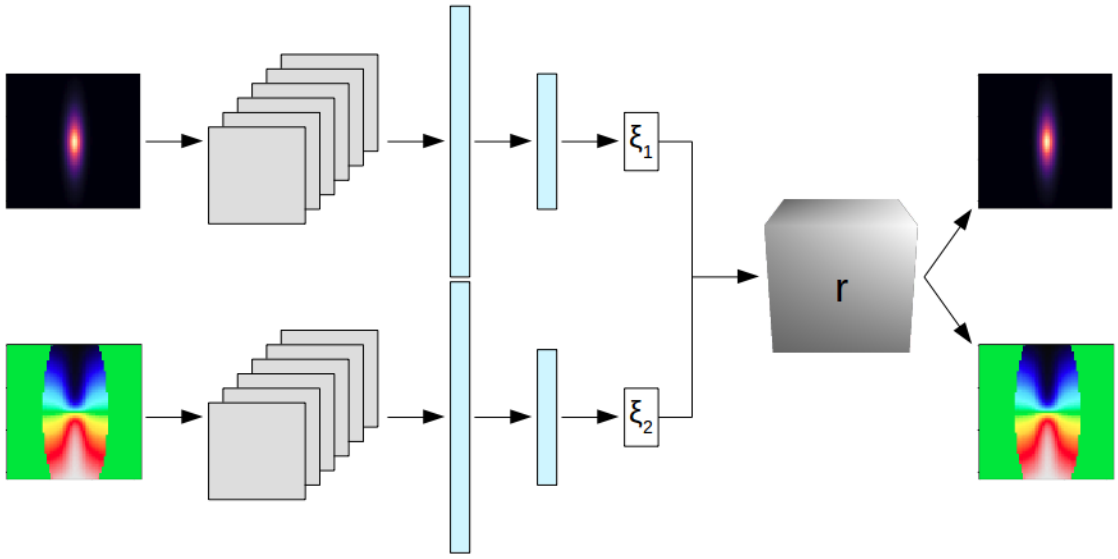


Figure 3.1. A simplified pictorial representation of the neural network used throughout this work. The model features two convolutional encoder subnets which concatenate learned features before passing them to a decoder subnet. The model receives moment maps as inputs and minimises the loss between decoder-generated moment map outputs and the inputs throughout training. In the diagram grey squares indicate convolutional layers, blue rectangles depict linearly connected layers, and the grey cube represents the auxiliary 3D cube containing the coordinate axes passed into the network.

training methods, instead allowing the network to train on all available data (including test data) in a self-supervised manner. An SAE becomes physics-aware once the assumption is made that the decoder function can be used to reveal physically meaningful information about the input. In this chapter, the physics-awareness of the model refers to our main focus of approximating parameterisations for rotation curves, intensity profiles and recovering galaxy inclinations (see §3.2.4).

For a more in-depth background to the use of autoencoders I refer the reader to Bourlard & Kamp (1988) and Hinton & Salakhutdinov (2006). For both a concise and thorough introduction to the use of self-supervised, physics aware, neural networks in astronomy I recommend Aragon-Calvo & Carvajal (2020).

3.2.3 The encoder subnets

Within the network, the encoders are two convolutional-classifier-like subnets. Each comprises a series of 4 convolutional and 2 fully connected layers, interspersed with pooling layers and activation functions. The encoders are used to extract and dimensionally reduce features from input images. The two subnets independently

Table 3.1 The SAE encoder subnet architecture used throughout this chapter. The first column lists the name of each layer/operation, the second column describes the type of layer/operation, the third column shows the dimensions of each layer's output tensors (hence the input shape to the next layer). The dimensions follow the PyTorch convention (batch size, number of channels, height, width). The filter column shows the dimensions (height, width) of kernels used to perform the convolution and pooling operations. The convolutional and linearly connected layer groups are separated by a blank row for clarity.

Name	Layer/Operation	Dimensions	Filter
Input	–	(64,1,64,64)	–
Conv	2D Convolution	(64,16,64,64)	(3,3)
Pool	2D Max Pooling	(64,16,32,32)	(2,2)
Conv	2D Convolution	(64,32,32,32)	(3,3)
ReLU	ReLU	–	–
Pool	2D Max Pooling	(64,32,16,16)	(2,2)
Conv	2D Convolution	(64,64,16,16)	(3,3)
ReLU	ReLU	–	–
Pool	2D Max Pooling	(64,64,8,8)	(2,2)
Conv	2D Convolution	(64,128,8,8)	(3,3)
ReLU	ReLU	–	–
Pool	2D Max Pooling	(64,128,4,4)	(2,2)
Lc1	Linear	(64,1,1,2048)	–
ReLU	ReLU	–	–
Drop	Dropout (p=0.1)	–	–
Lc2	Linear	(64,1,1,256)	–
Htanh	Hard tanh activation	–	–
Output	–	(64,1,1,2)	–

receive a moment zero map (a 2D intensity profile, normalized in the range 0–1) and a moment one map (a 2D velocity profile, normalized into the range -1–1) respectively. Throughout this work, I ensure that the input maps have size of 64×64 pixels. All input maps whose sizes are larger or smaller, like those discussed in §3.3.2 and §3.3.3, are subsequently up/down-sampled to a size of 64×64 using `PyTorch`'s `torch.nn.Upsample` class, in *bilinear* mode. Each moment map carries valuable information for the decoder functions as described in §3.2.4. With this in mind, the output of the encoders are two vectors which are concatenated before passing to the decoder subnet. For an in depth look at the encoder subnet structure see Table 3.1.

The encoders learn the following properties: *subnet 1*: observed galaxy inclination (i) and free parameters of the intensity profile which make up ξ_1 in Figure 3.1; *subnet 2*: the parameters of the velocity profile of the galaxy which make up ξ_2 in

Figure 3.1.

3.2.4 The decoder subnet

Here I detail the functions required for reconstructing the moment zero and moment one input maps from the concatenated feature representations ξ_1 and ξ_2 as shown in Figure 3.1. In recovering the moment maps, I am primarily interested in modelling two profiles. Firstly, the intensity:

$$\mathrm{I(r)} = \mathrm{I_0} \exp\left(-\frac{\mathrm{r_{x,y}}}{\mathrm{r_{scale}}}\right) \exp\left(-\frac{z}{\mathrm{r_{z\text{-}scale}}}\right), \tag{3.3}$$

where $\mathrm{I_0}$ is the intensity normalisation factor (set to 1 throughout, due to the global normalisation described above), $\mathrm{r_{x,y}}$ is the radius in the xy plane, in arcseconds, $\mathrm{r_{scale}}$ is the intensity scale length in the xy plane, z is the position in the z axis, and $\mathrm{r_{z\text{-}scale}}$ is the intensity scale length in the z axis set to a value of 1 spaxel throughout this work, to emulate a thin disk. Intensity values are determined by combining the integrals of Equation 3.3 across each spaxel in the xy and z planes.

Secondly, the rotational velocity:

$$\mathrm{V(r)} = \frac{2\mathrm{V_{max}}}{\pi} \arctan\left(-\frac{r}{\mathrm{r_{turn}}}\right), \tag{3.4}$$

where $\mathrm{V_{max}}$ is the asymptotic line of sight velocity, r is the radius in arcseconds, and $\mathrm{r_{turn}}$ is the velocity profile scale length.

Explicitly, in this work, i and $\mathrm{r_{scale}}$ are the parameters learned as ξ_1, with $\mathrm{V_{max}}$ and $\mathrm{r_{turn}}$ being the parameters learned as ξ_2 in Fig 3.1.

Here, our choice of exponential intensity profile and arctan velocity profile are entirely arbitrary (i.e. not driven by any physical theory), but are choices motivated by some of the simplest forms that can approximately fit the typical disks and rotation curves found in the Universe. Clearly objects that do not follow these functional forms will not be appropriately fit by this network and I discuss this further in §3.3.5. However, it should be noted that this analytical-style decoder implementation would be equally valid for other functional forms. For example, one could choose to fit bulge-disk models with such an architecture, or include the influence of central point masses or the effects of dark matter halos. These more realistic networks will be explored in future works.

An auxiliary 3D tensor of radii (labelled r in Figure 3.1) is passed into the network, cloned, and evaluated using Equations 3.3 and 3.4. The 2D moment maps are then created using Equations 3.1 and 3.2. The velocity profile is later converted

into a line of sight velocity map via an inclination projection and velocity weighting based on the pixel angles about the line of sight axis.

3.2.5 MODEL TRAINING PROCEDURE

The network is trained with minimal optimisation of hyperparameters in order to demonstrate the simple nature of this architecture. At all times the network utilises a `PyTorch`'s `MSELoss` function which computes the mean squared error:

$$\mathcal{L} = \frac{1}{\mathrm{N}} \sum_{\mathrm{i}=0}^{\mathrm{N}} \left(f\left(\mathrm{x_i}\right) - \mathrm{y_i} \right)^2 , \tag{3.5}$$

between the model outputs, $\mathrm{y_i}$, and inputs, $\mathrm{x_i}$, for every forward pass of a batch of size N. In this case, this is the squared difference between the moment zero and moment one inputs and decoder generated outputs. It is worth noting here that all synthetically generated moment maps have the same position angle and consequently any observational data used for training and testing have been de-rotated using published position angle measurements. I do this as position angle is a non-physical parameter which we can easily account for in preprocessing (with e.g. the `fit_kinematic_pa` routine of Krajnović et al. 2006).

I use an adaptive Adam learning rate optimiser (Kingma & Ba, 2014) which begins with a value of 10^{-4} and reduces via multiplication of 0.975 every 2 epochs. I find that the model converges well after 300 epochs for all training runs presented in this chapter.

Where synthetic training data is used, the network receives batches of 64 input moment map pairs. Initial tests showed the network to be largely unaffected by batch size and so 64 is arbitrarily chosen to increase training speed.

The models and Python training scripts used for the work presented in this chapter are publicly available on GitHub[6].

3.2.6 MODEL TESTING PROCEDURE

Testing the network can be done in three distinct ways, depending on the situation at hand. In order to test data, one can choose whether to train the network on the test data alone (I call this testing procedure *solo* testing), to train on the test data alongside other examples (I call this testing procedure *combined*), or to use the

[6] `https://github.com/SpaceMeerkat/Corellia/`

network in full test mode having only trained on examples not including those data that I wish to test (called *blind* testing).

One can imagine the case where sufficient training data has been passed through the network in a survey, such that in order to return rapid kinematic modelling of new observations one simply passes the new observations through the network with no prior exposure to the training procedure. This *blind* testing has the advantage of rapid testing speed but at the potential cost of lowered predictive accuracy, in an epistemic uncertainty dominated regime. One can also imagine the case whereby initial survey data has been collected and some sample of the dataset the network used to train is also in need of testing. As the network has seen these data during the training procedure, *combined* testing has the advantage of potentially higher accuracy at the expense of time needed to train the model. It should come as no surprise that the ideal testing scenario for this network is *combined*, with a sufficiently large training set in an aleatoric uncertainty dominated regime. However, there are cases (such as at first light of a survey) where the only test data available is that which the network was trained on. It is in this scenario that *solo* testing will occur and although this testing regime lacks the benefits afforded by *combined* testing, it has the potential advantage of predictions not being influenced by anomalous data whose population increases with training set size.

3.2.7 *Monte Carlo* dropout

In this section I summarise the use of *Monte Carlo* dropout (henceforth MC dropout; Gal & Ghahramani 2016) to provide quasi-statistical modelling uncertainties over learned parameters within the model.

In conventional neural network training circumstances *dropout* may be interpreted as permuting a trained model (Srivastava et al., 2014) via the probabilistic zeroing of weights in linearly connected layers. Traditionally, dropout layers are used throughout training in order to force the network to behave as an ensemble of architectures, increasing testing accuracy and generalisation power. In the case of MC dropout, after training, dropout is reapplied to the network in evaluation mode and inputs are passed through the model many times, effectively sampling a posterior where the model architecture is marginalised out. Gal (2016) first proposed the idea of approximating distributions over parameters learned in neural networks in this way and has since been used in astronomy (e.g. for the probabilistic labelling of galaxy morphologies, Walmsley et al. 2019).

For an input x (comprised of a moment 0 and moment one map), training

data $\mathcal{D}$, model weights w, T forward-pass evaluations, and encoder output k, the predicted parameter means and standard deviations are given by Equations 3.6 and 3.7 respectively.

$$\hat{\mathrm{k}} = \frac{1}{\mathrm{T}}\sum_{\mathrm{t}=1}^{\mathrm{T}} P(\mathrm{k}|\mathrm{x}, \mathrm{w_t}) \tag{3.6}$$

$$\sigma = \sqrt{\frac{\sum_{\mathrm{t}=0}^{\mathrm{T}} (\mathrm{k} - \mathrm{k_t})^2}{\mathrm{T}}} \tag{3.7}$$

For a comprehensive derivation of Equations 3.6 and 3.7, as well as the implications for using an arbitrary dropout probability, I refer the reader to Walmsley et al. (2019). Examples of the posterior distributions, p(k|w,$\mathcal{D}$), over learned parameters using MC dropout for a randomly selected synthesised galaxy are described further in §3.3.1.

It should be noted that, as the network does not use dropout to zero weights in the convolutional layers, σ does not represent a complete error over learned parameters. Instead one should consider σ as a lower limit error over parameters whose use becomes immediately obvious for pipeline flagging purposes or to generate relative errors within a test set. The errors produced through this technique are strictly errors due to the modelling technique, and will underestimate the true error in any parameter, which arises due to both modelling and observational uncertainties.

3.3 RESULTS AND DISCUSSION

In this section I present exemplar test results for highly spatially resolved galaxy observations. In each case I have trained new networks using the procedures described in §3.2.5.

3.3.1 SYNTHESISED EXAMPLES

Input-output

In order to explore the limitations of the network, I tested the model using synthetic galaxies generated using the `Python` based kinematic simulator `KinMS`[7] (KINematic Molecular Simulation, Davis et al. 2013; Davis et al. 2020).

Figure 3.2 shows the inputs and outputs as well as both known and predicted profiles for a galaxy generated using the same analytical functions described in §3.2.4

[7] `https://github.com/TimothyADavis/KinMSpy`

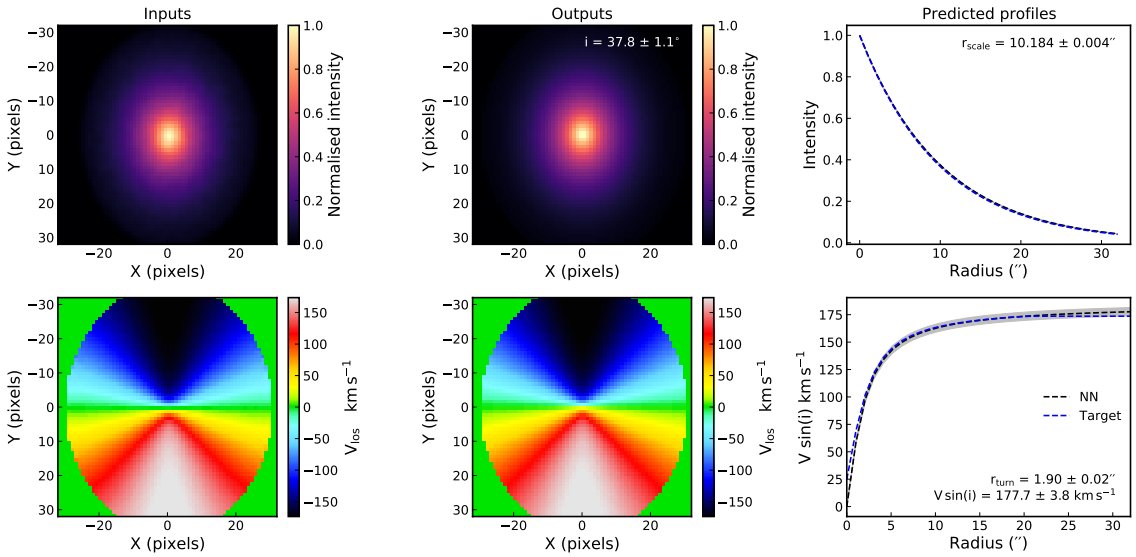


Figure 3.2. A randomly selected synthesised galaxy, created using `KinMS` and evaluated using the network in *blind* testing mode. The black dashed lines and grey areas show the mean and 1σ modelling uncertainties respectively for profiles predicted by the neural network model. The blue dashed lines show the target profiles which were used to create the input maps. The galaxy was created with the following known parameters: i= 37.2°, $r_{scale} = 10.0''$, $r_{turn} = 1.6''$, and $V_{max}\sin(i) = 173.6$ km s^{-1}. The network predicted parameters are shown as text in the upper-middle, upper-right, and lower-right subplots.

with inclination, maximum velocity, and scale lengths drawn randomly in the ranges shown in Table 3.2, and a fixed beam size of 2 resolution elements.

Table 3.2 Parameter values and ranges for all synthetically generated galaxies using the `KinMS` package. The units for r_{scale} and r_{turn} are absent due to both quantities being fractions of the input map size. The position angle of each galaxy is fixed at 0 as it is not a physically meaningful parameter. Throughout model training, parameters are drawn uniformly in the ranges listed.

Parameter	Size/range	Units
Position angle	0	deg
Inclination	10–90	deg
r_{scale}	0.1–0.35	–
r_{turn}	0.01–0.8	–
$V_{max}\sin(i)$	50–500	km s^{-1}

It is clear that the model is able to recover the galaxy's rotation curve (and other parameters) well in *blind* testing mode, whereby the model has not yet trained on the test data. The quasi-probabilistic distributions for each learned parameter

for this galaxy are shown in Figure 3.3, highlighting the Gaussian-like nature of the learned parameter distributions as well as an expected covariance between r_{turn} and $V_{max}\sin(i)$. For a discussion of the population accuracy, please see Figure 3.4 and associated text.

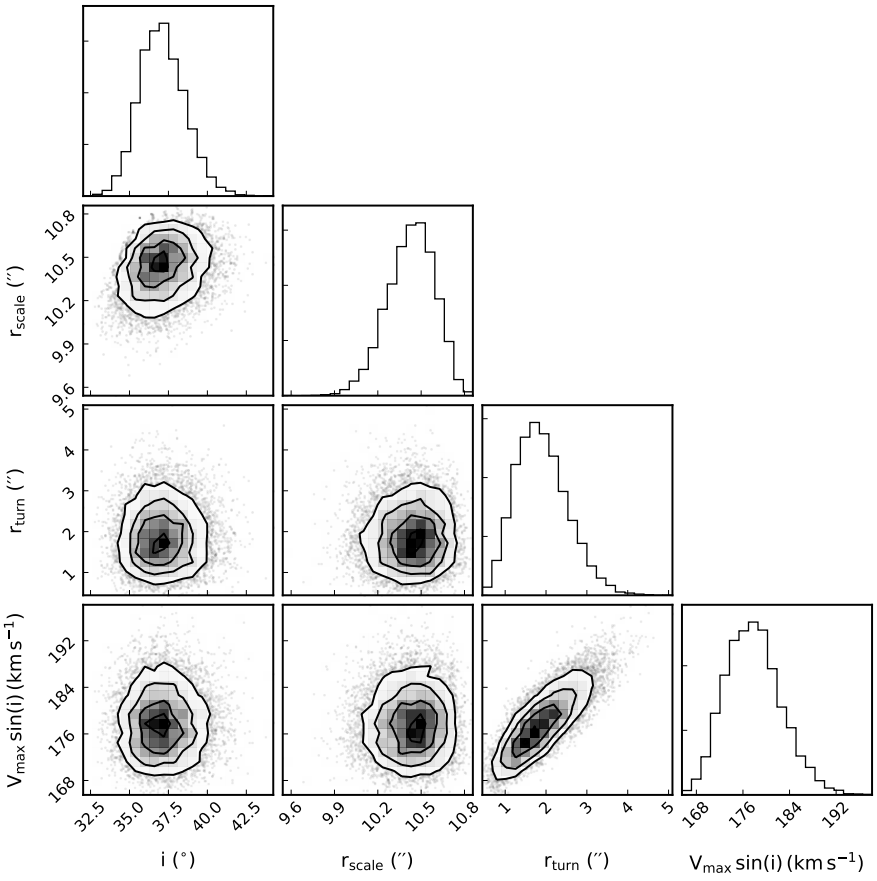


Figure 3.3. Corner plot showing the level of covariance between learned parameters for one randomly generated, synthesised galaxy (discussed further in §3.3.1). The accompanying histograms represent quasi-probabilistic distributions thanks to the use of *Monte Carlo* dropout. This galaxy was passed through the network in test mode 10 000 times in order to build the distributions. I observe well constrained learned parameters with Gaussian like profiles, allowing for quasi-probabilistic modelling errors for the parameters. The only strong covariance observed is that between the maximum line of sight velocity and the velocity profile scale length, which is entirely expected and present in traditional kinematic analyses.

As seen in Figure 3.4 the model is able to recover the desired physical parameters of synthesised galaxies well, heuristically. For the 1739 test galaxies shown in Figure 3.4 I measure the average deviation of parameters: i, r_{scale}, r_{turn}, and $V_{max}\sin(i)$, from the 1:1 line as $\sigma_i = 0.98°$, $\sigma_{r_{scale}} = 0.003$, $\sigma_{V_{max}\sin(i)} = 3.48\,\mathrm{km\,s^{-1}}$, and $\sigma_{r_{turn}} = 0.017$ respectively.

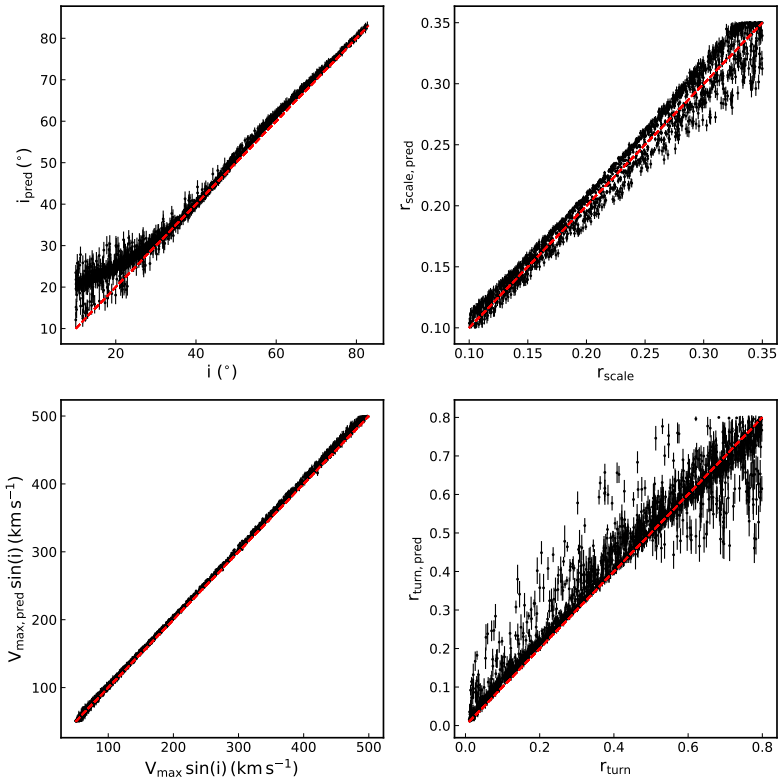


Figure 3.4. True versus predicted plots for each learnable parameter in the network. Black markers and error bars pertain to the tested galaxies and the red dashed line indicates the 1:1 line on which perfect predictions should ideally lie. This model was trained using purely synthetic data with a restoring beamsize of 2 resolution elements and only including well resolved examples as discussed in §3.3.1. Those galaxies whose projected r_{turn} fell below 1.5 times the restoring beamsize were removed in order to mimic the automated flagging of poorly resolved galaxies at high inclination in a survey. Of the 2000 synthesised galaxies tested, 261 (13%) were removed using this cut.

It is clear from Figure 3.4 that the generated error estimates for a population of test galaxies do not represent the total errors over the parameters and only encode the modelling errors, i.e. a fraction of the total error comprised of observational and modelling errors. This makes the presented errors strictly lower limit estimates, and mostly useful for comparing reliability within the dataset, rather than external use. This can be seen by the fact that on average only ∼35% of the data points in Figure 3.4 have errorbars which overlap with the 1:1 true-versus-predicted line. For the presented dataset these errors likely underestimate the total error by a factor of ∼2.5. Including errors in the observations themselves will help to narrow this gap and will be explored further in future work. Note that the systematic increase in modelling errors at lower inclinations is fully expected due to loss of line of sight

information as i approaches 0°.

The effect of resolution

One expects $r_{scale,pred}$ to artificially increase with beam size for a fixed r_{scale}. However, r_{scale} is not known for observations of galaxies whose values r_{scale} fall below some fraction of the beamsize. I see this effect happening as shown in Figure B.1 (removed from the main text here due to the use of a fixed beamsize throughout this chapter) in a non-complex manner. Therefore, I recommend enforcing flagging based on inclination which appears to be strongly linked with those galaxies whose r_{scale} is under predicted (along the minor axis). In the edge-on galaxy case, the minor axis is no longer well resolved resulting in a poor recovery of the intensity profile. However, this is a well-known issue in moment based kinematic modelling, in which the intensity profiles and kinematics can never be fully derived in edge-on galaxies due to line of sight effects.

As I have included no method for mitigating the changes induced by varying beam size, it comes as no surprise that the network will behave differently given a sufficiently large ratio of beam size to galaxy extent. Given that I do not have a mechanism for dealing with "beam smearing" in the current network architecture, I expect to see its influence, lowering the apparent line of sight velocities close to the center of galaxies where the iso-velocity contours are closest together. For minimising the effects of varying beam size I recommend convolving the 3D spatial cube r (see Figure 3.1), evaluated using Equation 3.3, with the restoring beam before creating the output maps. The advantage of this approach being that the restoring beam is often included in data-product header units, and so should be readily available for creating kernels with which to perform the aforementioned convolution. I consider this approach as beyond the scope of the work presented in this chapter, but will be included in future work focusing on retrieving the properties of marginally resolved galaxies.

Fill factor

In previous work I showed that the fill factor (i.e. the number of zeroed pixels) in a velocity map's field of view, impacts the behaviour of NN models which take them as inputs (Dawson et al., 2019). With the NN model presented in this work, I have seen little evidence that this has an effect on the galaxies' predicted parameters. I attribute this behaviour to the nature of the training procedure, whereby in *combined* and *solo* testing, the network does not rely solely upon inference of unseen data.

3.3.2 HI examples

The primary goal of developing a network like that presented in this chapter is to demonstrate the applicability of machine learning to SKA science. As such, in this section I show the network performs well when training and testing on self-contained HI observational data. In order to do this I present two example test galaxies, NGC 2403 and NGC 3198, observed using the **V**ery **L**arge **A**rray (VLA) as part of **T**he **Hi** **N**earby **G**alaxy **S**urvey (THINGS) (Walter et al., 2008), and showing a diversity of rotation curve shapes. These galaxies are two of 17 THINGS galaxies used for *mixed* training and testing using the network and chosen heuristically for the appearance of their well defined rotating HI disks. The names and publications for the galaxies used in this sample are shown in Table A.4.

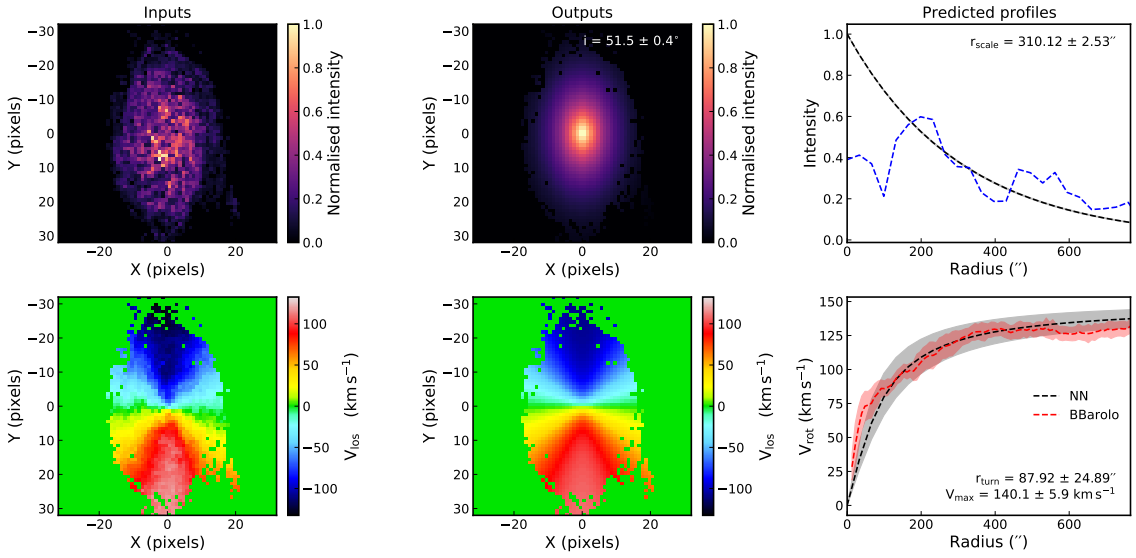


Figure 3.5. An example galaxy, NGC 2403, observed in HI and evaluated using the network in *combined* testing mode. Maps in the left and middle columns share x and y axis sizes of 64×64 pixels. In this way I am directly observing the input and output maps of the model. The right column has undergone an x-axis rescaling to match observational scales found in the literature. The black dashed lines and grey areas show the mean and 1σ modelling errors respectively for profiles predicted by the neural network. The blue dashed line shows a major axis cut of the input intensity map. The red dashed line and filled area show the best fit and associated errors modelled using `BBarolo` on the datacube. In order to make a direct comparison between the network's and `BBarolo`'s derived rotation curves, the network's velocity profile has been corrected for by the predicted inclination term. The network predicted parameters are shown as text in the upper-middle, upper-right, and lower-right subplots. I see that this galaxy has a velocity profile which can be roughly approximated by an arctan function meaning the kinematic parameters are well recovered by the model.

Figure 3.5 shows the derived intensity profile and rotation curve for NGC 2403. I include the rotation curve modelled using `BBarolo` (Di Teodoro & Fraternali, 2015) on the datacube (Di Teodoro & Lelli, private communication). In comparison, I see that the neural network's predicted rotation curve matches closely and so I am convinced that the network is able to recover physical information well. Although the galaxy's intensity profile does not strictly exhibit an exponential form, this has little impact in the recovery of the rotation curve which is the networks primary objective.

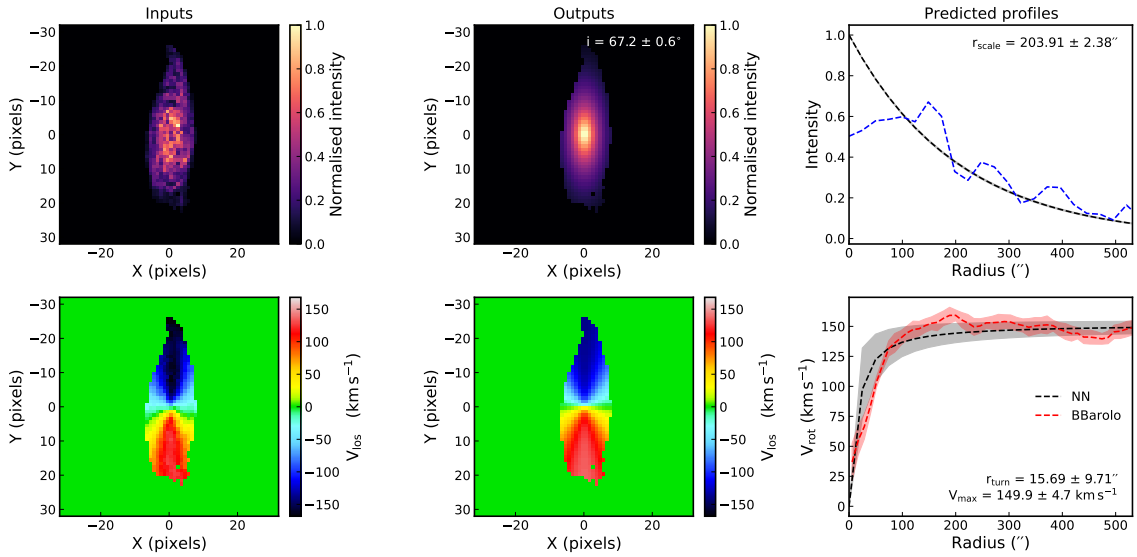


Figure 3.6. An example galaxy, NGC 3198, observed in HI and evaluated using the network in *combined* testing mode. Maps in the left and middle columns share x and y axis sizes of 64×64 pixels. In this way I am directly observing the input and output maps of the model. The right column has undergone an x-axis rescaling to match observational scales found in the literature. The black dashed lines and grey areas show the mean and 1σ modelling errors respectively for profiles predicted by the neural network. The blue dashed line shows a major axis cut of the input intensity map. The red dashed line and filled area show the best fit and associated errors modelled using `BBarolo` on the datacube. In order to make a direct comparison between the network's and `BBarolo`'s derived rotation curves, the network's velocity profile has been corrected for by the predicted inclination term. The network predicted parameters are shown as text in the upper-middle, upper-right, and lower-right subplots. I see that this galaxy has a velocity profile which can be roughly approximated by an arctan function meaning the kinematic parameters are well recovered by the model.

Figure 3.6 shows the derived intensity profile and rotation curve for NGC 3198. This galaxy exhibits a mild warp and a flat rotation curve (Gentile et al., 2013) with a slight rise at $\sim 200''$. Warped HI discs are not uncommon in the outer regions of galaxies. At present our network architecture is not set up to model these (however one could easily extend the model in order to do so). Again, I include

the rotation curve modelled using `BBarolo` on the datacube (Di Teodoro & Lelli, private communication) in Figure 3.6. Crucially, although this warping behaviour is not included in our model, in this case the network still returns reasonable parameter estimations, showing that it could still be usable for parameter estimations across a broadly diverse population of galaxies.

3.3.3 CO examples

In order to demonstrate the flexibility of this network architecture, I trained a model to recover the kinematic properties of galaxies observed in the CO line using the **A**tacama **L**arge **M**illimeter/submillimeter **A**rray (ALMA). My samples are drawn from the mm-**W**ave **I**nterferometric **S**urvey of **D**ark **O**bject **M**asses (WISDOM) project (see Table A.5 for more information) and have high spatial resolution. Due to the nature of these objects being targeted for their evidence of black hole influence on the gas kinematics, I expect to see small values of a_V for the sample. As seen in Figure 3.7, this effect is clearly visible, highlighting the predictable behavioural nature of the network. It is also clear in Figure 3.7, that NGC 1387 (FCC184, Zabel et al. 2020, Boyce et al., in prep), an exemplar galaxy from the WISDOM sample, exhibits an exponential intensity profile which the network can easily recover. It is worth noting the low inclination returned by the network for this example galaxy. This is not unexpected given the very circular appearance of NGC 1387. In fact, 10° is the lower limit of permitted inclinations provided by the network and provides the first inclination estimate for NGC 1387's CO gas disc.

Such an example demonstrates the *transferable* nature of this network architecture and training style but without the difficulties often associated with traditional transfer learning tasks. This means that such architectures and training styles can be applied to a multitude of different datasets with the possibility of architectural modifications suiting other types of data outside of interferometry and even astronomy.

3.3.4 Testing speed

The network can retrieve a mean field approximation for all learnable parameters, of a single galaxy observation, in 0.0025 seconds on a single Intel(R) Core(TM) i7-6700 CPU core. This time scales linearly with the number of MC dropout samples one wishes to collect (i.e. for a set of 1000 MC dropout samples, a typical test on an individual galaxy would take 2.5 seconds) to generate pseudo-probabilistic distributions. However as the batch throughput size is limited only by the available device

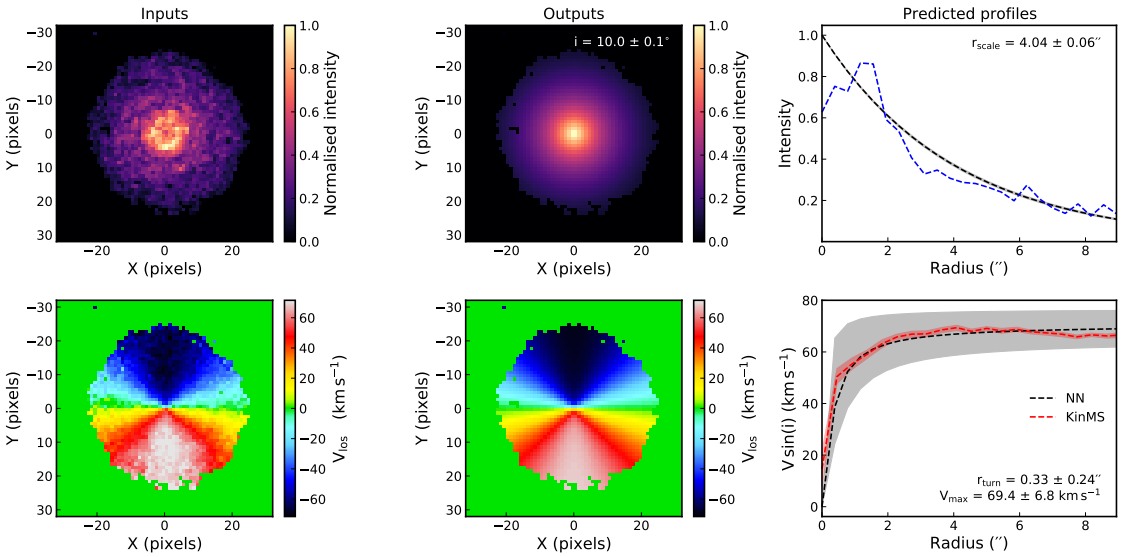


Figure 3.7. An example WISDOM galaxy, NGC 1387, observed in CO and evaluated using the network in *combined* testing mode. The left and middle columns share x and y axis sizes of 64×64 pixels. In this way I am directly observing the input and output maps of the model. The right column has undergone an x-axis rescaling to match observational scales found in the literature. The black dashed lines and grey areas show the mean and standard deviation respectively for profiles predicted by the neural network model. The blue dashed line shows a major axis cut of the input intensity map. The red dashed line shows the `KinMS` reconstructed rotation curve. The network predicted parameters are shown as text in the upper-middle, upper-right, and lower-right subplots. I easily see that this galaxy has an intensity profile and velocity profile which can be roughly approximated by an exponential and an arctan function respectively, meaning the kinematic parameters are well recovered by the model.

memory, it is possible to retrieve values for learnable parameters, and hence MC dropout samples, in the same time frames as listed above for multiple observations. This means that one could potentially return hundreds to thousands of parameterisations and associated pseudo-errors in a matter of seconds.

3.3.5 CAVEATS

There are a few caveats pertaining to the use of the model described in this work. These caveats may impact the way in which users handle the network and the confidence levels associated with parameter estimations.

A key factor in recovering sensible parameterisations using the network is the choice of decoder functions (see §3.2.4). In this work I have used simple, general, functions in the form of an exponential (see Equation 3.3) and an arctan (see Equation

3.4). However, should one wish to model specific emission line components of galaxies, it would be prudent to adopt more tailored functional forms. For example, it has been shown that HI discs can display depressions in their intensities in their central regions, typically filled by molecular gas (Wong & Blitz, 2002), for which a truncated Gaussian intensity profile (Martinsson et al., 2013) would be more appropriate when reconstructing the intensity maps. Additionally, when modelling the very outer regions of HI discs, one might consider adopting a more complex multi-parameter function capable of encoding the *sharpness* of the turnover at r_{turn} and the behaviour of the curve after this point (e.g. Rix et al. 1997), or even declining velocities in the central regions (Lelli et al., 2016). A declining rotation curve would be challenging for the current model to fit (and impossible to fully retrieve). However, due to the nature of the loss function chosen in this work (see Equation 3.5), the network will prioritise fitting to the higher velocity regions of galaxies.

As described in §3.3.1, the resolution of input images impacts the ability of the network to correctly predict a_{scale}, particularly in the high inclination regime. This places constraints on the user's confidence in parameter estimations when working in both the large-beam and high inclination cases combined. Additionally, we can see in Figure 3.4 that the network struggles to accurately recover inclinations at the very low inclination range. This is a predictable effect caused by the loss of line of sight velocity information for face on disks but again, in the case of survey pipelines, these low inclined galaxies will require additional flagging. In both the aforementioned caveat cases it is worth noting that traditional kinematic modelling methods also struggle to accurately estimate parameters, in particular when working with moment maps. Extensions of the network's framework presented here to kinematically model datacubes may alleviate these issues and will be explored in future work.

3.4 Conclusions

I have demonstrated the performance of a neural network model architecture which can be used to recover rotation curves of galaxies from their kinematics. The model was tested on synthetically generated galaxies as well as observations using both HI and CO emission lines.

Testing on synthetically generated galaxies has highlighted the powerful performance of the network as well as areas where the network's performance is sub optimal. For the latter areas I have discussed solutions including: an additional convolution with the restoring beam to counteract the effects of "beam smearing", and

flagging high inclination data in a large beam and high inclination regime.

Testing observational HI data from THINGS has shown that this style of network is well suited to work with data like that expected from the SKA in the near future. I have shown that the network is capable of estimating velocity curves for discs exhibiting a variety of profiles. In order to do this, I have directly compared the rotation curves estimated by the network to those modelled directly from the cubes using kinematic modelling tools. The network is able to perform adequate recovery of parameters even in cases where it would not be possible to reproduce the true rotation curves. These promising results give us confidence that adopting more flexible decoder functions will extend the applicability of the model for more specific use cases should one wish to model HI discs exclusively.

Testing observational CO data from the WISDOM project has shown that the network is suitable for a range of emission line observations. Unlike traditional ML models, the network architecture and training styles outlined in this work prevent the need for *transfer learning* which is often time consuming and fraught with ungainly challenges associated with systematic properties of training sets. I have shown that the model outlined in this work can recover rotation curves which heuristically match rotation curves extracted from ALMA observations using more time-consuming approaches.

As previously stated, improvements to the model architecture in this work include but are not limited to: adapting the model to use more complex intensity and velocity profiles in the decoder subnet, automatically accounting for large beam effects such as beam smearing and information loss either via systematic offsets in model predictions or via the incorporation of an extra convolutional layer in the decoder subnet, and reintroducing a position angle estimation step. An idealised improvement on the model would be to work directly with interferometric datacubes themselves, or even visibilities, without the need to generate moment maps prior to training and testing. However, I have found that the discretised nature of channels in interferometric datacubes presents a non-gradient-trackable step in the decoder's reconstruction of datacube inputs. This discontinuity in the gradient tree prevents back propagation via gradient descent and consequently halts model training. I propose adapting this self-supervised approach to work with datacubes as a lucrative avenue of research for challenging current kinematic modelling tools in preparation for the SKA and other upcoming large facilities.

Chapter 4
The stellar mass Tully-Fisher relation with SAMI & MaNGA using self-supervised, physics-aware, Bayesian neural networks

> "Too many dots," Miller said.
> "Not enough lines."
>
> James S.A. Corey, *"Leviathan Wakes"*

In this chapter, I adopt the use of self-supervised physics-aware Bayesian neural network models, to retrieve the kinematic properties of galaxies, observed with IFU instruments in a time-efficient manner. In doing so, I combine SAMI DR2 and MaNGA DR16 IFU survey data to subsequently recover the Hα Tully-Fisher relation (TFR) for 1834 low redshift galaxies. My best fit stellar mass Tully-Fisher relation has the form $\log(M_*/M_\odot) = 3.013 \pm 0.037 \times (V_{max}/\mathrm{km\,s^{-1}} - 2.256) + 10.054 \pm 0.008$. My fits are found to be in agreement with those presented in the wider literature, with the exception of a somewhat shallower slope, which arises because I here include corrections for asymmetric drift. I identify and quantify trends between position along (and perpendicular to) the TFR and galaxy properties, namely: age and mass-to-light ratio. I also quantify a clear relation between velocity turnover radius, r_{turn}/r_e, and stellar mass. This work represents the first time the Hα TFR has been recovered on such a large collection of IFU data-products, and using machine learning. Such methods are efficient, and will be powerful when applied to next generation IFU survey data

releases, with instruments such as HECTOR.

4.1 INTRODUCTION

The Tully-Fisher relation (TFR, Tully & Fisher 1977) is a fundamental scaling relation between luminosity (McGaugh et al., 2000) and the rotational velocity of galaxies. In the dark matter paradigm, the rotational motion in the outer parts of galaxies is primarily determined by the mass of the dark matter halo within which the galaxies reside. Therefore, the TFR can be considered as displaying a connection between the visible-baryonic and dark matter masses (Douglass et al., 2019). In the nearby Universe, the TFR has been well studied (e.g. Tully & Pierce 2000; Bell & de Jong 2001; Kassin et al. 2007; Masters et al. 2008; Lagattuta et al. 2013). The TFR has been widely used as a distance indicator, relating distance-independent velocity measurements of galaxies to absolute magnitude (e.g. Opik 1922; Roberts 1969; Bottinelli 1971; Balkowski et al. 1974; Shostak 1975). The TFR exists in several different forms, depending on which measures of mass, luminosity, or rotational velocity one chooses to use. Originally, Tully and Fisher used optical luminosity, but subsequent work showed the relation to be tighter when defined using longer wavelength radiation and even tighter when replacing luminosity by a galaxy's total baryonic mass (McGaugh et al., 2000). The latter form of the relation is known as the baryonic Tully-Fisher relation (BTFR). The TFR has traditionally been derived using observations of HI, however in recent decades, the use of optical spectroscopy has also played a role in constraining the TFR at different wavelengths.

In the past, spectroscopic measurements of rotational velocity and velocity line widths, were mostly taken using a single fibre or slit (e.g. York et al. 2000; Percival et al. 2001; Driver et al. 2009). However, single fibre, and slit-based, measurements are vulnerable to induced errors based on slit placement and aperture effects (Spekkens et al., 2005; Oh et al., 2011; Simons et al., 2015; Bloom et al., 2017b). Measuring spatial variations across an extended source is also challenging with a single slit. Therefore, 2-dimensional, spatially resolved, kinematics provide a better approach to circumvent this problem. Integral field spectrographs are one of several instruments capable of observing 2-dimensional, spatially resolved kinematics of extended sources. Example instruments leading the charge are the Sydney–AAO Multi-object Integral field spectrograph (SAMI), and the Mapping Nearby Galaxies at APO (MaNGA) spectrograph. Both instruments allow spatially resolved observations of multiple galaxies in a single pointing, drastically decreasing the timescale of survey coverage when compared with single fibre integral field spectrographs (Croom

et al., 2012).

The TFR has been explored using 2-dimensional spatially resolved spectroscopy in a variety of ways. At low redshifts, the ATLAS^{3D} survey has been used to show the carbon monoxide (CO) TFR for early type galaxies (Davis et al., 2011). The Hα stellar mass TFR was studied in the K-band Multi-Object Spectrograph (KMOS) Redshift One Spectroscopic Survey (KROSS), with galaxies out to redshifts of $z \sim 1$ (Tiley et al., 2016). Recently, the CALIFA Survey have produced a low redshift TFR using rotation curve fitting to stellar velocity fields (Bekeraité et al., 2016). Bloom et al. (2017b) studied the stellar mass TFR using a sample of galaxies from the SAMI galaxy survey, investigating the relationship between stellar mass and kinematic asymmetry and demonstrating the use of 2D spatially resolved kinematics for accurate TFR studies. The HI TFR has also been studied to great extent using kinematic maps from HI interferometric surveys (e.g. Begum et al. 2008; Stark et al. 2009; Trachternach et al. 2009; Oh et al. 2011).

In this chapter I make attempts to recover the Hα TFR, using *big data era* approaches for the first time. Combining datasets from multiple IFU surveys, I attempt to recover the TFR for the largest IFU sample to date, with the goal in mind of preparing for the next generation of IFU instruments. In the near future, IFU instruments (e.g. HECTOR-I and HECTOR-II, Bryant & Bland-Hawthorn 2016) will collect 2-dimensional, spatially resolved, spectroscopic observations for more galaxies than any other IFU survey to date. I therefore begin the process of building on the domain of existing machine learning applications for exploring the kinematics of IFU survey objects (e.g. Sarmiento et al. 2021). I adopt machine learning methods originally intended for use with resolved millimetre and radio observations, using radio interferometers, and explore their use with IFU measured kinematics. Specifically, I will use Hα integrated intensity and velocity maps of galaxies from SAMI and MaNGA surveys, to recover kinematic parameter estimations for more galaxies than achieved to date.

This chapter is divided into 3 main sections. §4.2 details the astronomical data used throughout the chapter and how that data is used to recover the Hα TFR. This includes the quality control measures applied to the chosen samples as well as the techniques used to model kinematic characteristics of galaxies. Emphasis is placed on recovering the TFR using machine learning approaches in a time-efficient manner. The results of this work are then discussed in §4.3, showing the TFR alongside comparisons to those calculated in the wider literature. Here I also discuss some model behaviour and relevant correlations between derived parameters. Finally, in §4.4 I

summarise the main outcomes of the work presented in this chapter and propose future work.

4.2 SAMPLE AND METHODS

4.2.1 THE SAMI INSTRUMENT AND SAMI GALAXY SURVEY DR2 SAMPLE

The SAMI Galaxy Survey (Bryant et al., 2015) is a spatially resolved survey of nearby ($z \lesssim 0.1$) galaxies. The SAMI instrument is a multi-object Integral Field Spectrograph (IFS) installed at the prime focus of the Anglo-Australian Telescope (AAT) feeding the double-beamed AAOmega spectrograph (Croom et al., 2012; Sharp et al., 2006). The SAMI instrument uses 13 fused optical imaging fibre bundles, or hexabundles (Bland-Hawthorn et al., 2011; Bryant et al., 2013), with each 61 fibre core subtending $\sim 1.6''$ on the sky. Together, this gives SAMI a total bundle diameter of $15''$ which can be deployed over a 1° field of view to simultaneously image 12 galaxies and 1 standard star at a time. An additional 26 fibres provide simultaneous blank sky observations.

Throughout the work presented in this chapter, I make use of resolved 2D Hα maps, from the SAMI survey. SAMI IFU observations of Hα line emission make use of the AAOmega spectrograph's 1000R grating. This red arm has a central wavelength of 680nm and a spectral range from 625 to 735 nm. The spectral resolution in the red arm is $R = 4500$ ($\lambda/\delta\lambda$). For a full description of the SAMI Galaxy Survey data reduction pipeline, I refer the reader to Allen et al. (2015); Sharp et al. (2015), and Scott et al. (2018).

The SAMI Galaxy Survey consists of two complementary samples of galaxies, with matched selection criteria. In total, 3400 galaxies make up SAMI data release 2 (DR2). The selection of galaxies for SAMI DR2 is described in detail in Bryant et al. (2015) and Owers et al. (2017). The two samples consist of a SAMI-GAMA sample drawn from the Galaxy And Mass Assembly (GAMA) survey (Driver et al., 2011) and an additional sample drawn from 8 galaxy clusters (Owers et al., 2017) for completeness. The SAMI-GAMA sample consists of a series of volume-limited samples, with galaxies drawn from the three 4×12 degree fields of the GAMA-I survey. Galaxies in these regions reside in a variety of environments from isolated up to massive groups, but do not include any galaxy cluster members within $z \lesssim 0.1$. The cluster sample galaxies used to supplement the total SAMI sample adhere to the same mass selection and redshift limits as the GAMA sample but, in practice, cluster

galaxies of different masses are targeted based on their redshifts (e.g. Scott et al. 2018).

Observations for the SAMI Galaxy Survey (up to DR2) began in March 2013 and were completed in May 2018. At the completion of the survey, ~ 3000 galaxies have been observed to a completeness of 80% and 84% for the GAMA and cluster samples respectively (Scott et al., 2018), with 1930 and 724 unique primary targets in each of the samples. Of those 3400 observations, data for 1559 targets has been released as the DR2 sample, comprising roughly 50% of the full survey.

It is worth noting that an additional sample of secondary targets with slightly lower mass cuts in each redshift bin were also included in the SAMI Galaxy Survey observing run. These targets were observed when a hexabundle could not be allocated to a primary target. In the final block of observations, an extra set of ancillary targets, primarily drawn from GAMA galaxies, was needed in order to fill all hexabundles. These targets did not meet the stellar mass cuts of the original selection criteria and are not included in DR2 and are therefore not included in this work either.

4.2.2 The MaNGA instrument and SDSS-IV MaNGA sample

MaNGA operates on the Sloan Digital Sky Survey (SDSS) $2.5\,m$ telescope (Gunn et al., 2006) at Apache Point Observatory (APO). The instrument itself has seventeen science IFU hexabundles, ranging in size from 19 to 127 fibres, with twelve 7-fibre minibundles used for calibration purposes (Yan et al., 2016) and 92 single fibres used for sky subtraction with diameters between $12.5''$ and $32.5''$ on the sky respectively (Drory et al., 2015; Law et al., 2015).

Optical fibres from the MaNGA instrument feed the twin Baryonic Oscillation Spectroscopic Survey (BOSS, Dawson et al. 2013; Smee et al. 2013) spectrographs, rebuilt from the original SDSS spectrographs, which were used for the SDSS Legacy and SEGUE surveys (York et al., 2000; Yanny et al., 2009). The spectral resolution ranges from $R = \lambda/\delta\lambda \sim 1800$ at 60 nm to $R \sim 2200$ at 1030 nm, in the red channel, encompassing Hα line emission (see Fig. 36 of Smee et al. 2013). The spectral resolution increases linearly towards longer wavelengths. Spectra are processed and reconstructed into 3-dimensional datacubes using a software pipeline (Law et al), adapted from the pipelines previously used for BOSS (idlspec2d, Bolton et al. 2012).

The goal of the SDSS target selection is to observe a sample of $\sim 10,000$ galaxies over a period of 6 years. SDSS is up to its sixteenth data release (DR16, Ahumada et al. 2020), the fourth data release of SDSS-IV. The proposed total sample

is divided into 'primary', 'secondary', and 'ancillary' groups, based on radial coverage goals. The primary selection, accounting for $\sim$ 5000 of the $\sim$ 10,000 target galaxies, have radial coverage of out to 1.5 effective radii (R_e) for $\sim$ 80% of the sample. The secondary selection accounting for $\sim$ 3300 galaxies, have radial coverage out to $2.5\,R_e$ for $\sim$ 80% of the sample. MaNGA targets have a fairly flat distribution of stellar masses with $M > 10^9\,M_\odot$, out to a redshift range of $0.03 < z < 0.1$. The remaining 10% of the total MaNGA sample is comprised of ancillary targets of high value (Bundy et al., 2015). Having observed more than one-third of the entire sky, there are currently 4824 3-dimensional spectroscopic cubes from observing galaxies with MaNGA as part of DR16.

4.2.3 PARAMETER ESTIMATIONS USING MACHINE LEARNING

I aim to recover kinematic parameters directly from Hα integrated intensity and velocity maps of galaxies from SAMI and MaNGA surveys combined. In order to do so in a time-efficient manner, I make use of machine learning models which can perform physics-aware, self-supervised learning via receiving moment maps as model inputs.

In order to construct the TFR (see §4.3), we must first recover the asymptotic value of the velocity profile, V_{max} , from the combined SAMI and MaNGA datasets. In order to retrieve V_{max} in a time-efficient manner, I employ the use of a self-supervised physics-aware neural network as described in Dawson et al. (2021b). The model builds on similar approaches by Aragon-Calvo & Carvajal (2020), where covolutional autoencoders (Masci et al., 2011) are used as *semantic autoencoders*, thanks to the use of physically motivated *decoder subnets*. The model used in this work is identical to that used by Dawson et al. (2021a), including the use of two decoder functions to perform input map reconstruction from the embeddings in the neural network model. Firstly, I assume that the ionised gas discs in galaxies can be described by a 3D exponential disc profile:

$$I(r) = I_0 \exp\left(-\frac{r_{x,y}}{r_{scale}}\right) \exp\left(-\frac{z}{r_{z\text{-}scale}}\right), \tag{4.1}$$

where I_0 is the intensity normalisation factor (set to 1 throughout). The intensity profile is applied to a 3-dimensional xyz spatial cube for which: $r_{x,y}$ is the radius in the xy plane, in arcseconds, r_{scale} is the intensity scale length in the xy plane, z is the position in the z axis, and $r_{z\text{-}scale}$ is the intensity scale length in the z axis set to a value of 1 cube-spaxel throughout this work, to emulate a thin disk. Intensity values

are determined by combining the integrals of Equation 4.1 across each cube-spaxel in the xy and z planes, with this 3-dimensional distribution being rotated and projected based on the fitted inclination and position angle of the galaxy to compare with the Hα integrated line intensity map.

Secondly I assume the rotation curves of my sample galaxies can be described by a simple velocity profile:

$$\mathrm{V(r)} = \frac{2\mathrm{V_{max}}}{\pi} \arctan\left(-\frac{r}{\mathrm{r_{turn}}}\right), \tag{4.2}$$

where $\mathrm{V_{max}}$ is the asymptotic line of sight velocity, r is the radius in arcseconds, and $\mathrm{r_{turn}}$ is the velocity profile scale length.

The choice of exponential intensity profile and arctan velocity profile are not driven by any physical theory. Instead, they are choices motivated by some of the simplest forms that can approximately fit the typical disks and rotation curves found in the Universe. Objects that do not follow these functional forms will not be appropriately fit by this network and as such, I make attempts to remove the effects of such poorly-fitting cases before constructing my results, as described in §4.2.4. Given that Bloom et al. (2017b) used arctan functions of similar form to Equation 4.2 (Courteau, 1997) for rotation curve fitting to SAMI galaxies, this approach allows us to make directly comparable results.

As a branched neural network (as presented in Chapter 3), the two *encoder subnets* independently receive a 2D intensity map (normalized in the range 0–1) and a line of sight velocity map (normalized into the range -1–1) respectively. Throughout this work, I ensure that the input maps have size of 64 $\times$ 64 pixels. All input maps whose sizes are larger or smaller, are subsequently up/down-sampled to a size of 64 $\times$ 64 using `PyTorch`[1]'s `torch.nn.Upsample` class, in *bilinear* mode, before being passed as input to the model.

4.2.4 Data preparation and model training

Prior to using moment maps as inputs to the machine learning models, I perform data quality control cuts, to ensure that the chosen datasets are suitably homogenised and the maps contain enough information for the models to learn robust physical parameterisations. As such, in this section, I detail the various quality control cuts applied to SAMI and MaNGA datasets before and after model training and testing respectively.

[1] `http://pytorch.org/`

SAMI

Initially I use a stellar mass cut in which all galaxies with a stellar mass $M_*/M_\odot < 10^{7.5}$ are removed. Additionally, galaxies with a maximum spaxel value in the Hα velocity field of $< 10\,\mathrm{km\ s^{-1}}$ are removed.

Galaxies are removed from the training set if they have a catalogued 'FLAG' value entry greater than zero, indicating the Lick index measurements (Faber, 1973; Worthey et al., 1994) for each aperture may be unreliable, predominantly due to low signal to noise. At the spaxel level, a signal to noise (SNR) cut is employed in which spaxels with SNR < 10 are masked. Those galaxies with activated spaxel numbers < 200 are then further removed. I employ a fixed cut here due to the IFU hexabundles having the same spaxel coverage per galaxy, unlike MaNGA which has varying coverage based on the hexabundle size.

An additional cut is performed when attempting to normalise the Hα maps to have one consistent value of position angle. I use the `PaFit` python package (Krajnović et al., 2006) with 1° angular precision to fit the position angle of each Hα velocity map, and rotate each map such that each is presented to the network with the same orientation. Where fitting of the position angle fails due to highly disordered or low coverage velocity fields, these galaxies are removed from the dataset.

These cuts left a total of 1075 galaxies from the original SAMI DR2 sample for model training and testing.

MaNGA

All MaNGA maps and physical properties are extracted from MaNGA SDSS-IV datacubes using Pipe3D (as described in Sánchez et al. 2016a,b). An accumulation of quality controls and cuts are used to filter the dataset, of Hα maps, into one considered suitable for use with the machine learning model described in §4.2.3. Firstly, all galaxies with a FLAG provided in the quality control catalogue were excluded from the training set. The quality control catalogue is produced as a result of visual inspection of central spectrum fitting, the generated 2D maps, and comparison with NASA-Sloan Atlas (NSA) results.

I also perform dataset quality control cuts on a spaxel level using the `Marvin` API (Cherinka et al., 2019). Throughout this work intensity, velocity, and velocity dispersion maps are all masked using one mask per galaxy. Each galaxy's mask is a consolidation of instrumental and post processing flags, marking spaxels that Marvin advises as '*do not use for science*'. Additionally, spaxels are removed using the same SNR threshold as described in §4.2.4 for SAMI galaxies. All galaxies with a resulting

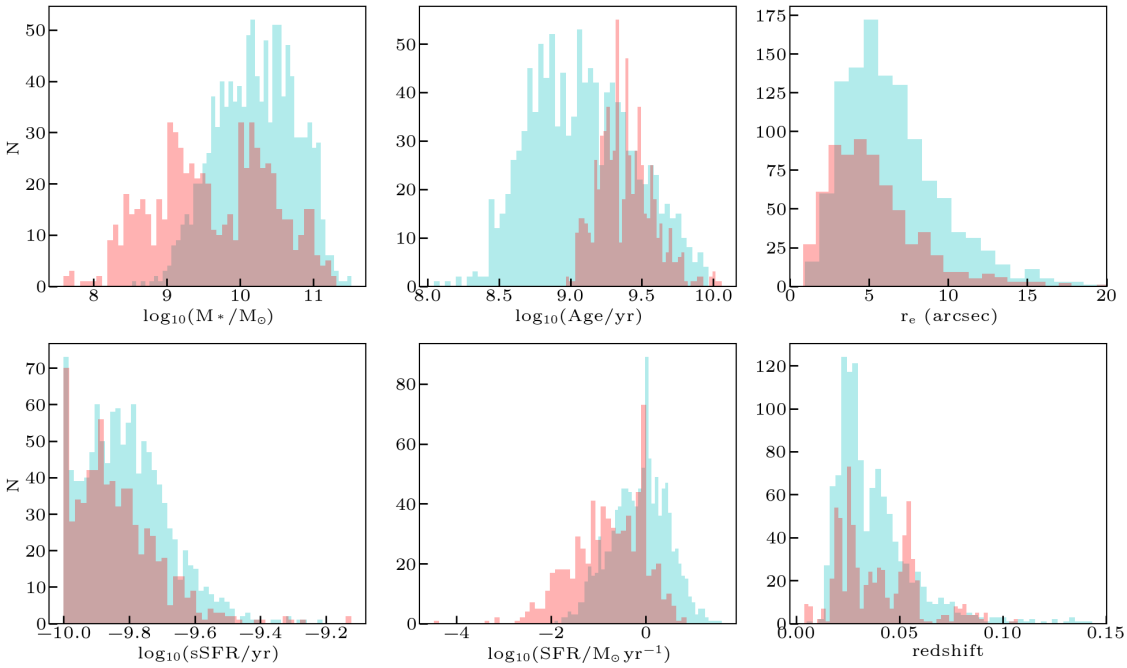


Figure 4.1. Histograms showing various galaxy properties for SAMI (red bins) and MaNGA (cyan bins) galaxies in the remaining samples after quality control cuts. From top left to bottom right: galaxy mass, age, effective radius, specific star formation, star formation, and redshift. In the bottom-left plot, the sSFR peaks at a value of 10^{-10} due to sources with catalogued sSFR's lower than the detection threshold (including negative values) being set to a control value of 10^{-10}. The total number of galaxies shown in each subplot is 1834, after pre-training and post-testing quality control cuts being applied to both datasets. The 1834 sources are comprised of 697 from the SAMI dataset and 1137 from the MaNGA dataset.

fill factor, after the aforementioned cuts, of less than 10% are removed from the dataset as well as those that do not meet the same minimum velocity requirement as for SAMI galaxies. An additional cut is performed when attempting to normalise the Hα maps to have one consistent value of position angle in the same way as described for the SAMI sample in §4.2.4.

These cuts left a total of 2645 galaxies from the original MaNGA DR16 sample for model training and testing, as described in §4.2.3. It is worth noting that these cuts inevitably induce certain biases. For example, making dataset cuts based on field of view coverage will bias the sample to nearby and bright sources which may limit the velocity profile characteristics for the network to learn from. However, this increase in sample bias is in direct opposition to the benefits of a sample which contains less controllable noise characteristics (due to e.g. sources being of low S/N). This balance between degraded model performance, due to spurious data quality, and unbiased

sample characteristics is a common problem in applying machine learning to real world tasks. In this work it is clear that we are leaning heavily more into the model performance at the expense of sample bias through the use of rather stringent cuts.

Model training & testing

Due to the different stellar mass ranges of SAMI and MaNGA samples, I trained a separate model on each dataset. This was to reduce the likelihood of over-densities in stellar mass bins from one sample biasing the parameter estimations of another sample due to physical properties being correlated with stellar mass. For example, if stellar mass were to be correlated with V_{max} (an expected relationship underpinning the TFR), then a more populous group of MaNGA galaxies in a high stellar mass bin could bias the estimated V_{max} of SAMI galaxies with stellar mass lower than the lower limit of the MaNGA sample. This could cause a systematic over-prediction of V_{max} at lower stellar mass ranges. An approach for avoiding this problem would be to train a single model with weighted sampling (Byrd & Lipton, 2018) of the training data, however thanks to the stellar mass distributions of both datasets being relatively flat, I instead avoid the problem entirely by simply training two separate models.

The models are trained in the same manner as in Chapter 3; for the purpose of clarity, I reiterate the important points of that procedure here. Each model is trained for 300 epochs (where 1 epoch comprises presentation of the entire dataset set to the model for parameter training), with input batches of size 16, during which time I observe clear plateauing of the reconstruction loss, marking convergence of the model. I use an adaptive Adam learning rate optimiser (Kingma & Ba, 2014) which begins with a value of 10^{-4} and reduces via multiplication by 0.975 every 10 epochs. The batch size, initial learning rate, and learning rate decay, were chosen arbitrarily in order to maximise training speed in the case of the batch size, and promote model convergence in the case of learning rate decay. I use the mean squared error (MSE) loss to train the models throughout (see Equation 4.3).

$$\mathcal{L} = \frac{1}{N} \sum_{i=0}^{N} \left(f\left(x_i\right) - y_i \right)^2 \tag{4.3}$$

During testing, the SAMI and MaNGA datasets are passed through the network again, in order to retrieve parameter estimations directly from the encoder subnet. Thanks to the use of *monte carlo dropout* (Gal, 2016; Gal & Ghahramani, 2016; Walmsley et al., 2019; Dawson et al., 2021a), modelling errors are found by passing each galaxy through the network 100 times, in order to build up distributions

over parameter estimations. Given that *dropout* is not used to zero the weights of the convolutional kernels, these errors are strictly modelling errors and are therefore not complete errors. One should consider these as lower limit errors over parameters; useful for pipeline flagging or generating relative errors within a self-contained dataset. The true, complete, errors will be a mixture of both modelling and observational errors.

The models and Python training scripts used for the work presented in this chapter are publicly available on GitHub[2].

Further quality control cuts

After testing, galaxies were further removed based on physical and model-behaviour motivated properties. Firstly, all galaxies passed through the trained model in testing mode, with a predicted inclination of $i < 30°$ were excluded from the final result datasets. Low inclination can create systematic line of sight problems when attempting to model the velocity profile of a galaxy. As galaxies approach inclinations of $i \to 0°$, velocity information is gradually hidden from the observer, until at 0°, when all line of sight velocity information is destroyed. Additionally, as the apparent ellipticity of a galaxy disc reduces with lower inclination, it also follows that fitting the inclination itself becomes more challenging for the model (see Fig. 4 of Dawson et al. 2021a). I therefore know that the model will have limited performance as galaxy inclinations approach 0°. During testing, a second population of galaxies with systematically higher predicted V_{max} due to low predicted values of inclination was found and removed via the chosen inclination cut described.

During testing, although I recover estimated parameters from the encoder subnet, I also pass these variable mean values on, to the decoder for image reconstruction. In doing so I can measure the mean squared error (MSE) loss between the true and reconstructed maps. In order to ensure I am not including galaxies with spurious reconstructed velocity fields, I first normalise MSE loss values for each galaxy reconstruction by V_{max}. I then proceed to remove galaxies whose normalised MSE values are $> 1\sigma$, where σ is the standard deviation of MSE values for each sample dataset. This is, of course, a rather strict cut with the resulting velocity profile parameterisations being sensitive to MSE threshold. As mentioned in §4.2.4, this imposes a focus on model performance versus sample bias.

After these post-testing cuts, I am left with two samples of size: 697 for the SAMI dataset and 1137 for the MaNGA dataset, which comprise my full results

[2] `https://github.com/SpaceMeerkat`

dataset of 1834 galaxies. A selection of properties for the combined dataset are shown in Figure 4.1.

Asymmetric drift correction

Due to non-perfect circular motions within galactic discs induced by turbulence the measured rotational velocities are lower than the circular velocities due to the gravitational potentials alone. As the TFR is a direct relation between galaxy mass and V_{max} , in order to reconstruct the TFR using my network I need to trace circular velocities as a function of the disc potential. Therefore, I need to correct the values of V_{max} by performing an *asymmetric drift correction* (Strömberg, 1946). i.e. increasing V_{max} as a function of the measured velocity dispersion. Asymmetric drift is defined as the difference between the mean stellar tangential velocities and the velocity as a result of the stellar density and velocity dispersion gradient (Binney & Tremaine, 2008).

I calculate the correction factor of V_{max} using an approximation of the Evans model (Binney & Tremaine, 2008; Weijmans et al., 2008) assuming a thin disc isotropic rotator (Neistein et al., 1999):

$$V^2_{max,\,new} = V^2_{max} + V^2_{disp}\left(2\left(\frac{r_{max}}{r_{scale}}\right) - 1\right), \tag{4.4}$$

where V_{max} and r_{scale} are the mean predicted values of the maximum rotational velocity and intensity profile scale length, and V_{disp} (calculated from maps provided by the respective surveys) is the average measured velocity dispersion of each galaxy from $2\,r_e$ out to the maximum radius at which Hα is detected, r_{max}. In the case where $r_{max} < 2r_e$, V_{disp} is instead measured from $r_{max}/2$ out to r_{max}. Corrections are applied to both SAMI and MaNGA datasets prior to constructing the TFR shown in §4.3, with typical corrections peaking at ~ 50km s^{-1}.

It is worth noting that the instrumental spectral line-spread function (LSF) values of Hα for MaNGA targets are systematically lower than in upcoming SDSS data releases (2021, private communication with Dr Federico Lelli). I therefore perform a correction to the average velocity dispersion measurements of MaNGA samples by increasing their LSF values by a factor of 1.045 (assuming LSF values are $\sim 95.5\%$ of those in updated data pipelines at Hα wavelength –see Fig. 13 in Law et al. 2021) prior to calculating the asymmetric drift corrections.

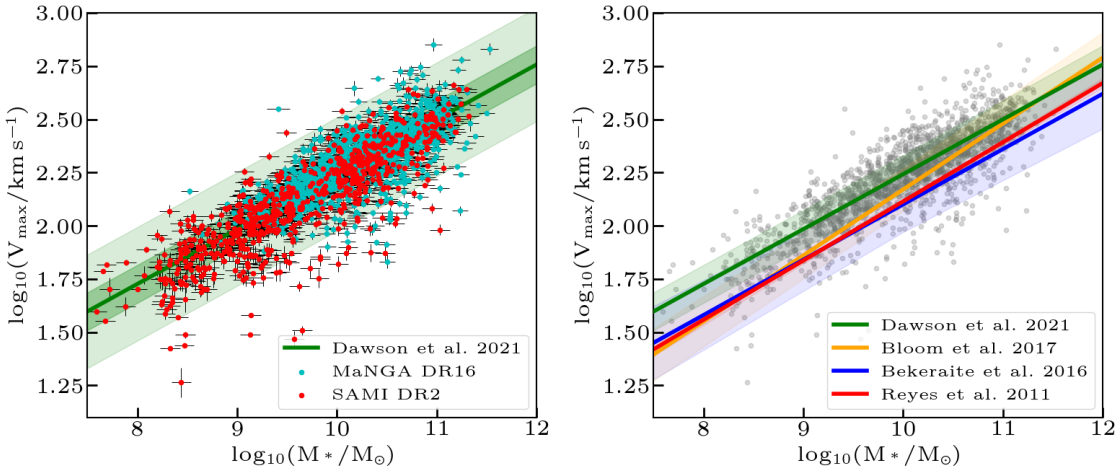


Figure 4.2. (Left) the stellar mass Tully-Fisher relation using all 1834 galaxies of my combined SAMI DR2 and MaNGA DR16 dataset. Velocities (V_{max} on the y-axes) are the estimates provided by the neural network model during testing. Masses are taken from their respective survey data-analysis-pipeline catalogues and velocities are estimated using the machine learning models described in §4.2.3. SAMI and MaNGA datapoints are shown with red and cyan markers respectively, complete with modelling errors thanks to the use of *monte carlo dropout*. Each galaxy is sampled 100 times to generate modelling errors. The green solid line shows the best fit, with green shaded regions showing the 1σ and 3σ confidence levels. (Right) the TFR with comparison fits from the literature and 1σ confidence levels. Data points matching those found in the left subplot are shown with grey markers. My best-fit TFR is systematically higher than the literature relation at low masses. This is because in this work I performed an asymmetric drift correction, which causes greater shifts in V_{max} at lower masses.

4.3 Results and Discussion

In this section I present the results of fitting the Tully-Fisher relation using V_{max} predictions from the models outlined in §4.2.3, and integrated stellar mass estimates from the survey catalogues. Parameter estimations for the entire dataset of 1834 galaxies sampling each galaxy 100 times using *monte carlo dropout*, took < 3 minutes on a single NVIDIA GTX TITAN Xp.

4.3.1 The reverse Hα TFR for combined SAMI and MaNGA samples

I create the 'reverse' Hα TFR using all 1834 galaxies from combined SAMI and MaNGA datasets described in §4.2.4 and §4.2.4. The reverse TFR consists of

fitting the predicted V_{max} based on known galaxy mass. In order to fit the predicted velocities from known masses, I use the python based least-squares fitting algorithm `LtsFit` (presented in §3.2 of Cappellari et al. 2013). The approach uses Least Trimmed Squares (LTS) robust technique (Rousseeuw & Driessen, 2006), which allows for intrinsic scatter and errors in all coordinates while converging to the correct solution even in the presence of significant outliers, which are excluded in the calculation of scatter around fits.

Performing a linear fit of the slope and intercept for a fixed pivot in the form $y = m(x - p) + c$, where m, p, and c are the gradient, pivot, and y-intercept respectively, I fit the reverse TFR:

$$V_{max}/\mathrm{km\,s^{-1}} = 0.258 \pm 0.003 \times (\log(M_*/M_\odot) - 10.09) + 2.265 \pm 0.002, \qquad (4.5)$$

using a pivot value of $p = 10.09$ –the median $\log_{10}(M_\odot)$ of the combined sample.

The results of fitting the reverse TFR to my combined dataset are shown in Figure 4.2. The observed rms scatter in V_{max} is 0.089 dex which I show in Figure 4.2 with shaded areas representing the 1σ and 3σ confidence levels on the fit.

In Figure 4.2 I display comparison fits for the reverse TFR from the wider literature. Namely I include fits from: Bloom et al. (2017b) who performed fits to 729 SAMI observations of galaxies drawn from the GAMA DR2 sample, Bekeraité et al. (2016), whose reverse TFR fit was calculated in Bloom et al. (2017b) and converted from a fit to the stellar kinematics of 199 Calar Alto Legacy Integral Field Area Survey (CALIFA) galaxies, and finally Reyes et al. (2011) who performed TFR fits to 189 SDSS DR7 disc galaxies, observed using long-slit spectroscopy. I summarise the derived fit parameters for my sample and for those found in the aforementioned publications in Table 4.1.

In Figure 4.2, I also show a comparison of my reverse Hα TFR fit to those found in the wider literature. I show the 1σ confidence levels for all fits, for comparison. It can be seen that at all times, my fit overlaps with another found in the literature within the examined mass range. My fit clearly has the highest y-intercept of all the fits shown, and a clean systematic offset across the entire mass range. This is due to my data being corrected for asymmetric drift (see §4.4), which when removed brings the fit back into agreement with the ancillary fits presented from the wider literature. None of the comparison fits from the wider literature, shown, perform a correction for asymmetric drift in this way. However, it is worth noting that my asymmetric drift corrections are approximate in nature and more meticulous calculations of the

Author/Paper	Sample	Sample size	Slope $\log(\mathrm{km\ s^{-1}})/\log(\mathrm{M_*/M_\odot})$	Intercept $\log(\mathrm{km\ s^{-1}})$
Dawson et al. 2021	SAMI DR2 & MaNGA DR16	1834	0.258 ± 0.003	2.265 ± 0.002
Bloom et al. (2017b)	SAMI DR1	729	0.310 ± 0.009	2.198 ± 0.100
Bekeraité et al. (2016)	CALIFA	199	0.260 ± 0.017	2.123 ± 0.130
Reyes et al. (2011)	SDSS DR7	189	0.278 ± 0.010	2.164 ± 0.004

Table 4.1 This table shows information regarding the fit parameters for a linear regression fit to the TFR for my combined sample and for those shown in the wider literature for comparison. All fits have been normalised to share the same pivot $p = 10.09$ as used in my TFR fit. Columns show (from left to right): The author/paper from which fit parameters were sourced, the sample from which galaxies were drawn for those corresponding studies, the size of those samples, the gradient of the reverse TFR fit for those samples, and the y-intercept for of the reverse TFR fit for those samples.

correction value are recommended as a way to improve the validity of the derived fit parameters.

4.3.2 The forward Hα TFR for combined SAMI and MaNGA samples

I create the 'forward' Hα TFR using all 1834 galaxies from combined SAMI and MaNGA datasets described in §4.2.4 and §4.2.4. This is a useful exercise as in practice, astronomers may have access to either galaxy mass or $\mathrm{V_{max}}$ and must infer the other. As such it is worth performing both forward and reverse fits. The forward TFR consists of fitting the predicted baryonic mass based on known $\mathrm{V_{max}}$ values for my combined sample. To perform the fit, I again use the python based least-squares fitting algorithm `LtsFit` (presented in §3.2 of Cappellari et al. 2013).

Performing a linear fit of the form $y = m(x - p) + c$, where m, p, and c are the gradient, pivot, and y-intercept respectively, I fit the reverse TFR for my sample:

$$\log(\mathrm{M_*/M_\odot}) = 3.013 \pm 0.037 \times (\log(\mathrm{V_{max}/km\,s^{-1}}) - 2.256) + 10.054 \pm 0.008, \quad (4.6)$$

using a pivot value of $p = 2.256$ –the median $\log(\mathrm{V_{max}})$ of the combined sample.

The results of fitting the forward TFR to my combined dataset are shown in Figure 4.3. The observed rms scatter is $0.31\,\mathrm{M_\odot}$ which I show with shaded areas representing the 1σ and 3σ confidence levels on the fit.

In Figure 4.3, I show comparison fits from the wider literature against my derived fit. As reverse TFR fits were exclusively calculated in those taken from the literature, I simply invert the parameters for the reverse TFR fits for display in Figure 4.3. As with the reverse TFR (see §4.2) I see good agreement between my fit and those

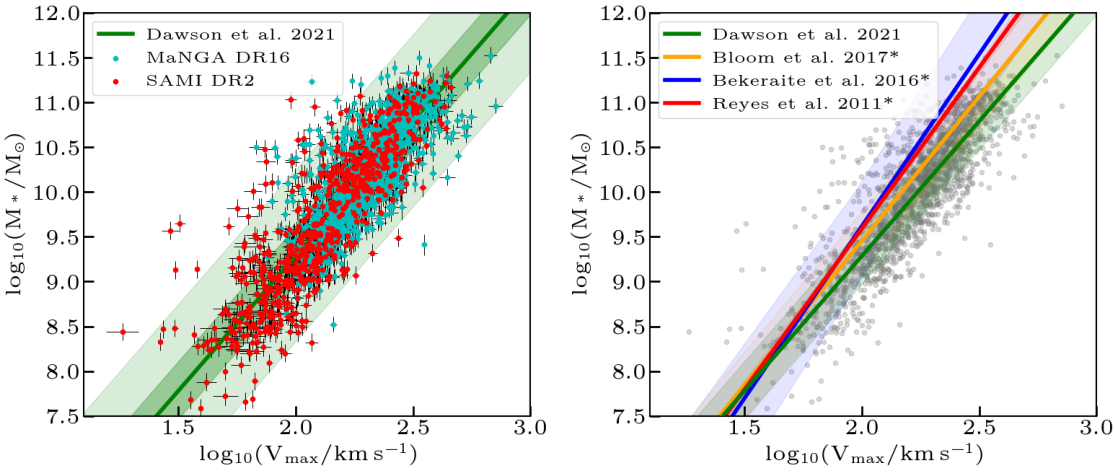


Figure 4.3. (Left) the stellar mass Tully-Fisher relation using all 1834 galaxies of my combined SAMI DR2 and MaNGA DR16 dataset. Values of V_{max} are estimated using the machine learning models described in §4.2.3 and stellar masses are taken from their respective survey data-analysis-pipeline catalogues. SAMI and MaNGA datapoints are shown with red and cyan markers respectively, complete with modelling errors thanks to the use of *monte carlo dropout*. Each galaxy is sampled 100 times to generate modelling errors. The green solid line shows the best fit, with green shaded regions showing the 1σ and 3σ confidence levels. (Right) the TFR with comparison fits from the literature and 1σ confidence levels. These fits are annotated with an asterisk to show that I have simply inverted the function parameters and uncertainties in Table 4.1 to display them in this figure. My best-fit TFR is lower than the literature relations displayed. This is because in this work I performed an asymmetric drift correction, which causes systematic shifts in V_{max} .

presented in the literature when working within a 1σ confidence region. I summarise the derived fit parameters for my sample and for those found in the aforementioned publications in Table 4.2.

Author/Paper	Sample	Sample size	Slope $\log(M_*/M_\odot)/\log(\mathrm{km\ s^{-1}})$	Intercept $\log(M_*/M_\odot)$
Dawson et al. (2021)	SAMI DR2 & MaNGA DR16	1834	3.013 ± 0.037	10.054 ± 0.008
Dawson et al. (2021)*	SAMI DR2 & MaNGA DR16	1834	$3.876^{+0.046}_{-0.045}$	$10.054^{+0.012}_{-0.123}$
Bloom et al. (2017b)*	SAMI DR1	729	$3.226^{+0.099}_{-0.093}$	$10.278^{+0.518}_{-0.549}$
Bekeraité et al. (2016)*	CALIFA	199	$3.846^{+0.269}_{-0.236}$	$10.602^{+0.516}_{-0.588}$
Reyes et al. (2011)*	SDSS DR7	189	$3.597^{+0.134}_{-0.284}$	$10.421^{+0.284}_{-0.305}$

Table 4.2 This table shows information regarding the fit parameters for a linear regression fit to the TFR for my combined sample and for those shown in the wider literature for comparison. Columns show (from left to right): The author/paper from which fit parameters were sourced, the sample from which galaxies were drawn for those corresponding studies, the size of those samples, the gradient of the reverse TFR fit for those samples, and the y-intercept for of the reverse TFR fit for those samples. All fits have been normalised to share the same pivot $p = 2.256$ as used in my fit. Author/paper information highlighted with an asterisk, indicates that the fit parameters and errors are calculated by inverting the 'reverse' TFR fit parameters for comparison. This is due to the fact that fits to the forward TFR are not presented in the literature examples chosen for comparison.

4.3.3 r-band magnitude TFR

Figure 4.4 shows the r-band magnitude TFR with V_{max} values taken from my neural network model. I fit the TFR using the same methods as in §4.3.1 and 4.3.2, giving:

$$M_R = -5.290 \pm 0.100 \times (\log(V_{max}/\mathrm{km\,s^{-1}}) - 2.256) + 19.708 \pm 0.022, \quad (4.7)$$

where M_R is the predicted r-band absolute magnitude. The rms scatter is found to be 0.92 mag. The r-band magnitude TFR serves the same purpose as the stellar mass TFR but is included in this work due to the convenience of r-band magnitude being readily available in the data reduction pipeline database. Should one wish to infer galaxy distances, rather than masses, based on measurements of V_{max}, then using this r-band TFR would be the ideal starting point.

4.3.4 Correlations along the stellar mass TFR

Thanks to the samples used throughout this chapter being from large surveys including data reduction and data analysis pipelines, we are able to explore various known (and unknown) corellations between parameters derived using my ML models and those from the pipelines themselves. This represents a good opportunity to further validate that the models are performing as we expect and to also search for new relationships which may impact future observing stratergies.

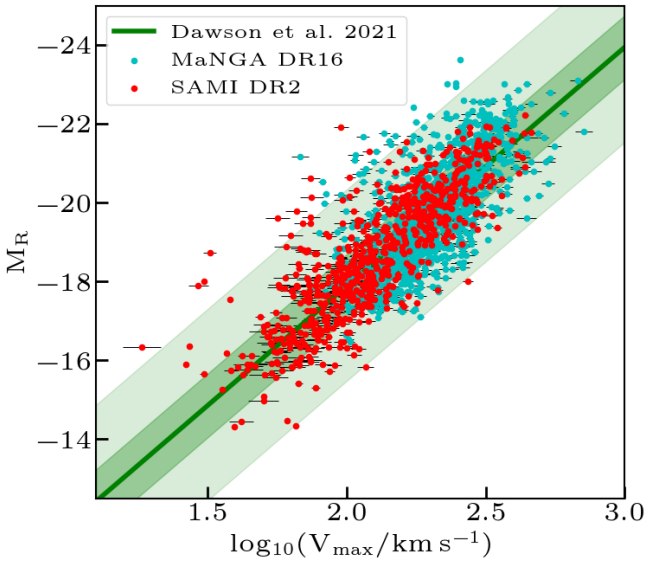


Figure 4.4. The r-band magnitude Tully-Fisher relation using all 1834 galaxies of my combined SAMI DR2 and MaNGA DR16 dataset. Values of V_{max} are estimated using the machine learning models described in §4.2.3 and absolute r-band magnitudes are taken from their respective survey data-analysis-pipeline catalogues. SAMI and MaNGA datapoints are shown with red and cyan markers respectively, complete with modelling errors thanks to the use of *monte carlo dropout*. Each galaxy is sampled 100 times to generate modelling errors. The green solid line shows the best fit, with green shaded regions showing the 1σ and 3σ confidence levels.

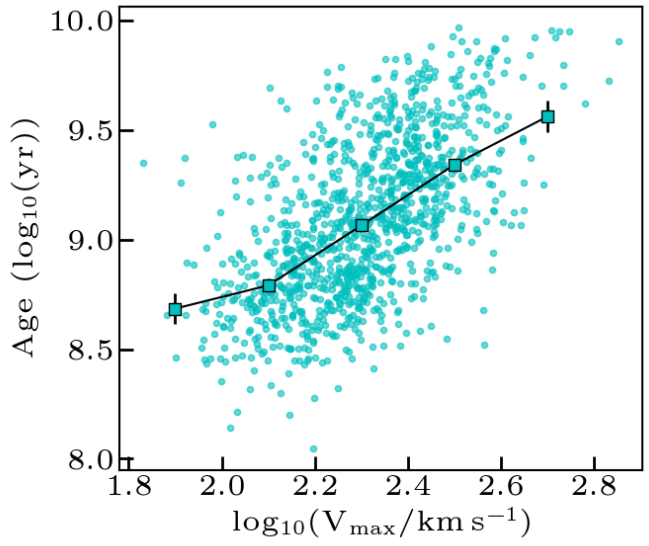


Figure 4.5. The distribution of galaxy ages as a function of V_{max} for the MaNGA sample tested in this work. I see a positive trend, in which higher mass galaxies tend to have older ages. I omit the SAMI sample from this plot as ages was fit in a manner which does not allow for suitable comparison between the two datasets. A running median and associated errors are shown as cyan squares connected by a black line.

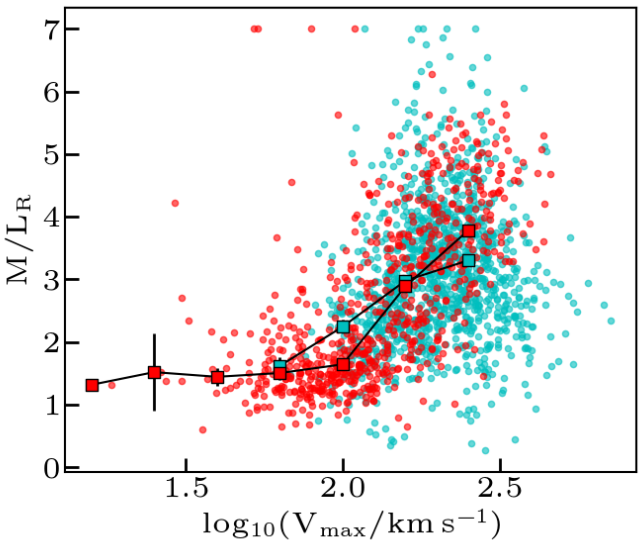


Figure 4.6. Mass to light ratios against V_{max} for all 1834 galaxies in the MaNGA (cyan) and SAMI (red) datasets after quality control cuts. Running medians for both datasets and associated errors are shown with square markers connected by black lines. I see that there is a positive trend in the mass to light ratios with increasing V_{max}. It is worth noting that the choice of using a running median is arbitrary and can be improved upon prior to publication by using more suitably chosen adaptive fitting routines which can account for sparsity in individual bins.

Figure 4.5 shows the distribution of galaxy ages (retrieved from the MaNGA database and calculated using SED fitting in the MaNGA data-reduction pipeline) against V_{max}, for the MaNGA sample. We see a clear positive trend between the two properties with spearman rank correlation coefficient and p value of $r_s = 0.605, p < 0.001$. In this case I have omitted the SAMI galaxies due to varying fitting models being used to recover ages between the two samples making comparisons unsuitable. It therefore follows that a trend exists between position along the TFR and galaxy age. We see that the galaxies with older stellar populations lie at the higher stellar mass and V_{max} range of the TFR. This makes sense given that we expect massive galaxies to be older, and therefore to be dominated by older stars (Gallazzi et al., 2005). This is consistent with hierarchical assembly due to '*assembly bias*', the theory that larger dark matter halos in the early Universe resulted in higher mass galaxies which then evolved through cosmic time hierarchically but with much older stellar populations (Neistein et al., 2006). Therefore, the trend of increased stellar population age with higher V_{max} agrees with prior studies.

The relationship between V_{max} and galaxy age, complements a visible trend between position along the TFR and the mass-to-light ratio (M/L) shown in Figure

4.6. Here we see that both datasets exhibit positive correlations between V_{max} and M_*/L with spearman rank correlation coefficients and p values of $r_s = 0.168, p < 0.001$ and $r_s = 0.693, p < 0.001$ for MaNGA and SAMI datasets respectively. It again follows that this relationship is present in the stellar mass TFR. Older, more massive galaxies are dominated by low mass populations of stars (Gallazzi et al., 2005), which increase the overall M/L of a galaxy; whereas younger, lower mass galaxies have luminosities dominated by brighter more massive stars, leading to lower global M/L's.

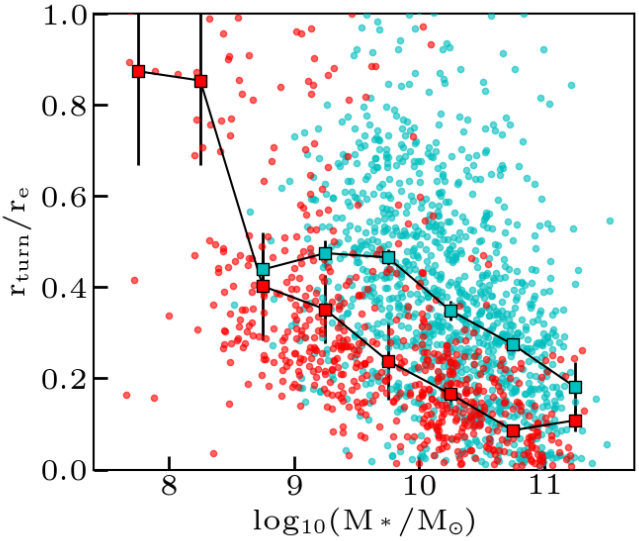


Figure 4.7. The velocity turn over radius r_{turn}, normalised by r_e, as a function of galaxy mass. I see a negative trend for both MaNGA (cyan) and SAMI (red) datasets. Running medians for both datasets and associated errors are shown with square markers connected by black lines. reviseAgain the choice of using a running median is arbitrary and can be improved upon in the future by using more suitably chosen adaptive fitting routines which can account for sparsity in individual bins A visible offset between the two populations is visible and requires further investigation, as explained in the main text.

In Figure 4.7, we can also see evidence of a negative trend between r_{turn} and V_{max} after normalising r_{turn} by r_e. This is an expected trend to arise during testing as higher mass galaxies have been shown to have their mass more centrally concentrated than lower mass galaxies (Caon et al., 1993; Yoon, 2017), which leads to a smaller turn over radius. I measure spearman rank correlation coefficients and p values of $r_s = -0.37, p < 0.001$ and $r_s = -0.69, p < 0.001$ for MaNGA and SAMI datasets respectively. If a relationship between r_{turn} and V_{max} can be modelled to a reasonable degree of accuracy, this could have implications for future target selection for kinematic analysis, by circumventing the need to explicitly model r_{turn} if one can

measure V_{max} alone, or having a priori knowledge of how far out to probe galaxies (to explore beyond r_{turn}) given known galaxy masses. It is clear in Figure 4.7 that the distributions of r_{turn}/r_e, of MaNGA and SAMI datasets, are offset from one another. Why this behaviour arises is currently unknown and must be fully addressed before such relationships can be modelled robustly.

It is also worth noting a few other factors which may effect the relationship we see between r_{turn}/r_e and V_{max} in Figure 4.7. Changes in the expected velocity profile due to "beam smearing", and the optically thick nature of Hα emission will play some role in the observed relationship. Additionally the known covariance between r_{turn} and V_{max} must influence this relationship in an as-of-yet unknown manner.

An investigation of the aforementioned behaviours and possible influences in the relationship seen in Figure 4.7 will be the next logical step in validating a strong relationship between r_{turn}/r_e and V_{max}. Unfortunately this could not be completed in the timescales required for this thesis. As such I leave work quantifying the true strength of this relationship as a promising avenue for future work.

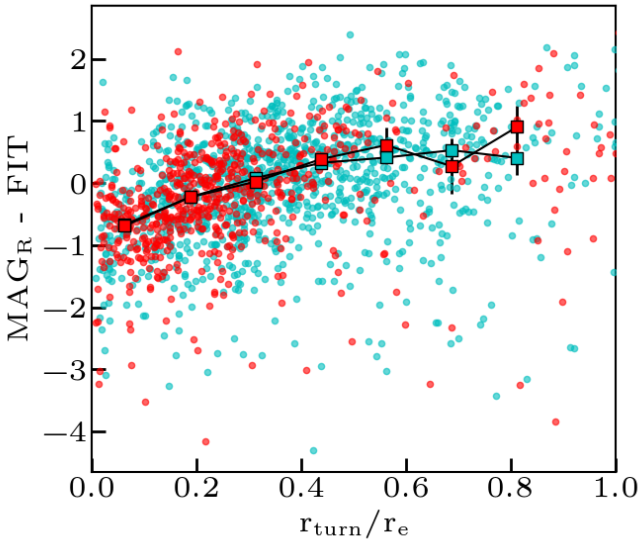


Figure 4.8. The residual r-band magnitudes of each galaxy plotted against the predicted velocity turn over radius r_{turn}, normalised by r_e. MaNGA galaxies are shown with cyan markers, while SAMI galaxies are shown with red markers. Running medians for both datasets and associated errors are shown with square markers connected by black lines. I observe a positive trend, with galaxies with lower values of r_{turn} having systematically more luminous than expected by my predictions of V_{max}. This is discussed in greater detail in §4.3.5.

In Figure 4.8, we see a clear trend between predicted r_{turn}/r_e and the residuals from my best fit TFR in magnitude space shown in Figure 4.4. Both SAMI and

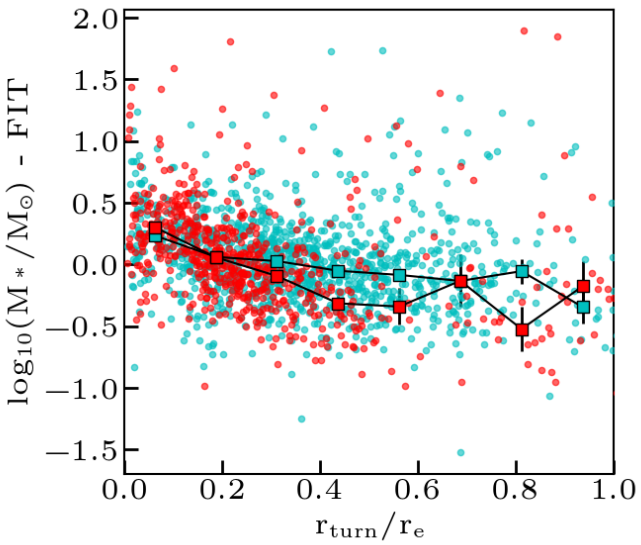


Figure 4.9. The residual masses of each galaxy plotted against the predicted velocity turn over radius r_{turn}, normalised by r_e. MaNGA galaxies are shown with cyan markers, while SAMI galaxies are shown with red markers. Running medians for both datasets and associated errors are shown with square markers connected by black lines. I observe a negative trend, with galaxies with lower values of r_{turn} having systematically higher mass than expected by my predictions of V_{max}. This is discussed in greater detail in §4.3.5.

MaNGA have spearman rank correlation coefficients and p values of $r_s = 0.39, p < 0.001$ and $r_s = 0.36, p < 0.001$ respectively. The negative offset of magnitudes at lower r_{turn}/r_e is unexpected and suggests that galaxies with low measured r_{turn}/r_e are more massive than expected given their rotation velocity.

We can make attempts to explain this behaviour by looking at the residuals of predicted mass as shown in Figure 4.9. Here we observe a negative trend, corresponding to an under-prediction of mass, and therefore of V_{max} also, for lower values of r_{turn}/r_e.

4.3.5 UNCERTAINTIES ON THE TFR

As I am determining the TFRs with a new technique, it is important that we search for any underlying uncertainties that could affect the results. I did this by checking for the presence of residual correlations between the position of galaxies on the TFR and their properties. In Figure 4.10, we see that the galaxies which have negative residual magnitude values (where $r_{turn}/r_e \lesssim 0.2$), can exhibit much lower values of r_{max}. This is particularly prevalent for the SAMI sample. This may suggest that the trends seen in Figures 4.8 and 4.9 arise from the model's inability

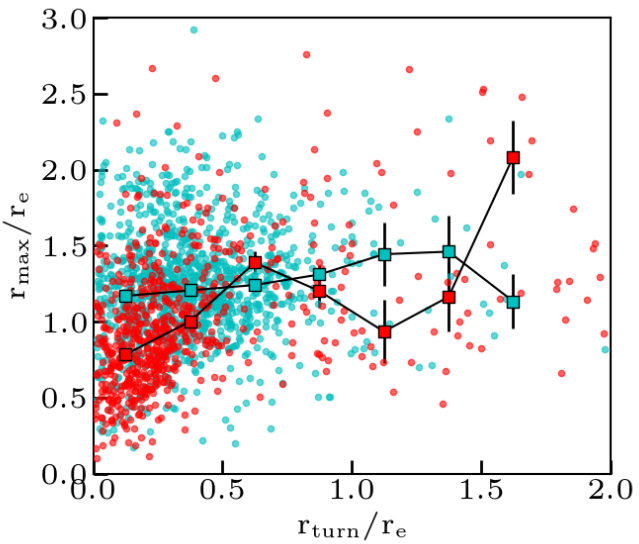


Figure 4.10. The predicted velocity turn over radius r_{turn} for each galaxy, normalised by r_e, against Hα extent normalised by r_e. MaNGA galaxies are shown with cyan markers, while SAMI galaxies are shown with red markers. Running medians for both datasets and associated errors are shown with square markers connected by black lines. I observe a sharp drop in the ratio of Hα extent for low values of r_{turn} for the SAMI sample. I discuss this issue in greater detail in §4.3.5 along with the implications for r_{turn}, and V_{max} predictions.

to appropriately fit either r_{turn} or V_{max} in cases where the galaxies are very compact. Given that the two parameters co-vary, the reduced ability to fit could be due to underpredictions of V_{max} which directly lead to lower estimations of r_{turn} in order to offset the MSE reconstruction loss, which dominates at higher velocity regions. Alternatively, an overprediction of r_{turn} leads to the opposite effect in which V_{max} is lowered in order to reduce the overall MSE reconstruction loss. Which of these two possible behaviours occurs is currently unknown and requires further investigation, however it is clear that the ratio of Hα extent and r_{turn} plays a significant role in the residual trends seen in Figures 4.8 and 4.9. Given that galaxies with a range of r_{turn} are present at all stellar masses, and my best fit TFRs are fully consistent with those in the literature, we expect this uncertainty to be increasing the scatter seen in my relations, rather than biasing them significantly. Further work will be required in order to fully understand the extent of this behaviour.

4.4 CONCLUSIONS

I have trained two self-supervised physics-aware neural networks to recover parameter estimations from 2-dimensional spatially resolved kinematic maps from IFU surveys. Using a combined dataset from observations using SAMI and MaNGA IFUs, I modelled the Hα stellar mass Tully-Fisher relation using the largest sample to date. My forward and reverse fits agree well with those found in the literature, with the exception of a notably shallower slope in my fits. I attribute this discrepancy to galaxies in my sample having higher values of V_{max} as I have performed asymmetric drift corrections to my dataset throughout. Both my forward and reverse TFR fits, overlap (within 1σ) with at least one other fit found in the literature within the mass range studied in this chapter.

I also observe some possible trends with position along the TFR and physical properties, taken from the respective survey data-analysis pipeline catalogues. In particular, I see trends in galaxy age and mass-to-light ratios with V_{max}, which agree with my expectations from the literature.

Although systematic uncertainties are currently present which I was not able to resolve on the timescales required for this thesis, a possible trend has been observed between position along the TFR and r_{turn}/r_e. Measuring the strength of this trend is beyond the scope of this work, but promises interesting implications for the future of kinematic studies of galaxies. Should one already know the masses of galaxies in a sample, then it may be possible to know, prior to observing, how far out to observe galaxies in order to target regions beyond r_{turn}. Alternatively, if one is able to build a robust relationship between V_{max} and r_{turn}/r_e, then it may be possible to use new survey target selection approaches based purely on known values of V_{max} . This could boost observing efficiency and help build more representative survey samples. I propose the analysis of this relationship as a promising avenue of future work.

I have investigated and discussed the trends seen when measuring residuals between TFR fits and combined SAMI and MaNGA datapoints. Of particular note is the increased discrepancy between expected and measured r-band magnitude. I see an increase in brightness for galaxies for values of $r_{turn}/r_e \lesssim 0.2$. I posit that this arises in cases where objects are very compact. Further investigation is required so solidify this hypothesis and differentiate the behaviour from other possible reasons, such as insufficient noise removal which could could exacerbate fitting issues if r_{turn}/r_{max} approaches 0.

From a machine learning perspective, I have demonstrated that it is possible to reuse model architectures to solve problems and uncover physical properties of galaxies, observed with IFU instruments, in a time-efficient manner. Given that the models used were designed for use with millimetre/radio observations, the results outlined in this work demonstrate the generalised nature of the models. This gives us greater confidence in the use of such models moving into the future in order to prepare for next generation IFU surveys.

Chapter 5
Thesis Conclusions

"Just use an LSTM..."

Hardie Pienaar, *The answer to all of life's problems*

5.1 Conclusions

In this thesis I studied the use of modern machine learning approaches in tackling challenges for kinematically characterising galaxies beyond the Milky Way. These pieces of work focused on preparing for next generation millimetre and radio interferometers, such as ALMA and the SKA, by specifically targeting applications to large-scale emission line surveys of galaxies. A variety of emission lines were studied throughout the thesis, including cold dense molecular gas (CO), warm neutral hydrogen (HI), and hotter ionised hydrogen (Hα). Each are important for different scientific purposes; but the applicability of the approaches studied in this thesis to different emission lines, using different telescopes, indicates that this work is based on a more generalised approach to tackling big data era challenges in astronomy than previously attempted.

As a quick-look reference, Figure 5.1 shows a rough guide to the path taken in this thesis from the broad topic of extragalactic astronomy, through to the machine learning approaches utilised to solve specific problems.

In order to understand the compatibility of machine learning with characterising kinematic properties of galaxies, I first used hydrodynamical simulations to

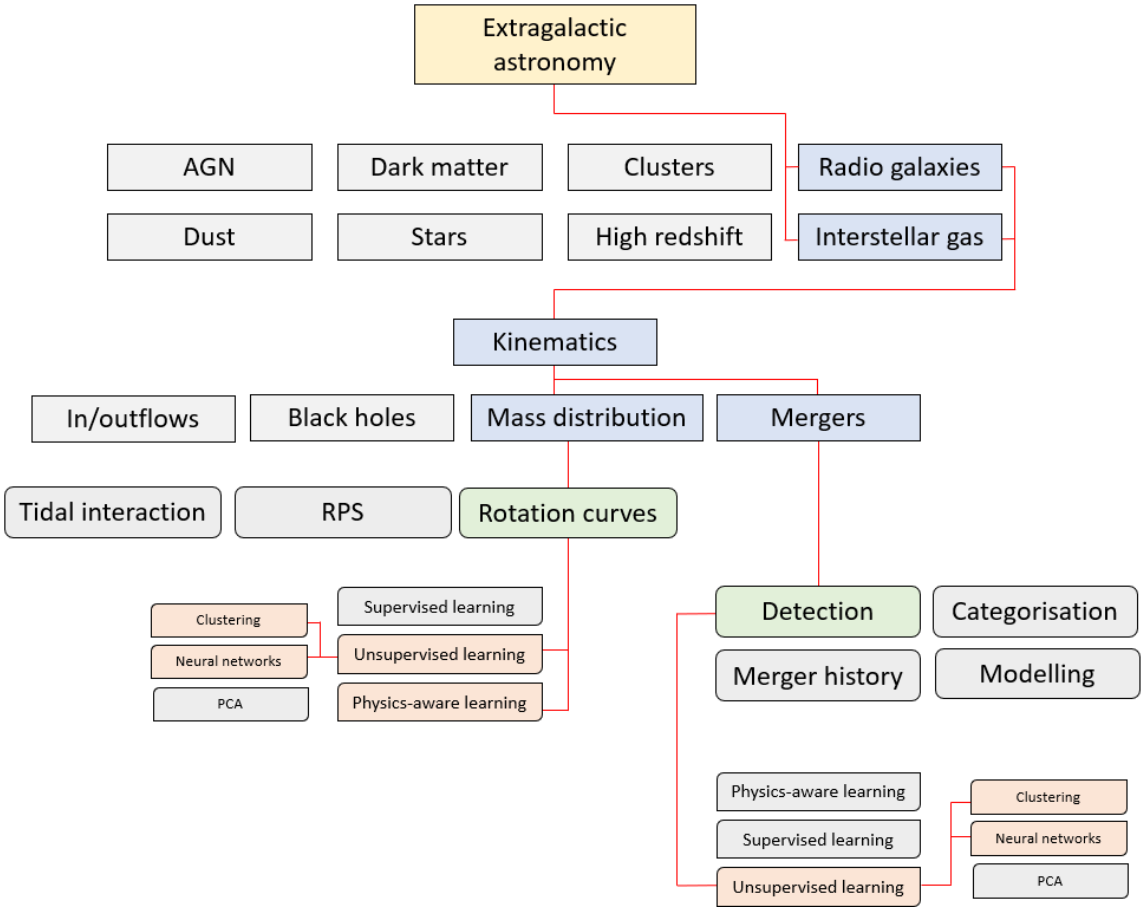


Figure 5.1. A generalised view of the research-path described in this thesis. Topics move from the overarching extragalactic astronomy, through to the specific machine learning approaches used to tackle problems.

evaluate machine learning models in an empirical fashion. Using unsupervised learning I embedded kinematic features from simulated moment one maps of galaxies onto a 3-dimensional latent manifold. This structure was then used to discriminate between disturbed and regularly rotating cold gas structures. By using moment one maps I was working at a very high level in the data reduction pipeline. At this stage the data is lower dimensional which therefore made it easier to develop and evaluate novel data-handling techniques. Using unsupervised learning while being able to empirically calculate kinematic features of the simulated data allowed for even greater power in in assessing the performance of machine learning models. The simulated nature of the training data also granted complete flexibility in augmenting the data presented to the model for training. With nearly infinite augmentations available, this was an exercise in evaluating machine learning models with spectacular parameter coverage and an opportunity to discover the limitations of using vast quantities of simulated data to produce models with high generalisation power.

The model had one clear goal from the outset: to embed kinematic features from moment one maps in a minimal time-frame, while maintaining high generalisation power. The model was trained and tested on millions of simulated cold molecular gas disc observations. The moment maps were embedded onto a 3-dimensional spheroidal-like manifold where a correlation between level of ordered rotation and position was clearly visible. Using a maximum-likelihood boundary line, disordered and ordered rotators were separated into two distinct categories based on an empirical measure of (dis)orderly rotation κ. The performance of the model was then dictated by its performance as both an embedding network and binary classifier.

During testing on simulated data, the model showed a mean accuracy of 85% when attempting to categorise the circularity κ, with minimal degradation in this performance when injecting noise. The model was also tested on real HI observations using the VLA, from the Local Volume HI Survey (LVHIS, Koribalski et al. 2018). The outcome of testing on HI observations is important because it shows the generalisation power of using a model such as this, and it shows that using machine learning to study the kinematics of HI is possible and therefore applicable to SKA science. The accuracy of testing on HI data was found to be 97%. This is obviously a heuristic accuracy as we have no handle on the true circularity of galaxies from direct observation alone. Additionally, the model was tested on CO observations from surveys using ALMA. Given that CO is generally found in much more centrally concentrated regions of galactic discs than HI , and the ALMA observations had greater spatial and spectroscopic resolution when compared with the LVHIS data, this presented the opportunity to test the generalisation power of the model. With the ALMA data, again the model performed well, with a mean accuracy of 90%.

Despite the positive results from this work, some caveats could not be ignored and dictated the progression of this line of research away from more traditional unsupervised learning approaches, and into the realm of self-supervised, physics-aware machine learning.

I then went on to build models which were capable of learning physical parameters without the need for a throw-away training set. This presented two major positive steps in developing machine learning models for next generation instruments. The ability to learn physically meaningful results directly from input data removes the need for secondary steps such as the binary classifier presented in the first project. Secondly, removing the need for a throw-away training set circumvents transfer learning issues when using a simulated backend, increases the fraction of 'testable' data (by removing the need for train-test split approaches), and eliminates the need for

separate training data storage.

In order to accomplish these feats, a new type of machine learning architecture was needed. Building upon similar approaches used with simulated optical images of galaxies (Aragon-Calvo & Carvajal, 2020), I created a novel convolutional autoencoder architecture in which the decoder had been replaced with native `PyTorch` (Paszke et al., 2017) functions. These functions forced the network to learn physically motivated parameters in the embedding layer prior to image reconstruction. Again inputs to the network were moment zero and moment one maps, but this time, all training data consisted of observational survey data. Simulated data was used; however, only to check for spurious model behaviour. During both training and testing, the inputs were moment one maps from the mm-Wave Interferometric Survey of Dark Object Masses (WISDOM) survey and The HI Nearby Galaxy Survey (THINGS, Walter et al. 2008).

I trained networks to recover the intensity and rotational velocity curves from resolved velocity maps for both CO and HI data. The modelling of these galaxies assumed simple decaying exponential intensity profiles and arctan velocity profiles throughout. The recovered velocity profiles matched well with those found in the literature, using modelling tools such as `KinMS` (Davis et al., 2013; Davis et al., 2020) and `BBarolo` (Di Teodoro & Fraternali, 2015). The significance of these comparison tests lie in a comparison of the testing speed itself. The trained self-supervised model is able to recover rotation curves of galaxies in a fraction of a second, limited only by the quantity of available GPU memory for batch loading. In comparison, current state of the art modelling tools (e.g. `BBarolo`) take at least several minutes per target to recover rotation curves. Therefore, this shows a major improvement in physical parameterisation speed.

In addition to learning physical parameters, I also employed the use of *monte carlo dropout* to transform a network, which would otherwise calculate mean field approximations of parameters, into one capable of producing modelling errors. Using synthesised moment one maps I show that the network learns physical parameterisations with associated modelling errors without evidence of learned parameters covarying. This is with the exception of the velocity turn over radius and the asymptotic velocity whose covarying nature is fully expected. In order to show the ability of the model in fitting rotation curves to a variety of observations I present 3 galaxies and their fit results. NGC 2403 and NGC 3198 are both HI observations from THINGS (Walter et al., 2008; de Blok et al., 2008). In both cases the fits agree with those using other more analytically motivated methods, even for NGC 3198 which exhibits a slight rise in velocity at high radii and has a warped HI disc. The CO observation

of NGC 1387 from the WISDOM survey (Zabel et al. 2020, Boyce et al. in prep), is included in order to show that the network is capable of fitting galaxies whose rotation curves asymptote at very small radii as well as demonstrating the ability of the network to handle observations of different emission lines, with different instruments.

The effects of instrument resolution on direct learnable physical properties, from moment maps, was also discussed in this work. No attempts were made to mitigate the effects of "beam smearing" as it was beyond the scope of the project, yet it is of vital importance to do so before these methods can be used on large scale interferometric surveys in the future. Other artificial errors in predicted parameters due to beam size and inclination were also discussed and resulted in advised flagging and/or cuts, at the post testing stages of the project.

Finally, in Chapter 4, I go on to use self-supervised physics-aware algorithms to retrieve V_{max} for 1834 galaxies observed in Hα line emission. Using a combination of SAMI survey DR2 and MaNGA DR16 IFU data, V_{max} is calculated and used to fit the stellar mass Tully-Fisher relation (TFR). This represents the largest Hα TFR study at the time of writing, and delivers interesting results. Firstly, the ability to use the same models as those used on mm-radio interferometric data products highlights their generalisation power, one of the key goals of developing such models as mentioned previously. With upcoming instruments like Hector-I, which will double the number of galaxies observed with IFUs, when compared with the SAMI survey, it is reassuring to see that previous work shown in this thesis can be adapted for use with very different instruments and observations at different wavelengths. The derived TFR matched well with those found in the wider literature. Notable was the behaviour of my fit having a more shallow gradient. This was attributed to the fact that I performed an asymmetric drift correction to the model's predicted values of V_{max} unlike the comparison fits shown alongside. Using galaxy properties from the respective data analysis pipeline catalogues, expected physical correlations were visible on along the TFR. This includes the relationship between V_{max} and age, as well as mass to light ratio. Interestingly, we find that the strength of these relationships differ between the two surveys.

In summary, in this thesis I have shown that it is possible, and in fact scientifically beneficial, to use machine learning in order to derive kinematic properties from extragalactic line emission surveys. I have shown that galaxies can be broadly categorised based on their kinematics using simulated backends in conjunction with unsupervised learning models. Improvements on these models have shown that it is

also possible to create physics-aware machine learning models which have comparable power to existing kinematic fitting routines, but with far less required computational costs. These models have been used to fit the velocity profiles of cold molecular, warm neutral, and singly ionised gas, the latter being useful in recovering the Hα TFR using the largest IFU sample to date. Consequently, expected correlations between galaxy mass (and rotational velocity) and derived physical properties from Hα observations have been measured on a large scale basis.

From a more data science oriented perspective, throughout this thesis I have demonstrated the successful use of ML and AI techniques to create proof-of-concept models. These models have shown great promise for use with future observational data but also have been shown to deliver new scientific results in the present, in particular in Chapter 4.

5.2 Ongoing & future work

Throughout this thesis, I have focused on applying machine learning to high level interferometric data products. Moment maps are highly abstracted from the raw visibilities that are captured by mm-radio interferometers. Therefore, it stands to reason that in order to maximise the efficiency of kinematic characterisation with upcoming surveys, one must begin investigating the performance of such methods lower in the data reduction pipeline. Ideally, an investigation of machine learning applications in the uv-plane would be optimal. However it makes more sense to progressively solve the challenges faced when working with higher dimensional and more complex data as we move away from moment maps, through to clean datacubes, dirty data cubes, and back to the raw visibilities. I have already begun investigating the use of self-supervised physics-aware learning on clean datacubes. The challenges faced in doing this involve accounting for changes in both brightness and velocity across voxels of 4-dimensional hypercubes, and memory issues when using larger input-output products.

In the near future, large surveys of galaxy clusters imaged at 21 cm with ASKAP (Koribalski et al., 2020) and MeerKAT (Jonas & MeerKAT Team, 2016) will be made available (Serra et al., 2016). These images will provide the perfect testing ground for overcoming modelling issues mentioned throughout the thesis. For example, given the low spatial and spectroscopic resolution of MeerKAT images, the effects of 'beam smearing' and heavily downsampled data products will have to be tackled in order for the models to return physically meaningful results. Therefore, the most logical step in developing self-supervised physics-aware models is to introduce

ways to convolve reconstructed images (or at least model the convolution operation) prior to the final output layer of the neural network. This approach could force the network to model the underlying, pre-convolved, physics –harbouring scientifically useful information.

A common theme among the later chapters of this thesis, is the use of simple generalised descriptions of galaxy profiles. The use of exponential intensity profiles and arctan velocity profiles was deliberate in order to reduce the complexity of the neural network models and avoid inducing parameter degeneracies (a fast way to break neural network training). In the future, it may possible to use these simple functions in a way that allows the networks to behave as outlier detection models. Those galaxies whose profiles cannot be modelled well with these simple functions will exhibit different reconstruction losses which may also exhibit spatial dependence. One can imagine a ram pressure stripped dwarf galaxy may not be modelled well by an arctan velocity profile or exponential intensity profile; such an example should have a very large reconstruction error. However, a symmetrically warped disc may have reasonably low reconstruction error due to the choice of a loss function which balances discrepancies (between input and reconstructed output) across the field of view. Therefore, using these models as outlier detection algorithms clearly requires further investigation in order to gauge whether they deliver performant and interpretable results. Alternatively, one could upgrade the network's ability to model galaxy profiles by investigating the use of different decoder profiles altogether. Adopting the use of funcitons only slightly more complex than an arctan (e.g. Rix et al. 1997) are a possible route to take. Or complete overhauls in an attempt to model complex warps and asymmetries may provide fruitful results. The degree to which we parameterise profiles depends on the science case, yet it offers many avenues for research moving into the future.

5.3 Concluding remarks

With the advent of next generation radio interferometers such as SKA on the near horizon, astronomers are poised to receive vast quantities of extra galactic observations. The scientific gain from such datasets promises interesting insights into the interactions and evolution of galaxies through time. The challenges of extracting robust science in a fast efficient manner will be tackled in many different ways and at different stages in data reduction pipelines. Here we have touched upon just a few approaches, focused on a few aspects of extragalactic galaxy studies, among many. The application of machine learning to kinematically characterise galaxies is a nascent

area of research, demonstrated by the simplicity of models discussed in this thesis and near-endless paths to take in adapting such models. With the ever increasing uptake of machine learning to solve problems in astronomy, I am excited to see how developments progress in machine learning assisted studies of galaxy kinematics going in the future. And I am pleased to have done my part in advancing the field.

Appendix A
Ancillary tables

	Layer	Layer Type	Units/number of filters	Size	Padding	Stride
Encoder	Input	Input		(64,64)		
	Conv1	Convolutional	8	(3,3)	1	1
	ReLU	Activation				
	Conv2	Convolutional	8	(3,3)	1	1
	ReLU	Activation				
	MaxPool	Max-pooling		(2,2)		1
	Conv3	Convolutional	16	(3,3)	1	1
	ReLU	Activation				
	Conv4	Convolutional	16	(3,3)	1	1
	ReLU	Activation				
	MaxPool	Max-pooling		(2,2)		1
	Linear	Fully-connected	3			
Decoder	Linear	Fully-connected	3			
	Up	Partial inverse max-pool		(2,2)		1
	ReLU	Activation				
	Trans1	Transposed Convolution	16	(3,3)	1	1
	ReLU	Activation				
	Trans2	Transposed Convolution	16	(3,3)	1	1
	Up	Partial inverse max-pool		(2,2)		1
	ReLU	Activation				
	Trans3	Transposed Convolution	8	(3,3)	1	1
	ReLU	Activation				
	Trans4	Transposed Convolution	8	(3,3)	1	1
	Ouput	Output		(64,64)		

Table A.1 Architecture for our autoencoder, featuring both encoder and decoder subnets. The decoder is a direct reflection of the encoder's structure.

OBJECT ID	SURVEY	Author Prediction (disturbed=0, regular=1)	Model Prediction ($\kappa < 0.5$=0, $\kappa > 0.5$=1)	Heuristic Result (TP=true positive, FP=false positive TN=true negative, FN=false negative)
ESO358-G063	AlFoCS	1	1	TP
ESO359-G002	AlFoCS	0	0	TN
FCC207	AlFoCS	1	1	TP
FCC261	AlFoCS	0	0	TN
FCC282	AlFoCS	0	1	FP
FCC332	AlFoCS	0	0	TN
MCG-06-08-024	AlFoCS	0	0	TN
NGC1351A	AlFoCS	1	0	TN
NGC1365	AlFoCS	1	1	TP
NGC1380	AlFoCS	1	1	TP
NGC1386	AlFoCS	1	1	TP
NGC1387	AlFoCS	1	1	TP
NGC1436	AlFoCS	1	1	TP
NGC1437B	AlFoCS	1	1	TP
PGC013571	AlFoCS	0	1	FP
NGC0383	WISDOM	1	1	TP
NGC0404	WISDOM	0	0	TN
NGC0449	WISDOM	1	1	TP
NGC0524	WISDOM	1	1	TP
NGC0612	WISDOM	1	1	TP
NGC1194	WISDOM	1	1	TP
NGC1574	WISDOM	1	1	TP
NGC3368	WISDOM	1	1	TP
NGC3393	WISDOM	1	1	TP
NGC4429	WISDOM	1	1	TP
NGC4501	WISDOM	1	1	TP
NGC4697	WISDOM	1	1	TP
NGC4826	WISDOM	1	1	TP
NGC5064	WISDOM	1	1	TP
NGC7052	WISDOM	1	1	TP

Table A.2 ALMA galaxies selected from the WISDOM and AlFoCS surveys. WISDOM targets have beam major axes ranging from 2.4″ to 6.7″ with a mean of 4.4″ and pixels/beam values ranging from 2.42 to 6.68 with a median value of 4.46. ALL WISDOM targets have channel widths of 2 $\mathrm{km\,s^{-1}}$ bar one target which has a channel width of 3 $\mathrm{km\,s^{-1}}$. AlFoCS targets have beam major axes ranging from 2.4″ to 3.3″ with a mean of 2.9″ and pixels/beam values ranging from 5.25 to 7.85 with a median value of 6.46. AlFoCS targets have channel widths ranging from 9.5 to 940 $\mathrm{km\,s^{-1}}$, with a median channel width of 50 $\mathrm{km\,s^{-1}}$. Of all 30 galaxies in the test set, 7 were identified by eye as most likely to be classified as $\kappa < 0.5$ and their associated model predictions are shown. 27 (90%) of the galaxies are classified as predicted by human eye. NGC1351A is the only false negative classification owing to its disconnected structure and edge on orientation.

LVHIS ID	OBJECT ID	Author Prediction (disturbed=0, regular=1)	Model Prediction ($\kappa < 0.5$=0, $\kappa > 0.5$=1)	Heuristic Result (TP=true positive, FP=false positive TN=true negative, FN=false negative)
LVHIS 001	ESO 349-G031	1	1	TP
LVHIS 003	ESO 410-G005	0	1	FP
LVHIS 004	NGC 55	1	1	TP
LVHIS 005	NGC 300	1	1	TP
LVHIS 007	NGC 247	1	1	TP
LVHIS 008	NGC 625	1	1	TP
LVHIS 009	ESO 245-G005	1	1	TP
LVHIS 010	ESO 245-G007	0	0	TN
LVHIS 011	ESO 115-G021	1	1	TP
LVHIS 012	ESO 154-G023	1	1	TP
LVHIS 013	ESO 199-G007	1	1	TP
LVHIS 015	NGC 1311	1	1	TP
LVHIS 017	IC 1959	1	1	TP
LVHIS 018	NGC 1705	1	1	TP
LVHIS 019	ESO 252-IG001	1	1	TP
LVHIS 020	ESO 364-G?029	1	1	TP
LVHIS 021	AM 0605-341	1	1	TP
LVHIS 022	NGC 2188	1	1	TP
LVHIS 023	ESO 121-G020	1	1	TP
LVHIS 024	ESO 308-G022	1	1	TP
LVHIS 025	AM 0704-582	1	1	TP
LVHIS 026	ESO 059-G001	1	1	TP
LVHIS 027	NGC 2915	1	1	TP
LVHIS 028	ESO 376-G016	1	1	TP
LVHIS 029	ESO 318-G013	1	1	TP
LVHIS 030	ESO 215-G?009	1	1	TP
LVHIS 031	NGC 3621	1	1	TP
LVHIS 034	ESO 320-G014	1	1	TP
LVHIS 035	ESO 379-G007	1	1	TP
LVHIS 036	ESO 379-G024	0	0	TN
LVHIS 037	ESO 321-G014	1	1	TP
LVHIS 039	ESO 381-G018	1	1	TP
LVHIS 043	NGC 4945	1	1	TP
LVHIS 044	ESO 269-G058	1	1	TP
LVHIS 046	NGC 5102	1	1	TP
LVHIS 047	AM 1321-304	0	0	TN
LVHIS 049	IC 4247	0	1	FP
LVHIS 050	ESO 324-G024	1	1	TP
LVHIS 051	ESO 270-G017	1	1	TP
LVHIS 053	NGC 5236	1	1	TP
LVHIS 055	NGC 5237	1	1	TP
LVHIS 056	ESO 444-G084	1	1	TP
LVHIS 057	NGC 5253	0	0	TP
LVHIS 058	IC 4316	0	0	TP
LVHIS 060	ESO 325-G?011	1	1	TP
LVHIS 063	ESO 383-G087	0	0	TN
LVHIS 065	NGC 5408	1	1	TP
LVHIS 066	Circinus Galaxy	1	1	TP
LVHIS 067	UKS 1424-460	1	1	TP
LVHIS 068	ESO 222-G010	1	1	TP
LVHIS 070	ESO 272-G025	0	0	TN
LVHIS 071	ESO 223-G009	1	1	TP
LVHIS 072	ESO 274-G001	1	1	TP
LVHIS 075	IC 4662	1	1	TP
LVHIS 076	ESO 461-G036	1	1	TP
LVHIS 077	IC 5052	1	1	TP
LVHIS 078	IC 5152	1	1	TP
LVHIS 079	UGCA 438	0	0	TN
LVHIS 080	UGCA 442	1	1	TP
LVHIS 081	ESO 149-G003	1	1	TP
LVHIS 082	NGC 7793	1	1	TP

Table A.3 LVHIS galaxies chosen from the LVHIS database as suitable for testing. The targets have beam major axes ranging from 5.3″ to 34.7″ with a mean of 13.2″ and have pixels/beam values ranging from 5.25 to 34.74 with a median value of 12.78. The channel widths are 4 $\mathrm{km\,s^{-1}}$ bar one target which has a channel width of 8 $\mathrm{km\,s^{-1}}$. Of all 61 galaxies in the test set, 10 (16%) were identified by eye as most likely to be classified as $\kappa < 0.5$ and their associated model predictions are shown. Of these 10 galaxies 8 were correctly identified as low κ by the binary classifier with no false negative predictions.

Table A.4 Information regarding the THINGS sample galaxies used throughout his work. Columns give the following information: *Object*, the target name as given in THINGS project publications, *Publication*, records the relevant publication in which the THINGS targets appear.

Object		Publication
DDO 53	NGC 3621	All from Walter et al. (2008); de Blok et al. (2008)
NGC 925	NGC 4736	
NGC 2403	NGC 4826	
NGC 2841	NGC 5055	
NGC 2903	NGC 5236	
NGC 3184	NGC 6946	
NGC 3198	NGC 7331	
NGC 3351	NGC 7793	
NGC 3521		

Table A.5 Information regarding the WISDOM project sample used throughout this work. Table columns give the following information: *Object*, the target name as given in WISDOM project publications, *Observation type*, gives the emission line ALMA observed for the target, *Publication*, records the relevant publication in which ALMA observations of the targets appear.

Object	Observation type	Publication
NGC 3665	^{12}CO(2-1)	Onishi et al. (2017)
NGC 0383	^{12}CO(2-1)	North et al. (2019)
NGC 0524	^{12}CO(2-1)	Smith et al. (2019)
NGC 1387	^{12}CO(2-0)	Zabel et al. (2020), Boyce et al. (in prep)
NGC 4429	^{12}CO(3-2)	Davis et al. (2017b)
NGC 4697	^{12}CO(2-1)	Davis et al. (2017a)

Appendix B
Ancillary plots

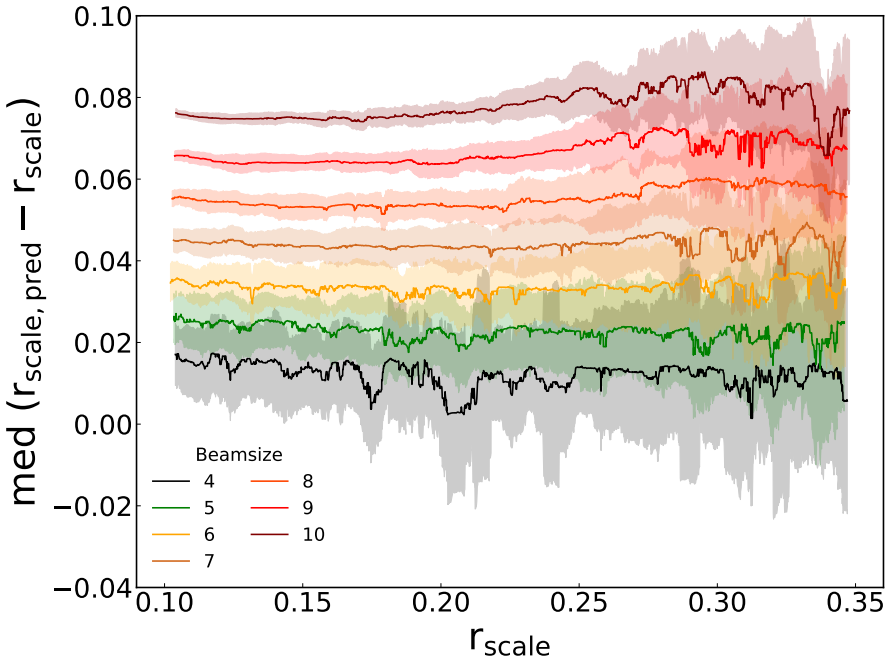


Figure B.1. The effects of varying the ratio of beam size to galaxy extent. It is clear to see that an increased beam size results in an artificial lengthening of the intensity profile scale length. It can also be seen that the spread in median offset increases with r_{scale}, which occurs due to information loss as the convolved flux is "smeared" out beyond the field of view. The value of r_{scale} at which this effect begins to take hold is clearly inversely proportional to the beamsize.

Bibliography

Ackermann S., Schawinski K., Zhang C., Weigel A. K., Turp M. D., 2018, MNRAS, 479, 415

Ahumada R., et al., 2020, ApJS, 249, 3

Alger M. J., et al., 2018, Monthly Notices of the Royal Astronomical Society, 478, 5547

Alladin S., Hasan S., 2007, Resonance, 12, 13

Allen J. T., et al., 2015, MNRAS, 446, 1567

Andrianomena S., Rafieferantsoa M., Davé R., 2019, arXiv e-prints, p. arXiv:1906.04198

Aragon-Calvo M. A., Carvajal J. C., 2020, Monthly Notices of the Royal Astronomical Society, 498, 3713

Armstrong D. J., et al., 2016, MNRAS, 456, 2260

Armstrong D. J., Pollacco D., Santerne A., 2017, MNRAS, 465, 2634

Balkowski C., Bottinelli L., Chamaraux P., Gouguenheim L., Heidmann J., 1974, A&A, 34, 43

Beckman J., Carretero C., Vazdekis A., 2008, Chinese Journal of Astronomy and Astrophysics Supplement, 8, 77

Begum A., Chengalur J. N., Karachentsev I. D., 2005, A&A, 433, L1

Begum A., Chengalur J. N., Karachentsev I. D., Sharina M. E., 2008, MNRAS, 386, 138

Bekeraité S., et al., 2016, A&A, 593, A114

Bekki K., 2019, Monthly Notices of the Royal Astronomical Society, 485, 1924

Bell E. F., de Jong R. S., 2001, The Astrophysical Journal, 550, 212

Bhandare A., Bhide M. V., Gokhale P., Chandavarkar R., 2016.

Binney J., Tremaine S., 2008, Galactic Dynamics: Second Edition

Blais-Ouellette S., Carignan C., Amram P., Côté S., 1999, AJ, 118, 2123

Bland-Hawthorn J., et al., 2011, Opt. Express, 19, 2649

Bloom J. V., et al., 2017a, MNRAS, 465, 123

Bloom J. V., et al., 2017b, MNRAS, 472, 1809

Bluck A. F. L., Mendel J. T., Ellison S. L., Moreno J., Simard L., Patton D. R., Starkenburg E., 2014, MNRAS, 441, 599

Bolatto A. D., Wolfire M., Leroy A. K., 2013, ARAA, 51, 207

Bolton A. S., et al., 2012, AJ, 144, 144

Bosma A., 1978, PhD thesis, -

Bosma A., 2016, Proceedings of the International Astronomical Union, 11, 220–222

Bottinelli L., 1971, A&A, 10, 437

Bourlard H., Kamp Y., 1988, Biological cybernetics, 59, 291

Bournaud F., Duc P. A., Amram P., Combes F., Gach J. L., 2004, A&A, 425, 813

Bournaud F., Dekel A., Teyssier R., Cacciato M., Daddi E., Juneau S., Shankar F., 2011, ApJ, 741, L33

Breiman L., 2001, Machine Learning, 45, 5

Bremer M. N., et al., 2018, MNRAS, 476, 12

Brittain J. E., 1984, IEEE Proceedings, 72, 709

Bryant J. J., Bland-Hawthorn J., 2016, in Skillen I., Balcells M., Trager S., eds, Astronomical Society of the Pacific Conference Series Vol. 507, Multi-Object Spectroscopy in the Next Decade: Big Questions, Large Surveys, and Wide Fields. p. 291

Bryant J. J., Bland-Hawthorn J., Fogarty L. M. R., Lawrence J. S., Croom S. M., 2013, Monthly Notices of the Royal Astronomical Society, 438, 869

Bryant J. J., et al., 2015, Monthly Notices of the Royal Astronomical Society, 447, 2857

Bundy K., et al., 2015, ApJ, 798, 7

Butcher H., Oemler A. J., 1978, ApJ, 219, 18

Byrd J., Lipton Z. C., 2018, arXiv e-prints, p. arXiv:1812.03372

Cacciato M., Dekel A., Genel S., 2012, Monthly Notices of the Royal Astronomical Society, 421, 818

Caon N., Capaccioli M., D'Onofrio M., 1993, Monthly Notices of the Royal Astronomical Society, 265, 1013

Cappellari M., et al., 2013, MNRAS, 432, 1709

Carrasco-Davis R., et al., 2018, arXiv e-prints, p. arXiv:1807.03869

Cherinka B., et al., 2019, AJ, 158, 74

Colless M., et al., 2003, arXiv e-prints, pp astro–ph/0306581

Cortes C., Vapnik V., 1995, Machine Learning, 20, 273

Courteau S., 1997, AJ, 114, 2402

Crain R. A., et al., 2015, MNRAS, 450, 1937

Croom S. M., et al., 2012, MNRAS, 421, 872

Dame T. M., 2011, arXiv e-prints, p. arXiv:1101.1499

Davis T. A., et al., 2011, Monthly Notices of the Royal Astronomical Society, 414, 968

Davis T. A., Bureau M., Cappellari M., Sarzi M., Blitz L., 2013, Nature, 494, 328

Davis T. A., Greene J., Ma C.-P., Pandya V., Blakeslee J. P., McConnell N., Thomas J., 2015, Monthly Notices of the Royal Astronomical Society, 455, 214

Davis T. A., Bureau M., Onishi K., Cappellari M., Iguchi S., Sarzi M., 2017a, Monthly Notices of the Royal Astronomical Society, 468, 4675

Davis T. A., et al., 2017b, Monthly Notices of the Royal Astronomical Society, 473, 3818

Davis T. A., van de Voort F., Rowlands K., McAlpine S., Wild V., Crain R. A., 2019, MNRAS, 484, 2447

Davis T. A., Zabel N., Dawson J. M., 2020, KinMS: Three-dimensional kinematic modelling of arbitrary gas distributions (ascl:2006.003)

Dawson K. S., et al., 2013, AJ, 145, 10

Dawson J. M., Davis T. A., Gomez E. L., Schock J., Zabel N., Williams T. G., 2019, Monthly Notices of the Royal Astronomical Society, 491, 2506

Dawson J. M., Davis T., Gomez E., Schock J., 2021a, Monthly Notices of the Royal Astronomical Society, 503

Dawson J. M., Davis T. A., Gomez E. L., Schock J., 2021b, MNRAS, 503, 574

Dekel A., Sari R., Ceverino D., 2009, ApJ, 703, 785

Dewdney P. E., Hall P. J., Schilizzi R. T., Lazio T. J. L. W., 2009, IEEE Proceedings, 97, 1482

Di Teodoro E. M., Fraternali F., 2015, MNRAS, 451, 3021

Diaz J. D., Bekki K., Forbes D. A., Couch W. J., Drinkwater M. J., Deeley S., 2019, MNRAS, p. 1007

Dickey J. M., 1990, in Thronson Harley A. J., Shull J. M., eds, Astrophysics and Space Science Library Vol. 161, The Interstellar Medium in Galaxies. pp 473–481, doi:10.1007/978-94-009-0595-5_19

Dieleman S., Willett K. W., Dambre J., 2015, MNRAS, 450, 1441

Dinh L., Pascanu R., Bengio S., Bengio Y., 2017, arXiv e-prints, p. arXiv:1703.04933

Domínguez Sánchez H., Huertas-Company M., Bernardi M., Tuccillo D., Fischer J. L., 2018a, MNRAS, 476, 3661

Domínguez Sánchez H., et al., 2018b, MNRAS, 484, 93

Douglass K. A., Smith J. A., Demina R., 2019, ApJ, 886, 153

Draine B. T., 2011, Physics of the Interstellar and Intergalactic Medium

Dressler A., 1980, ApJ, 236, 351

Driver S. P., et al., 2009, Astronomy and Geophysics, 50, 5.12

Driver S. P., et al., 2011, MNRAS, 413, 971

Drory N., et al., 2015, AJ, 149, 77

Duffy A. R., Meyer M. J., Staveley-Smith L., Bernyk M., Croton D. J., Koribalski B. S., Gerstmann D., Westerlund S., 2012, MNRAS, 426, 3385

Dumoulin V., Visin F., 2016, arXiv e-prints, p. arXiv:1603.07285

ESO 2017, ALMA Inauguration Heralds New Era of Discovery - Revolutionary telescope will enable unprecedented views of the cosmos, `https://www.eso.org/public/news/eso1312/`

Eskridge P. B., Frogel J. A., 1999, Astrophys. Space Science, 269, 427

Faber S. M., 1973, ApJ, 179, 731

Fang J. J., Faber S. M., Koo D. C., Dekel A., 2013, The Astrophysical Journal, 776, 63

Fukushima K., 1980, Biological Cybernetics, 36, 193

Gabbard H., Williams M., Hayes F., Messenger C., 2018, Phys. Rev. Lett., 120, 141103

Gaia Collaboration et al., 2021, A&A, 649, A1

Gal Y., 2016, PhD thesis, University of Cambridge

Gal Y., Ghahramani Z., 2016, in Lee D. D., Sugiyama M., Luxburg U. V., Guyon I., Garnett R., eds, , Advances in Neural Information Processing Systems 29. Curran Associates, Inc., pp 1019–1027

Gallazzi A., Charlot S., Brinchmann J., White S. D. M., Tremonti C. A., 2005, Monthly Notices of the Royal Astronomical Society, 362, 41

Galloway M. A., et al., 2015, Monthly Notices of the Royal Astronomical Society, 448, 3442

Gavazzi G., Fumagalli M., Cucciati O., Boselli A., 2010, A&A, 517, A73

Gentile et al., 2013, A&A, 554, A125

George D., Huerta E. A., 2018, Physics Letters B, 778, 64

Glover S. C., Clark P. C., 2016, Monthly Notices of the Royal Astronomical Society, 456, 3596

Goodfellow I. J., Shlens J., Szegedy C., 2014, arXiv e-prints, p. arXiv:1412.6572

Grand R. J. J., Bovy J., Kawata D., Hunt J. A. S., Famaey B., Siebert A., Monari G., Cropper M., 2015, MNRAS, 453, 1867

Gunn J. E., Gott III J. R., 1972, Astrophys. J., 176, 1

Gunn J. E., et al., 2006, AJ, 131, 2332

Hansen S., Conselice C. J., Fraser-McKelvie A., Ferreira L., 2020, Research Notes of the American Astronomical Society, 4, 185

He K., Zhang X., Ren S., Sun J., 2015, arXiv e-prints, p. arXiv:1512.03385

Heald G., et al., 2011, A&A, 526, A118

Hibbard J. E., van der Hulst J. M., Barnes J. E., Rich R. M., 2001, AJ, 122, 2969

Hinton G. E., Salakhutdinov R. R., 2006, Science, 313, 504

Ho S. H., Martin C. L., Kacprzak G. G., Churchill C. W., Turner M. L., 2019, in American Astronomical Society Meeting Abstracts #233. p. 239.04

Hubble E. P., 1929, ApJ, 69, 103

Hubble E. P., 1936, Realm of the Nebulae

Johnston S., et al., 2007, PASA, 24, 174

Johnston S., et al., 2008, Experimental Astronomy, 22, 151

Jonas J., MeerKAT Team 2016, in MeerKAT Science: On the Pathway to the SKA. p. 1

Józsa G. I. G., Kenn F., Klein U., Oosterloo T. A., 2007, A&A, 468, 731

Jurić M., et al., 2017, in Lorente N. P. F., Shortridge K., Wayth R., eds, Astronomical Society of the Pacific Conference Series Vol. 512, Astronomical Data Analysis Software and Systems XXV. p. 279 (arXiv:1512.07914)

Kamphuis P., Józsa G. I. G., Oh S.-H., Spekkens K., Urbancic N., Serra P., Koribalski B. S., Dettmar R.-J., 2015, Monthly Notices of the Royal Astronomical Society, 452, 3139

Kassin S. A., et al., 2007, The Astrophysical Journal, 660, L35

Kauffmann G., Colberg J. M., Diaferio A., White S. D. M., 1999, Monthly Notices of the Royal Astronomical Society, 303, 188

Kawaguchi K., Pack Kaelbling L., Bengio Y., 2017, arXiv e-prints, p. arXiv:1710.05468

Kellermann K. I., Bouton E. N., Brandt S. S., 2020, Open Skies; The National Radio Astronomy Observatory and Its Impact on US Radio Astronomy, doi:10.1007/978-3-030-32345-5.

Kelvin L. S., et al., 2014, MNRAS, 444, 1647

Kingma D. P., Ba J., 2014, arXiv e-prints, p. arXiv:1412.6980

Kodirov E., Xiang T., Gong S., 2017, arXiv e-prints, p. arXiv:1704.08345

Kohonen T., 1990, Proceedings of the IEEE, 78, 1464

Koribalski B. S., et al., 2018, MNRAS, 478, 1611

Koribalski B. S., et al., 2020, Astrophys. Space Science, 365, 118

Krajnović D., Cappellari M., de Zeeuw P. T., Copin Y., 2006, MNRAS, 366, 787

Krizhevsky A., Sutskever I., Hinton G. E., 2012, in Pereira F., Burges C. J. C., Bottou L., Weinberger K. Q., eds, , Advances in Neural Information Processing Systems 25. Curran Associates, Inc., pp 1097–1105

Krizhevsky A., Sutskever I., Hinton G. E., 2017, Commun. ACM, 60, 84–90

Lagattuta D. J., Mould J. R., Staveley-Smith L., Hong T., Springob C. M., Masters K. L., Koribalski B. S., Jones D. H., 2013, ApJ, 771, 88

Lagos C. d. P., et al., 2015, MNRAS, 452, 3815

Law D. R., et al., 2015, AJ, 150, 19

Law D. R., et al., 2021, AJ, 161, 52

Lawrence A., Kerton C., Struck C., Smith B. J., 2020, The Astrophysical Journal, 891, 11

Lecun Y., Bottou L., Bengio Y., Haffner P., 1998, Proceedings of the IEEE, 86, 2278

Leitherer C., Robert C., Drissen L., 1992, ApJ, 401, 596

Lelli F., McGaugh S. S., Schombert J. M., 2016, AJ, 152, 157

Li D., Pan Z., 2016, Radio Science, 51, 1060

Licquia T. C., Newman J. A., 2015, ApJ, 806, 96

Lim T.-S., Loh W.-Y., Shih Y.-S., 2000, Mach. Learn., 40, 203–228

Liu X., Zhang F., Hou Z., Wang Z., Mian L., Zhang J., Tang J., 2020, arXiv e-prints, p. arXiv:2006.08218

Lotz J. M., Jonsson P., Cox T. J., Primack J. R., 2008, Monthly Notices of the Royal Astronomical Society, 391, 1137

Lukic V., Brüggen M., Banfield J. K., Wong O. I., Rudnick L., Norris R. P., Simmons B., 2018, MNRAS, 476, 246

Ma Z., Zhu J., Li W., Xu H., 2018, arXiv e-prints, p. arXiv:1806.00398

Madau P., Dickinson M., 2014, ARAA, 52, 415

Mancillas B., Combes F., Duc P.-A., 2019, Astronomy & Astrophysics, 630, A112

Martinez-Valpuesta I., Shlosman I., Heller C., 2006, The Astrophysical Journal, 637, 214

Martinsson T. P. K., Verheijen M. A. W., Westfall K. B., Bershady M. A., Andersen D. R., Swaters R. A., 2013, A&A, 557, A131

Masci J., Meier U., Cireşan D., Schmidhuber J., 2011, in Honkela T., Duch W., Girolami M., Kaski S., eds, Artificial Neural Networks and Machine Learning – ICANN 2011. Springer Berlin Heidelberg, Berlin, Heidelberg, pp 52–59

Masters K. L., Springob C. M., Huchra J. P., 2008, AJ, 135, 1738

McGaugh S. S., Schombert J. M., Bothun G. D., de Blok W. J. G., 2000, ApJ, 533, L99

Mihalas D., Routly P. M., 1968, Galactic astronomy

Mistani P. A., et al., 2015, Monthly Notices of the Royal Astronomical Society, 455, 2323

Moore B., Katz N., Lake G., Dressler A., Oemler A., 1996, nature, 379, 613

Moustakas J., et al., 2013, ApJ, 767, 50

NRAO 2022, Very Large Array, https://public.nrao.edu/telescopes/vla/

Neistein E., Maoz D., Rix H.-W., Tonry J. L., 1999, AJ, 117, 2666

Neistein E., van den Bosch F. C., Dekel A., 2006, MNRAS, 372, 933

Nikolaou N., et al., 2020, arXiv e-prints, p. arXiv:2010.15996

North E. V., et al., 2019, Monthly Notices of the Royal Astronomical Society, 490, 319

Oemler Augustus J., 1974, ApJ, 194, 1

Oh S.-H., de Blok W. J. G., Brinks E., Walter F., Kennicutt R. C., 2011, The Astronomical Journal, 141, 193

Oh S.-H., Staveley-Smith L., Spekkens K., Kamphuis P., Koribalski B. S., 2017, Monthly Notices of the Royal Astronomical Society, 473, 3256

Onishi K., Iguchi S., Davis T. A., Bureau M., Cappellari M., Sarzi M., Blitz L., 2017, Monthly Notices of the Royal Astronomical Society, 468, 4663

Oort J. H., 1997, The letters and papers of Jan Hendrik Oort : as archived in the University Library, Leiden. Astrophysics and space science library ; 213, Kluwer Academic Publishers, Dordrecht ; London

Oosterloo T., Verheijen M., van Cappellen W., 2010, in ISKAF2010 Science Meeting. p. 43 (arXiv:1007.5141)

Opik E., 1922, ApJ, 55, 406

Owers M. S., et al., 2017, Monthly Notices of the Royal Astronomical Society, 468, 1824

Parry O. H., Eke V. R., Frenk C. S., 2009, MNRAS, 396, 1972

Paszke A., et al., 2017, in NIPS-W.

Percival W. J., et al., 2001, MNRAS, 327, 1297

Plaut E., 2018, arXiv e-prints, p. arXiv:1804.10253

Polletta M., Nesvadba N. P. H., Neri R., Omont A., Berta S., Bergeron J., 2011, A&A, 533, A20

Qiu Y., Bogdanović T., Li Y., Park K., Wise J. H., 2019, ApJ, 877, 47

Reber G., 1940, ApJ, 91, 621

Reiman D. M., Göhre B. E., 2019, MNRAS, 485, 2617

Ren J., et al., 2020, Monthly Notices of the Royal Astronomical Society, 499, 3399

Reyes R., Mandelbaum R., Gunn J. E., Pizagno J., Lackner C. N., 2011, Monthly Notices of the Royal Astronomical Society, 417, 2347

Rix H.-W., Guhathakurta P., Colless M., Ing K., 1997, Monthly Notices of the Royal Astronomical Society, 285, 779

Roberts M. S., 1969, AJ, 74, 859

Robotham A. S. G., et al., 2011, MNRAS, 416, 2640

Rosenblatt F., 1958, Psychological review, 65 6, 386

Rousseeuw P., Driessen K., 2006, Data Min. Knowl. Discov., 12, 29

Rumelhart D. E., Hinton G. E., Williams R. J., 1986, Nature, 323, 533

Rumelhart D. E., Hinton G. E., Williams R. J., 1988, Learning Representations by Back-Propagating Errors. MIT Press, Cambridge, MA, USA, p. 696–699

Russakovsky O., et al., 2014, arXiv e-prints, p. arXiv:1409.0575

Sales L. V., Navarro J. F., Theuns T., Schaye J., White S. D. M., Frenk C. S., Crain R. A., Dalla Vecchia C., 2012, MNRAS, 423, 1544

Sánchez S. F., et al., 2016a, Rev. Mexicana Astron. Astrofis., 52, 21

Sánchez S. F., et al., 2016b, Rev. Mexicana Astron. Astrofis., 52, 171

Sancisi R., Fraternali F., Oosterloo T., van der Hulst T., 2008, A&A Rev., 15, 189

Sarmiento R., Huertas-Company M., Knapen J. H., Sánchez S. F., Domínguez Sánchez H., Drory N., Falcón-Barroso J., 2021, arXiv e-prints, p. arXiv:2104.08292

Scaife A. M. M., Porter F., 2021, MNRAS, 503, 2369

Schawinski K., Thomas D., Sarzi M., Maraston C., Kaviraj S., Joo S.-J., Yi S. K., Silk J., 2007, MNRAS, 382, 1415

Schaye J., et al., 2015, MNRAS, 446, 521

Scott N., et al., 2018, Monthly Notices of the Royal Astronomical Society, 481, 2299

Serra P., et al., 2013, MNRAS, 428, 370

Serra P., et al., 2016, in MeerKAT Science: On the Pathway to the SKA. p. 8 (arXiv:1709.01289)

Shallue C. J., Vanderburg A., 2018, AJ, 155, 94

Shapley H., Curtis H. D., 1921, Bulletin of the National Research Council, 2, 171

Sharp R., et al., 2006, Proc. SPIE, 6269

Sharp R., et al., 2015, MNRAS, 446, 1551

Shen A. X., Bekki K., 2020, MNRAS, 497, 5090

Shen H., George D., Huerta E. A., Zhao Z., 2017, arXiv e-prints, p. arXiv:1711.09919

Shostak G. S., 1975, ApJ, 198, 527

Simons R. C., Kassin S. A., Weiner B. J., Heckman T. M., Lee J. C., Lotz J. M., Peth M., Tchernyshyov K., 2015, MNRAS, 452, 986

Smee S. A., et al., 2013, AJ, 146, 32

Smith M. D., et al., 2019, Monthly Notices of the Royal Astronomical Society, 485, 4359

Spekkens K., Sellwood J. A., 2007, ApJ, 664, 204

Spekkens K., Giovanelli R., Haynes M. P., 2005, The Astronomical Journal, 129, 2119

Springel V., et al., 2005, Nature, 435, 629

Srivastava N., Hinton G., Krizhevsky A., Sutskever I., Salakhutdinov R., 2014, Journal of Machine Learning Research, 15, 1929

Stark D. V., McGaugh S. S., Swaters R. A., 2009, AJ, 138, 392

Stark D. V., et al., 2018, MNRAS, 480, 2217

Steinkraus D., Buck I., Simard P., 2005, in Eighth International Conference on Document Analysis and Recognition (ICDAR'05). pp 1115–1120 Vol. 2, doi:10.1109/ICDAR.2005.251

Strömberg G., 1946, ApJ, 104, 12

Swaters R. A., 1999, PhD thesis, -

Tang H., Scaife A. M. M., Leahy J. P., 2019, MNRAS, 488, 3358

The EAGLE team 2017, arXiv e-prints, p. arXiv:1706.09899

Tiley A. L., et al., 2016, MNRAS, 460, 103

Trachternach C., de Blok W. J. G., Walter F., Brinks E., Kennicutt R. C. J., 2008, AJ, 136, 2720

Trachternach C., de Blok W. J. G., McGaugh S. S., van der Hulst J. M., Dettmar R. J., 2009, A&A, 505, 577

Tully R. B., Fisher J. R., 1977, A&A, 500, 105

Tully R. B., Pierce M. J., 2000, ApJ, 533, 744

Turner M. S., 1999, Philosophical Transactions of the Royal Society of London Series A, 357, 7

Vaughan S. P., et al., 2020, Monthly Notices of the Royal Astronomical Society, 496, 3841

Walmsley M., et al., 2019, Monthly Notices of the Royal Astronomical Society, 491, 1554

Walter F., Brinks E., Blok W., Bigiel F., Kennicutt R., Thornley M., Leroy A., 2008, The Astronomical Journal, v.136, 2563-2647 (2008), 136

Warren B. E., Jerjen H., Koribalski B. S., 2004, AJ, 128, 1152

Weijmans A.-M., Krajnović D., van de Ven G., Oosterloo T. A., Morganti R., de Zeeuw P. T., 2008, MNRAS, 383, 1343

Wong T., Blitz L., 2002, ApJ, 569, 157

Wootten A., Thompson A. R., 2009, Proceedings of the IEEE, 97, 1463

Worthey G., Faber S. M., Gonzalez J. J., Burstein D., 1994, ApJS, 94, 687

Wu C., et al., 2019, MNRAS, 482, 1211

Yan R., et al., 2016, AJ, 152, 197

Yanny B., et al., 2009, AJ, 137, 4377

Yoon I., 2017, MNRAS, 466, 1987

York D. G., et al., 2000, AJ, 120, 1579

Zabel N., et al., 2019, MNRAS, 483, 2251

Zabel N., et al., 2020, MNRAS, 496, 2155

Zabludoff A. I., Mulchaey J. S., 1998, ApJ, 496, 39

Zevin M., et al., 2017, Classical and Quantum Gravity, 34, 064003

de Blok W. J. G., Walter F., Brinks E., Trachternach C., Oh S. H., Kennicutt R. C. J., 2008, AJ, 136, 2648

ngVLA 2022, ngVLA, `https://ngvla.nrao.edu/`

van Albada T. S., Bahcall J. N., Begeman K., Sancisi R., 1985, ApJ, 295, 305

van Cittert P., 1934, Physica, 1, 201