

## Research Article

# Galaxy 3D shape recovery using mixture density network

Suk Yee Yong<sup>1,2</sup> , K.E. Harborne<sup>3,2</sup>, Caroline Foster<sup>4,2</sup>, Robert Bassett<sup>5,2</sup>, Gregory B. Poole<sup>6,5</sup> and Mitchell Cavanagh<sup>3</sup>

<sup>1</sup>CSIRO Space and Astronomy, Epping, NSW, Australia, <sup>2</sup>ARC Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D), Canberra, Australia, <sup>3</sup>International Centre for Radio Astronomy (ICRAR), M468, The University of Western Australia, Crawley, WA, Australia, <sup>4</sup>School of Physics, University of New South Wales, Sydney, NSW, Australia, <sup>5</sup>The Centre for Astrophysics & Supercomputing, Swinburne University of Technology, Hawthorn, VIC, Australia and <sup>6</sup>Astronomy Data and Computing Services (ADACS), Swinburne University of Technology, Hawthorn, VIC, Australia

### Abstract

Since the turn of the century, astronomers have been exploiting the rich information afforded by combining stellar kinematic maps and imaging in an attempt to recover the intrinsic, three-dimensional (3D) shape of a galaxy. A common intrinsic shape recovery method relies on an expected monotonic relationship between the intrinsic misalignment of the kinematic and morphological axes and the triaxiality parameter. Recent studies have, however, cast doubt about underlying assumptions relating shape and intrinsic kinematic misalignment. In this work, we aim to recover the 3D shape of *individual* galaxies using their projected stellar kinematic and flux distributions using a supervised machine learning approach with mixture density network (MDN). Using a mock dataset of the EAGLE hydrodynamical cosmological simulation, we train the MDN model for a carefully selected set of common kinematic and photometric parameters. Compared to previous methods, we demonstrate potential improvements achieved with the MDN model to retrieve the 3D galaxy shape along with the uncertainties, especially for prolate and triaxial systems. We make specific recommendations for recovering galaxy intrinsic shapes relevant for current and future integral field spectroscopic galaxy surveys.

**Keywords:** Galaxies: kinematics and dynamics; galaxies: fundamental parameters; galaxies: statistics; methods: data analysis; methods: statistical

(Received 22 February 2024; revised 28 March 2024; accepted 5 April 2024)

### 1. Introduction

The intrinsic, three-dimensional (3D) shape of a galaxy is a fundamental property that has been shown to closely relate to a variety of other physical parameters (e.g. Ryden 2006; Weijmans et al. 2014; van de Sande et al. 2018). Indeed, projected visual shape has been a key discriminator of galaxy type since the early classification schemes of Hubble (1926). A key challenge in linking observed galaxy shapes with other physical properties, however, is the reliable translation of two-dimensional (2D) projected information into intrinsic shape properties (e.g. Padilla & Strauss 2008; Weijmans et al. 2014; Foster et al. 2016). Under the simplifying assumption that all galaxies can be reasonably approximated as 3D ellipsoids, a galaxy with any particular 3D shape can be projected into a wide range of 2D elliptical shapes. A perfectly axisymmetric disc galaxy, for example, may appear as circular when viewed face-on, or highly elliptical when viewed edge-on (with the axis ratio limited only by the disc's intrinsic thickness). Due to this degeneracy, the problem of inferring the intrinsic shape of galaxies has long vexed astronomers.

Throughout this work, and as is common in the literature (e.g. Contopoulos 1956; Binggeli 1980; Franx et al. 1991; Lambas et al. 1992; Weijmans et al. 2014; Foster et al. 2017), we assume that a galaxy's 3D shape can be described by an ellipsoid and that its 2D projection can similarly be estimated as an ellipse. The projected

ellipse shape can be fully described with the lengths of its principal axes as follows:

$$\frac{x^2}{A^2} + \frac{y^2}{B^2} = 1, \quad (1)$$

where the semi-major and semi-minor axes,  $A$  and  $B$ , are aligned with the  $x$ - and  $y$ -directions, respectively. Similarly, an ellipsoid can be fully described as follows:

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1, \quad (2)$$

where  $a$ ,  $b$ , and  $c$  are the lengths of the principal semi-axes aligned with the associated Cartesian directions ( $x$ ,  $y$ ,  $z$ ) such that  $a > b > c$ . In galaxy shape studies, we are typically interested in the axis ratios rather than the absolute length of any axis and can thus reduce the parameterisation of shape down to one or two values for an ellipse and an ellipsoid, respectively. Without loss of generalisation, we may thus quantify 2D shapes through the axis ratio,  $B/A$ , and ellipsoidal shapes via two axis ratios  $p = b/a$  and  $q = c/a$ . As  $A$  and  $a$  represent the longest axis, in both the 2D and 3D cases, axis ratios are constrained to  $< 1$  by definition.

Since observations of individual galaxies provide a single measure of the projected shape, 3D shape inferences may be obtained through statistical methods applied to galaxy samples (e.g. Binney 1978; Vincent & Ryden 2005; Padilla & Strauss 2008). The simplest methods rely on a comparison between the observed axis ratio distributions of a given galaxy sample and the expected distribution for projections of randomly oriented ellipsoids (e.g. Sandage et al. 1970; Fasano 1991; Kimm & Yi 2007). Using this method

**Corresponding author:** Suk Yee Yong; Email: [sukyee.yong@csiro.au](mailto:sukyee.yong@csiro.au)

**Cite this article:** Yong SY, Harborne KE, Foster C, Bassett R, Poole GB and Cavanagh M. (2024) Galaxy 3D shape recovery using mixture density network. *Publications of the Astronomical Society of Australia* 41, e033, 1–16. <https://doi.org/10.1017/pasa.2024.32>

one can infer the 3D shape distribution of a given galaxy sample by varying the assumed underlying shape distribution of the randomly oriented ellipsoids until the resultant 2D shapes are well matched to the observations. There remains a significant degeneracy underlying such methods since a 2D shape distribution may be produced by a variety of underlying 3D shapes due to projection effects. Axisymmetry (i.e.  $a = b$ ) may be assumed to break this degeneracy.

Resolved galaxy kinematics represent a hopeful avenue to break longstanding degeneracies (e.g. Binney 1985; Franx *et al.* 1991; Statler 1994). Early attempts at utilising kinematic maps in shape recovery relied on radio interferometry, a technique that is limited to very nearby galaxies (e.g. Bak & Statler 2000). Detailed computationally intensive modelling of the kinematics and flux distribution is often required to constrain the intrinsic shape of individual galaxies (e.g. Statler 1994; van den Bosch & van de Ven 2009). These usually require expert know-how and assumptions to break degeneracies.

With the more recent proliferation of integral field spectroscopic (IFS) instruments, resolved kinematics now exist for large samples of galaxies covering a broad range of galaxy properties (e.g. Cappellari *et al.* 2011; Croom *et al.* 2012; Sanchez *et al.* 2012; Foster *et al.* 2021). Indeed, data from IFS surveys have been used to infer 3D shape distributions from a variety of samples (e.g. Weijmans *et al.* 2014; Foster *et al.* 2017; Li *et al.* 2018a; Ene *et al.* 2018).

One major difficulty in testing the reliability of such efforts, however, is that the intrinsic shape distributions of observational samples is not known a priori. In this sense, the results of 3D shape recovery must place their faith in the dependability of the theoretical underpinning of the kinematics based methods when presenting the resulting 3D shape distributions. One can, however, provide meaningful tests of the method by performing a series of mock observations of galaxies drawn from large scale, cosmological simulations from which the true, 3D shapes are already known. Such a test was performed on the methods of Franx *et al.* (1991) in Bassett & Foster (2019), who found that while  $q$  can be recovered reasonably well for most galaxy types,  $p$  is poorly recovered with the output distribution strongly favouring  $p = 1$  in all cases.

The reason for the disproportionate recovery of  $p = 1$  in Bassett & Foster (2019) was identified to be the assumed relationship between the intrinsic kinematic misalignment angle ( $\psi_{int}$ ) and underlying galaxy shape (Weijmans *et al.* 2014). To reduce the number of unknown parameters within the fitted model, it was assumed that  $\psi_{int}$  was solely determined by the underlying 3D shape of the galaxy, something that was not borne out in the simulations, and thus need not be true in nature. We are free from making such assumptions when using a non-linear, machine learning approach. Using a variety of photometric and kinematic measurements commonly derived for IFS observations of galaxies, we can probe the shape information locked into various parameters, and quantify their shape-determining ‘power’ relative to one another.

In recent works, we have seen an increased uptake in the use of machine learning to explore astrophysical questions. For example, the primary drivers for individual cosmological parameters is being explored with the Cosmology and Astrophysics with Machine Learning Simulations (CAMELS; Villaescusa-Navarro *et al.* 2021); for studies of hierarchical accretion histories and unlocking the un-observable merger history of a given galaxy (e.g. Bottrell *et al.* 2022); and for classifying morphology in large

astronomical data sets in efficient ways (e.g. Cavanagh *et al.* 2021). The machine learning approach is particularly useful when combined with large cosmological models of galaxy evolution, providing ground truth for labels that can then be applied to real data.

In this work, we revisit the findings of Bassett & Foster (2019) and attempt to use machine learning to recover the underlying 3D shape of individual galaxies without underlying assumptions. As was done by Bassett & Foster (2019), we use galaxy simulations, in which the true underlying shapes of galaxies are known, and mock IFS observations of these objects, such that commonly used kinematic measurements can be derived. With these data, we train a neural network model, called a mixture density network (MDN), to recover the underlying axis ratios, and hence the 3D shape, of an observed system.

In this paper, we describe the construction of the training set in Section 2, outline the development of the machine learning algorithm in Section 3, highlight our results in Section 4, discuss the ability of the algorithm to recover  $p$  and  $q$  and future directions in Section 5, and provide a summary in Section 6.

## 2. Construction of the training dataset

We perform post-processing of  $z = 0$  subhalo particle data from the EAGLE cosmological, hydrodynamic simulation (Schaye *et al.* 2015; Crain *et al.* 2015) to produce both 3D measurements and 2D mock-IFS observations. Mock observations of this simulation are produced using the open-source code, SIMSPIN (Harborne *et al.* 2023), for 2 519 galaxies extracted from EAGLE. We measure common kinematic and photometric observables from these mocks, which are then used along with their ground truth 3D measurements, to train, validate and test the machine learning model. A schematic diagram of our pipeline is illustrated in Fig. 1.

Below, we begin by describing the methodology of computing the 3D measurements (Section 2.1) and constructing the mock observations (Section 2.2). We discuss how we have controlled for bias in our training data (Section 2.3) and how the individual kinematic and photometric properties have been calculated for the training data set (Section 2.4). We then outline the selection of parameters to be used for the MDN that have been chosen using a principal component analysis (PCA) technique (Section 2.5).

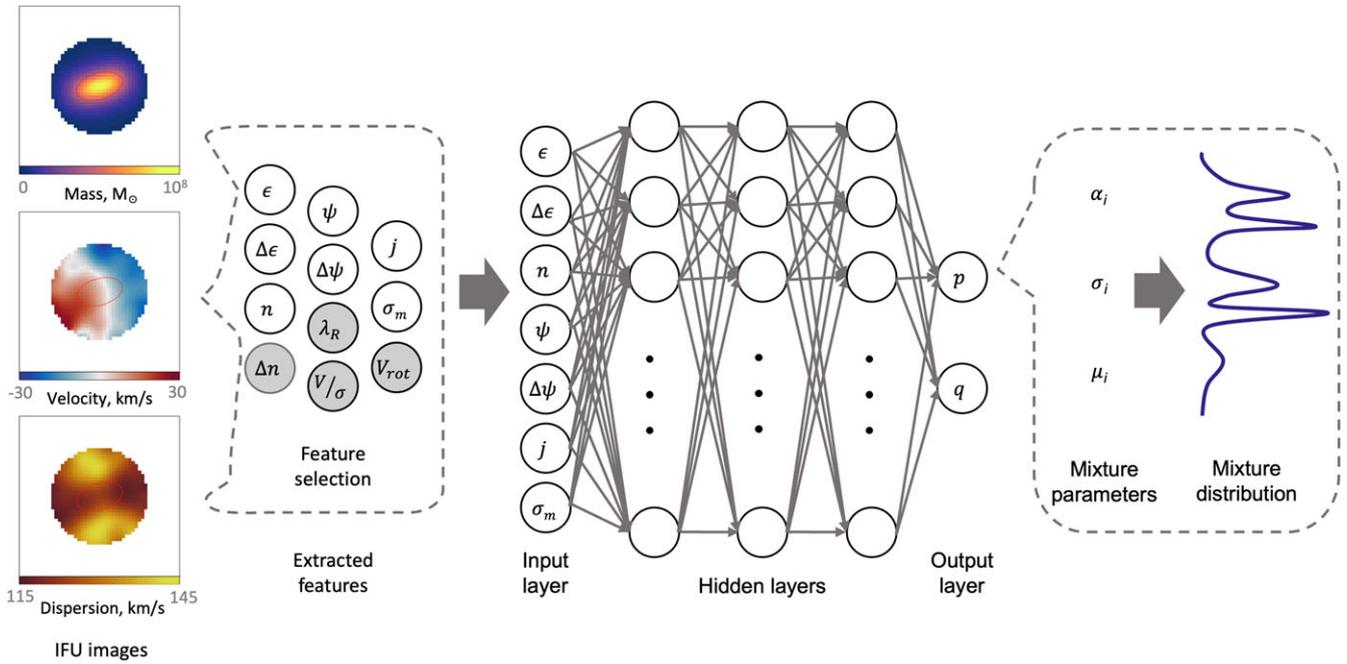
### 2.1 3D measurements

We perform 3D shape measurements following the method described in Bassett & Foster (2019), which is based on Bak & Statler (2000). Under the assumption that galaxies are well described by a 3D ellipsoid, we determine the axes lengths of each galaxy by calculating the eigenvectors and eigenvalues of the reduced inertia tensor (similar to e.g. Allgood *et al.* 2006; Li *et al.* 2018b). The reduced inertia tensor,  $I$ , is defined as:

$$I_{ij} \equiv \sum_n \frac{x_{i,n}x_{j,n}}{\tilde{r}_n}, \quad (3)$$

where the tensor sum is performed for each combination of orthogonal axis directions,  $ij$ , and  $\tilde{r}_n$  is the 3D, ellipsoidal radius measured for stellar particle  $n$ ,

$$\tilde{r}_n = \sqrt{x_n^2 + (y_n/p)^2 + (z_n/q)^2}, \quad (4)$$



**Figure 1.** Flowchart of the intrinsic shape determination pipeline. For each simulated galaxy, mock IFS images are constructed from which the kinematic and photometric features are extracted. Principal component analysis is applied for feature selection to choose a number of important features (those not selected are in grey). These are then fed to the mixture density network with 3 dense hidden layers of 128 nodes each. In the last layer, the MDN outputs a linear combination of Gaussian mixture parameters given by the weights  $\alpha_i$ , standard deviations  $\sigma_i$ , and means  $\mu_i$  to predict the  $p$  and  $q$  distributions.

where the major, intermediate, and minor axes are aligned with the  $x$ ,  $y$ , and  $z$  directions, respectively. The axis ratios  $p = b/a$  and  $q = c/a$  are computed as the ratios of the corresponding eigenvalues.

In practice, we must exclude particles at large radii, which can skew measurements of  $p$  and  $q$ . We therefore measure the shapes of galaxies only using particles within the ellipsoidal half mass radius,  $\tilde{r}_e$ , for consistency with previous works (e.g. Bassett & Foster 2019) and for fair comparison with the radius at which the kinematic and photometric properties will be measured within the mock observations.

This requires us to measure the 3D shapes as an iterative process because the ellipsoidal half mass radius must be determined prior to measuring the eigenvalues of the reduced inertia tensor. As an initial guess, we assume that each galaxy is spheroidal and measure the galaxy shape within the spherical half mass radius. We determine the new set of particles within the ellipsoidal half mass radius defined by this shape measurement. This process is repeated until the shape measurement has converged, giving the final values of  $p$  and  $q$  for each galaxy. The resulting range of galaxy shapes, as measured for the systems extracted from EAGLE, are shown in Fig. 2.

## 2.2 EAGLE mock observations

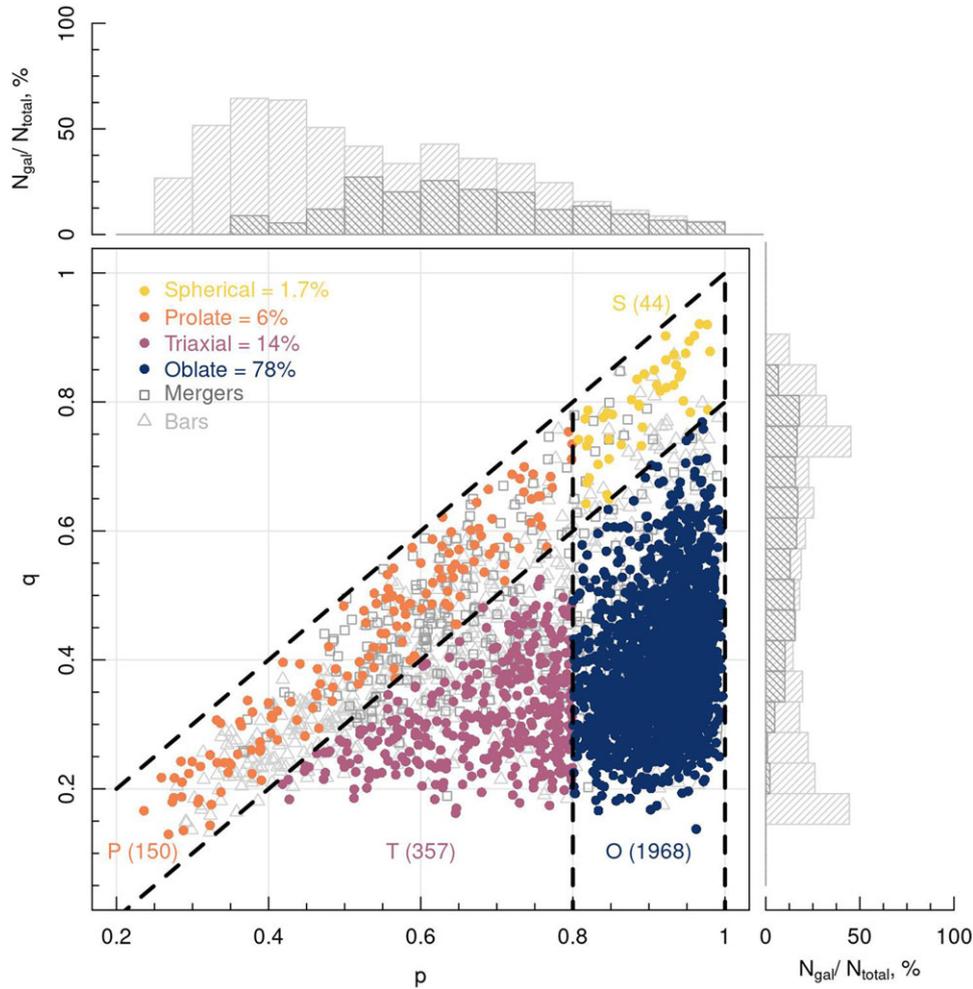
The EAGLE simulation suite is a set of hydrodynamical cosmological simulations designed to explore the necessary physical ingredients to form the galaxies we observe in the Universe today. A full description of these simulations can be found in Crain et al. (2015) and Schaye et al. (2015). Individual galaxies have been extracted from snapshot 28 ( $z=0$ ) of the publicly available RefL0100N1504 simulation box, a medium-resolution run of the ‘reference’ model within a  $100^3$  comoving Mpc box. Using the EAGLE database (McAlpine et al. 2016), we have selected a

sample of 3 638 galaxies above a stellar mass limit of  $M_* \geq 10^{10} M_{\odot}$ . For each of these galaxies, we begin by measuring their shapes as described in Section 2.1.

The sample of galaxies is trimmed to remove any objects problematic for the algorithm design. Galaxies marked as ‘spurious’ or those which contain less than 50% of the total stellar mass of the galaxy within a 50 kpc-radius sphere are also removed, leaving 3 626 objects. Of these, we further restrict our sample to systems that have not experienced a major or minor merger (where we consider merger ratios  $\geq 0.1$ ) in the previous 5 Gyr of lookback time in order to avoid objects with significant disturbed features (Morales et al. 2018), leaving us with a sample of 3 199 galaxies. As in Bassett & Foster (2019), we find that disturbed systems predominantly have low measured  $q$  parameters and evenly distributed  $p$  values, as shown in Fig. 2, and hence are more often interpreted as overly flat objects. Finally, following the methodology of Bassett & Foster (2019), we remove any barred galaxies from the sample as the shapes of these systems within  $1 R_{\text{eff}}$  will often be mistaken for prolate structures. To find these systems, we measure the radial shape profiles for each galaxy and flag any systems whose shape changes from prolate or spherical to triaxial with increasing radius. Flagged galaxies are then visually inspected for any strong internal stellar features, such as bars or spiral arms. This leaves us with a sample of 2 519 galaxies. The distributions of these barred and disturbed systems are demonstrated in the histograms in Fig. 2.

Mock IFS observations of the remaining EAGLE galaxies are performed using the publicly available code SIMSPIN.<sup>a</sup> While the methodology behind this code is outlined in Harborne et al. (2023), we summarise the necessary details followed to construct each mock IFS observation here.

<sup>a</sup>SIMSPIN is an open-source R-package at <https://github.com/kateharborne/SimSpin>.



**Figure 2.** The distribution of galaxy shapes measured from the RefL0100N1504 box of the EAGLE simulation. In dark grey squares we show galaxies that have undergone a major or minor merger within the last 5 Gyr, which we class as ‘disturbed’. Light gray outline triangles show systems with significant bar structures. The histograms show the distribution of these barred and merger systems in  $p$  and  $q$  in light and dark grey respectively. Coloured circles represent galaxies that we have selected for our investigation. Each colour demonstrates the shape of the system, with spherical objects in yellow, prolate objects in orange, triaxial objects in pink and oblate objects in blue.

We begin by initialising an observing telescope with properties equivalent to the SAMI survey (Croom *et al.* 2012). We have selected SAMI for comparison with the work presented in Foster *et al.* (2017) and Bassett & Foster (2019). This is a predefined telescope type within the SIMSPIN code. The spatial pixels of a SAMI observation are  $0.5''$  in width, with a equivalent velocity bin width of  $1.04 \text{ \AA}$  (Green *et al.* 2017). The field-of-view of a SAMI observation is circular with a diameter of  $15''$ . For kinematics, observations from the blue arm of the AAOmega Spectrograph are used. The line spread function for this is well-approximated by a Gaussian of size  $2.65 \text{ \AA}$  measured at full-width at half-maximum (van de Sande *et al.* 2017).

With the telescope defined, each galaxy is placed at a specified distance from the observer. A mass-weighted LOSVD is built for each spaxel dependent on the underlying velocity distribution of the particles at that pixel position. This information is then used to generate mock images of the projected mass, line-of-sight velocity and velocity dispersion in the style of a SAMI observation. From these maps, we measure a series of parameters in order to train the model.

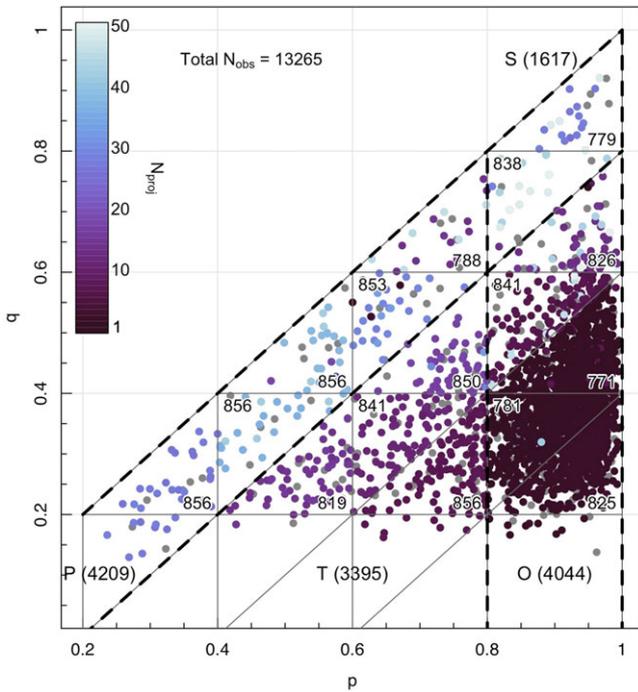
To complement the kinematic maps, we also produce a series of mock  $r$ -band images using SimSpin. This is done by associating

stellar population templates with star particles of a given age and metallicity and computing the received flux through an  $r$ -band filter. These images are constructed with a spatial resolution equivalent to KiDS with  $0.2''/\text{pixel}$ . Photometric properties, such as the observed half-light radii, ellipticity and Sérsic index, are measured from each of these images.

### 2.3 Building the unbiased training set

Using the defined sample of 2 519 unique galaxies extracted from EAGLE, we take a number of observations for each system from different orientation angles using this ‘mock’ SAMI instrument in order to train and validate our algorithm. There are 13 265 independent observations, of which 85 percent are used in the training set and the remaining 15 percent are reserved for the validation set. A further 9 599 observations of EAGLE galaxies are later used to test the algorithm. No unique galaxy appears in more than one of these sets (training, validation, or testing) in order to avoid confirmation bias.

In each case, we wish to uniformly sample the  $p - q$  parameter space so as not to bias our machine learning algorithm to one specific shape. If we provide a non-uniform distribution of galaxy



**Figure 3.** Demonstrating the equal sampling of  $p - q$  space within the training set. Each point shows an individual EAGLE galaxy within the full sample. Coloured points show galaxies selected for the training and validation sets, while grey points demonstrate galaxies that have been left for the testing phase. The colour of each point denotes the number of times that galaxy is observed in order to keep that  $p - q$  region equally sampled. The number in the corner of each  $p - q$  region demonstrates the total number of observations within that region.

shapes within the training set, the algorithm is likely to preferentially assign a new observation to an over-sampled parameter due to its prevalence in the training data, rather than due to its similarity with a specific shape’s kinematic properties. Hence, we must ensure that the training set includes a uniformly distributed density of galaxies in the  $p - q$  parameter space.

However, galaxy shapes are not equally prevalent in the Universe, or our simulations, as evident from Fig. 1. The majority of systems are oblate, while spherical systems are comparatively rare. In order to uniformly sample the parameter space, we have divided the  $p - q$  space into 16 regions. We take repeat observations from different angles of galaxies in lower density  $p - q$  regions in order to have a large enough data set to sufficiently train the algorithm. No two observations of the same galaxy are from the same angle.

In Fig. 3, we demonstrate how we have sampled each region within  $p - q$  space uniformly. It is noted that the  $p - q$  space regions are equal sized triangles in the region  $q > 0.4$  only, as in the lower  $q < 0.2$  regime, particularly for prolate objects, we have significantly fewer available galaxies to sample. Objects that exist in the region  $q < 0.2$  have been combined with a neighbouring region above as demonstrated in Fig. 3.

Further, we need to consider how the kinematic and photometric properties we measure for a given 3D shape is impacted by the chosen mock observational set up. With a simulated system, we have the flexibility to re-orient the system in angle, projected distance and we can change the level of atmospheric seeing, in hope of recreating a reasonable range of mocks that reflect a real survey. We must take care that these choices do not introduce biases

into the end result. In a follow up paper, we will demonstrate the importance of this unbiased sampling to produce a successful machine learning algorithm that uses kinematics from mock observations as input. For the sake of brevity, we present the optimal parameters over which to sample here.

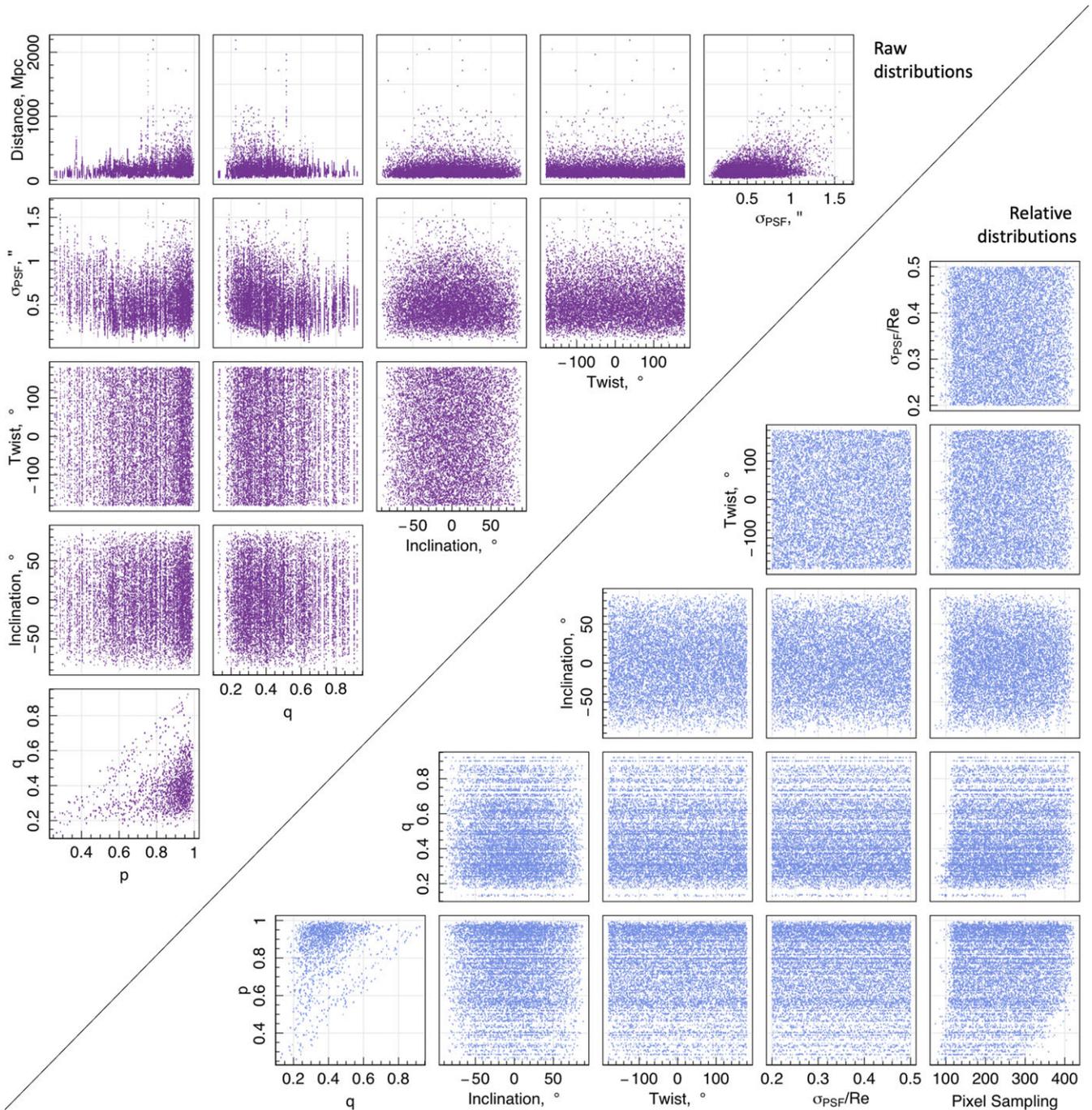
We create an unbiased kinematic training set by controlling ‘raw’ observational parameters including: (1) the orientation from which the galaxy is viewed, (2) spatial scales (adjusted by ‘moving’ simulated galaxies closer and further away), and (3) the level of blurring caused by seeing conditions. These raw observation properties are cross-correlated with the observed kinematics and one another, such that it is necessary to perform multiple observations in a series of ordered steps to produce one unbiased element of training data.

It is important that these cross-correlated properties, the ‘relative’ observational parameters, are fairly sampled. Of course, within observational surveys, such properties cannot be controlled making it difficult to train such an algorithm with observations alone. In reality, the ground truth properties are also expensive and difficult to acquire. For this reason, our simulations and SIMSPIN are incredibly useful.

We select these parameters within reasonable limits based on the SAMI survey. By ‘reasonable limits’ we mean that we restrict our parameter space to values that would be considered reliable by the survey, rather than the full possible range of the instrument.

**Projected orientation** The observing angle for each run is randomly selected from the uniformly sampled surface of a sphere within each  $p - q$  region. No unique galaxy will be observed from the same observation angle more than once. Within SIMSPIN, this is controlled using two parameters: the *inclination* about the  $x$ -axis of the projection, and the *twist* about the  $y$ -axis of the projection. We ensure that the combination of these parameters uniformly samples the surface of a sphere such that we are not biased towards a single observing angle in regions where more re-observations of single galaxies are necessary. This also ensures that all possible 2D projections are evenly sampled within the training set.

**Spatial scales** Physical spatial scales, i.e. the width of each spatial pixel in units of kpc, are varied by moving each galaxy to different distances. This distance is selected to sample a uniform distribution of pixels per measurement radius (otherwise called the ‘spatial pixel scale’). This is a ‘relative’ observational property that we need to consider in relation to the projected angle above. If we were to require a uniform distribution of distances (the ‘raw’ property) for each galaxy shape, we would inadvertently cause the distribution of spatial pixel scales to be lower for oblate systems (which, when viewed edge on have far fewer pixels within their measurement radius than an edge-on elliptical). The spatial pixel scale is known to correlate with the uncertainty of observed kinematic parameters such as the spin parameter  $\lambda_R$  (see Appendix C in Harborne et al. 2020). We know that galaxy shape is correlated with the effects of seeing conditions on the observed kinematics, and the size of this effect in turn is dependent on the number of pixels within the measurement radius. We do not want to train our algorithm to learn that oblate systems are characterised solely in this way, else any observation with poor seeing and low spatial pixel sampling may be labelled ‘oblate’. Hence, once a projected orientation is chosen for a galaxy, we adjust the projected distance to that object to recover a given spatial pixel scale selected from a uniform distribution.



**Figure 4.** (Top left) Considering the *raw* distributions of tuneable observation properties controlled in each mock observation using a corner plot. The relationship between each property is demonstrated in purple. By ‘raw’ we mean the values modified per observation i.e. viewing angle, size of the PSF and projected distance to the object. (Bottom right) Considering the *relative* distributions of the important tuneable observation properties to ensure mock observations are uniform in the important ratios, as shown in the blue corner plot. These ratios, i.e. the size of the PSF relative to the size of the object, and the number of pixels within the measurement radius, are important for measuring observable kinematics to produce an unbiased training set. The approximately uniform distribution shown in blue demonstrates that our sample selection is not biasing our algorithm results.

In the lower right-hand corner of Fig. 4, there is a slight remaining trend with pixel sampling and very low  $p$ . This is due to the field of view limit of the SAMI observations. There is a limit to how large a flat, edge on disc can be projected within a circular aperture. However, to remedy this, we have successfully flattened the distribution of pixel sampling with respect to the observational seeing conditions. As a result, the effect on the observed kinematic parameters is minimised.

**Seeing conditions** Seeing conditions are uniformly sampled between 0.2 and  $0.5 \sigma_{\text{PSF}}/R_{\text{maj}}$ , where we have chosen to define the size of the PSF relative to the measurement radius of the observation across a similar range to SAMI observations. Again, it has been shown in a number of works that the impact of seeing conditions on the observed kinematics of a galaxy is significant and varies with galaxy shape (Harborne et al. 2020; Graham et al. 2018; Greene et al. 2018; van de Sande et al. 2017; D’Eugenio et al. 2013).

We use a Latin Hypercube Sampling (LHS)<sup>b</sup> procedure (Stein 1987; McKay et al. 1979) to ensure that we do not introduce a correlation between these properties. We demonstrate the success of this approach in the relative distributions on the right-hand side of Fig. 4.

Of course, while this careful selection produces a uniform distribution in the chosen ‘relative’ parameters of importance to kinematic parameter recovery, this does result in skewed distributions for certain ‘raw’ observed parameters, as shown by the distributions on the left-hand side in Fig. 4. We must put flatter objects with lower  $p$  values closer to the observer (as shown by *Distance*, *Mpc* in the final column). These closer objects see higher  $\sigma_{PSF}$  values as a result.

However, it is the relative relationship between  $\sigma_{PSF}$  and the half-light radius of the galaxy that has impact on the measured kinematics and so it is this relative parameter that we must uniformly sample. Throughout the course of developing this MDN, several runs with different input training sets demonstrated how these biases propagate through the predictions of the network. This final set of relative distribution controls were found to best reflect reality of the shape distributions.

## 2.4 Computing the training parameters

Using these unbiased mock IFS observations, we measure common kinematic and photometric parameters used in IFS surveys. For each parameter, we demonstrate the ability of this term to distinguish unique 3D shapes using a histogram in Fig. 5. The reason for selection and method of measuring each of these parameters is discussed below. While this is a comprehensive list of all the parameters explored, we note that the list of features used within the MDN model is trimmed down using PCA. Only a selected number of key parameters are eventually used that best describe the features required to recover the intrinsic 3D shape distribution. A list of the explored parameters as well as those selected for the MDN is summarised in Table 1.

### 2.4.1 Ellipticity, $\epsilon$

The ellipticity,  $\epsilon$ , of a galaxy in projection is the first parameter considered. Franx et al. (1991) demonstrated how the observed ellipticity of a projected density distribution depends primarily on the intrinsic underlying galaxy shape and the viewing angle. For each intrinsic 3D shape, however, the mapping of a projected  $\epsilon$  is not unique. This is seen in the first panel of Fig. 5 where all 3D shapes may be projected to low apparent ellipticities ( $\epsilon \leq 0.3$ ). Conversely, spherical systems cannot be projected to high ellipticities regardless of the inclination angle. This suggests that apparent ellipticity is likely to be a useful feature for certain regions of the  $p$ - $q$  parameter space.

To measure  $\epsilon$ , each mock KiDS  $r$ -band observation is fit using PROFOUND (Robotham et al. 2018). We extract isophotal ellipses for each projection. Photometric and kinematic measurements are computed within the average ellipse that contains 50% of the total light of the galaxy, at which point the ellipticity of the object is also quoted. The ellipticity of this region is measured,

$$\epsilon = 1 - (b/a), \quad (5)$$

where the axis ratio ( $b/a$ ) is computed by diagonalising the inertia tensor for all pixels within the half-light isophotal ellipse. The

<sup>b</sup>We use the `lhs` R-package for this process, which can be found at <https://bertcarnell.github.io/lhs/index.html>.

error on this parameter,  $\Delta\epsilon$ , is computed by taking the standard deviation of the ellipticity of concentric isophotes within 40–60% of the total flux, as in Bassett & Foster (2019). This is done for every inclination at which the galaxy is measured, hence why a disc may appear with a low ellipticity when viewed nearly face-on.

It is worth noting that we see very few highly elliptical objects from this sample of EAGLE galaxies. This is not unexpected. As shown in Lagos et al. (2018) and Sande et al. (2019), EAGLE does not produce the observed high ellipticity objects we may expect. This is due in part to the temperature cooling floor imposed in cosmological hydrodynamical simulations that sets a minimum disc scale height of 1 kpc. There is also the effect of numerical heating (Ludlow et al. 2019, 2021; Wilkinson et al. 2023), which further increases the dispersion perpendicular to the plane of discs within the model, causing more disc-like objects to appear more elliptical in shape and kinematics. We can see that this is evident in the distribution of ellipticities. However, there is still a relative difference between each shape, so although the values may not reach as high as we may expect, their relative distribution seems reasonable.

### 2.4.2 Sérsic index, $n$

As a commonly measured structural parameter, we consider the Sérsic index,  $n$ , in the set of training parameters. The Sérsic index describes the slope of the light profile of a galaxy as fitted with a Sérsic profile (Sérsic 1963). Discy systems correspond to low Sérsic indices ( $0.3 \leq n \leq 1.5$ ), while elliptical systems have higher values  $n \sim 4$ . Despite its clear link to morphology and galaxy structure, Fig. 5 suggests that the power of Sérsic index in distinguishing 3D shapes is limited as we find the majority of shapes occurring in  $0.3 \leq n \leq 4$ . However, there is still a relative difference visible where spherical objects are seen to preferentially exist at the upper end of the limit and oblate objects dominate at the lower end of the limit, as we might expect.

The Sérsic profile is defined as:

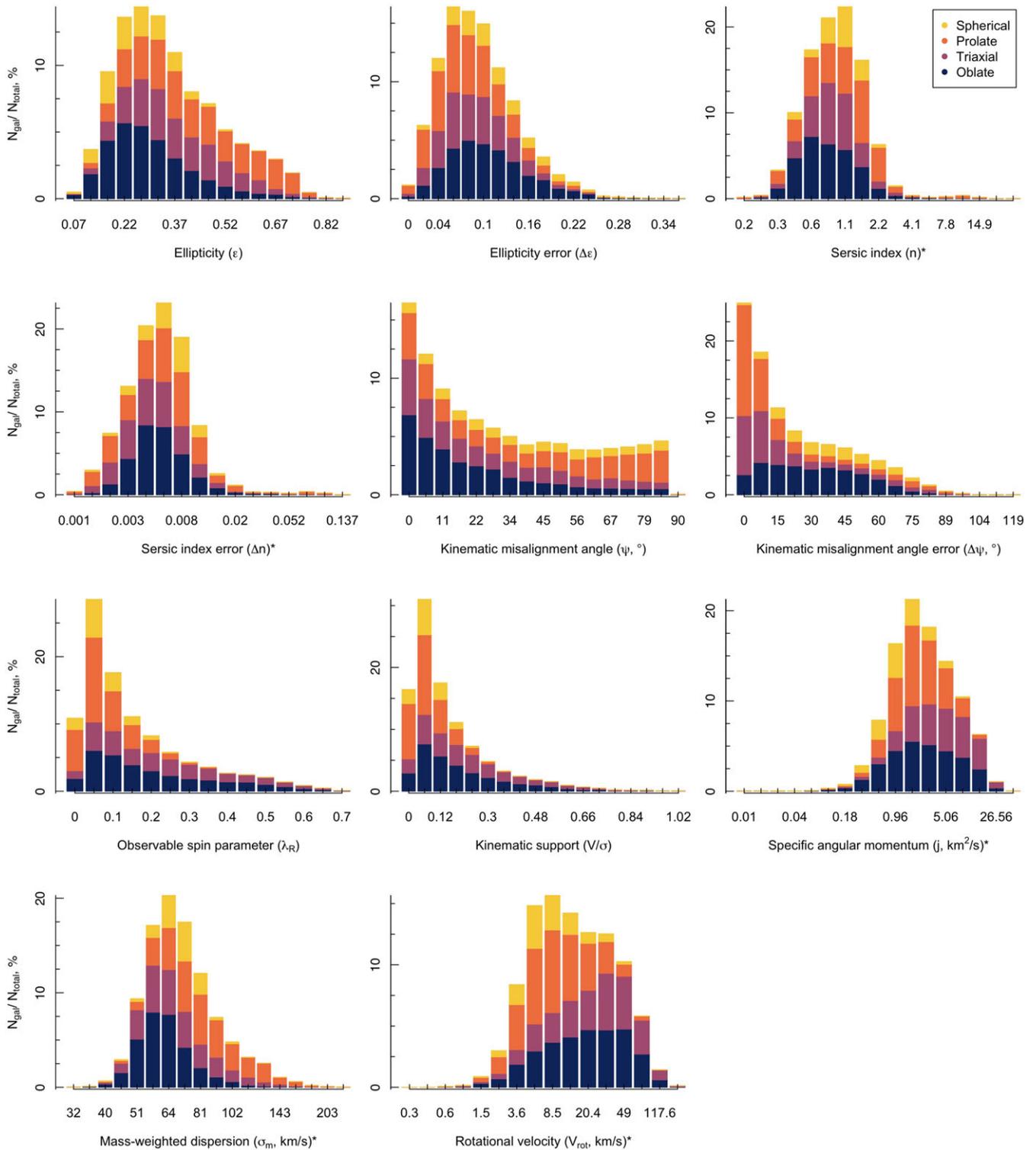
$$I(R) = I_e \exp \left\{ -b_n \left[ \left( \frac{R_m}{R_e} \right)^{1/n} - 1 \right] \right\}, \quad (6)$$

where the  $I_e$  is the profile intensity measured at the half-light radius,  $R_e$ . The Sérsic index,  $n$ , is a parameter that describes the shape of the light profile, while  $b_n$  is used to ensure that, for a given Sérsic index  $n$ , the correct integration properties occur at the half-light radius (i.e. that the function integrates to 0.5 for a given  $n$ ). Finally,  $R_m$  is the ‘modified’ radius at which an isophote is defined. Within the ProFit code used to compute isophotal ellipses, this has a number of functional forms depending on whether the fitted isophotes are circular, elliptical or boxy. For more details on each of the definitions, please refer to Section 2.1.1 of Robotham et al. (2017). We use elliptical isophotes for this analysis.

The associated error on  $n$  is computed as the standard deviation of the residual between the fitted Sérsic profile and the surface brightness profile.

### 2.4.3 Kinematic misalignment angle, $\psi$

As suggested by a number of works (Franx et al. 1991; Weijmans et al. 2014; Foster et al. 2017; Ene et al. 2018; Bassett & Foster 2019), the kinematic misalignment angle,  $\psi$ , should also enable us to map back to the underlying 3D shape of the galaxy in question. Again, using  $\psi$  alone is insufficient to invert back to an underlying 3D shape directly, as the kinematic misalignment is also dependent on the intrinsic axis of rotation. For a triaxial galaxy, this rotation axis can lie anywhere in a plane that connects the longest



**Figure 5.** Histograms showing how different intrinsic shapes of EAGLE galaxies within our training data populate each observable parameter. In each case, the spherical systems are shown in yellow, prolate systems in orange, triaxial systems in purple and oblate systems in blue. The overall height of the bar shows the distribution of each kinematic parameter within the full training set. The coloured regions then demonstrate the percentage of each bar that is made up of each intrinsic shape. Starred (\*) axis labels have been divided into equally-sized log10 bins to more clearly delineate between the groups, though bar labels are shown as the raw values for clarity. This plot demonstrates that, in none of the single measurements can we cleanly distinguish between the intrinsic shapes directly. This justifies the machine learning approach.

**Table 1.** List of investigated galaxy parameters in column 2 and their notation in column 1. A short description in column 3 and relevant sections where each parameter is described in the text are listed in column 4. Parameters that are selected as input features to the mixture density network (MDN) indicated in column 5.

Notation (1)	Feature (2)	Short description (3)	Section (4)	Used in MDN (5)
$\epsilon$	Ellipticity	Measured from each projected $r$ -band luminosity map at the isophotal ellipse containing half the total flux of the galaxy	Section 2.4.1	✓
$\Delta\epsilon$	Ellipticity error	Associated ellipticity error measured using the standard deviation between the ellipticities of the isophotal ellipses containing 40–60% of the total galaxy flux	Section 2.4.1	✓
$n$	Sérsic index	Measured in the observed $r$ -band light	Section 2.4.2	✓
$\Delta n$	Sérsic index error	Associated error on the fitted Sérsic index where the error is computed as the standard deviation of the residuals to the fit	Section 2.4.2	✓
$\psi$	Kinematic misalignment angle	Measured between the principle axis of rotation and the principle axis of the light distribution	Section 2.4.3	✓
$\Delta\psi$	Kinematic misalignment angle error	Associated error on the kinematic misalignment angle	Section 2.4.3	✓
$\lambda_R$	Observable spin parameter	Measured within 1 Re that describes the relative importance of rotation over dispersion as a function of radius in supporting a galaxy's structure	Section 2.4.4	
$V/\sigma$	Kinematic support	Measured within 1 Re of the relative importance of rotation over dispersion in supporting a galaxy's structure	Section 2.4.5	
$j$	Specific angular momentum	Measured within 1 Re	Section 2.4.6	✓
$\sigma_m$	Mass-weighted average velocity dispersion	Measured within 1 Re	Section 2.4.7	✓
$V_{\text{rot}}$	Observed rotational velocity	Measured along the principle kinematic axis	Section 2.4.8	

and shortest axis of the system. In the second row of Fig. 5, we consider the distribution of  $\psi$  and its associated error with different intrinsic shape classifications for observations in our training set. It is clear that triaxial objects make up a similar percentage of every bin. Despite this, there are a number of clear trends with oblate and prolate objects predominantly occupying the low  $\psi$  and high  $\psi$  bins respectively, providing some level of distinguishing power. Using machine learning, we can explore the relative importance of this parameter with respect to other photometric and kinematic features.

To compute  $\psi$ , we use the following equation from Franx et al. (1991):

$$\sin(\psi) = \left| \sin(PA_{\text{phot}} - PA_{\text{kin}}) \right|. \quad (7)$$

The photometric position angle,  $PA_{\text{phot}}$ , is measured as the orientation of the major axis of the isophotal ellipse in degrees from vertical. Similarly, the error on this value is computed using the standard deviation of isophotes' major axes angles for the isophotes 40–60% of the total light.

The kinematic position angle,  $PA_{\text{kin}}$ , describes the stellar angular rotation vector along which the rotational velocity is maximised. To compute  $PA_{\text{kin}}$ , we use the Kinemetry technique outlined in Appendix C of Krajnovic et al. (2006).

#### 2.4.4 Observable spin parameter, $\lambda_R$

The observable spin parameter  $\lambda_R$  was defined by Emsellem et al. (2007) to describe the level of coherent rotation in galaxies observed as part of the SAURON survey (Bacon et al. 2001).  $\lambda_R$  is now commonly measured by IFS survey teams as a method of distinguishing fast rotators from slow rotators within a  $\lambda_R - \epsilon$  plane. Given the underlying relation between the intrinsic shape and rotational support, as well as the commonality of this measure, it is an obvious parameter to include within the machine learning training. We can see strong evidence for this in the first panel of the third row of Fig. 2.4. While there are relatively fewer high-spin

objects observed in the EAGLE training set, we can see that any objects above  $\lambda_R$  of 0.3 are either oblate or triaxial systems.

We use the following equation to define the spin parameter within the half-light isophotal ellipse defined in Section 2.4.1:

$$\lambda_R = \frac{\sum_i^N F_i R_i |V_i|}{\sum_i^N M_i R_i \sqrt{V_i^2 + \sigma_i^2}}, \quad (8)$$

where  $F_i$  is the flux of stellar particles contained per pixel,  $i$ ,  $V_i$  is the LOS velocity,  $\sigma_i$  is the LOS velocity dispersion and  $R_i$  is the ellipsoidal radius corresponding to the semi-major axis of the ellipse at this  $i$ th pixel location. Only pixels contained within the half-light radius (described in Section 2.4.1) are included within this calculation. We note that this ellipsoidal radius definition is slightly different to the original  $\lambda_R$  definition in Emsellem et al. (2007), which uses circularised radial weighting. The choice of an elliptical aperture is made to be consistent with the values measured by the SAMI survey (Sande et al. 2017).

#### 2.4.5 Kinematic support, $V/\sigma$

Along a similar vein,  $V/\sigma$  is another common dynamical parameter used to gauge the balance of rotational vs. pressure support of galaxies (Illingworth 1977; Binney 1978; Davies et al. 1983). With the advent of IFS, this value was reformulated by Binney (2005) using the tensor virial theorem in order to relate kinematics back to the intrinsic flattening. For this reason, it is included within the training parameters.

We use the light-weighted kinematics to compute  $V/\sigma$  which is given by the equation:

$$V/\sigma = \sqrt{\frac{\sum_i^N F_i V_i^2}{\sum_i^N M_i \sigma_i^2}}, \quad (9)$$

where, as before,  $F_i$  is the stellar flux contained per pixel,  $i$ ,  $V_i$  is the LOS velocity at that pixel and  $\sigma_i$  is the LOS velocity dispersion at that pixel. As with the measurement of  $\lambda_R$ , only pixels within

the measurement radius of the half-light ellipse are used in this computation. When considering the distribution of this parameter across different galaxy shapes, in the central panel of the third row in Fig. 5, we find qualitatively similar distributions to the  $\lambda_R$  parameter.

#### 2.4.6 Specific angular momentum, $j$

In Bassett & Foster (2019), it was found that the intrinsic flattening of a galaxy was most strongly correlated with the stellar specific angular momentum,  $j$ . Considering the distribution of galaxy shapes in the right panel of the third row in Fig. 5, we can already see that high  $j$  galaxies do exhibit flatter structures for our training galaxies taken from the EAGLE simulation.

We compute the specific angular momentum,  $j$ , using the formula:

$$j = \frac{\sum_i^N F_i R_i |V_i|}{\sum_i^N F_i}, \quad (10)$$

where the terms are the same as those used above. Again, only pixels contained within the half-light radius are included within this calculation.

#### 2.4.7 Mass-weighted velocity dispersion, $\sigma_m$

To then consider the level of dispersion support, in contrast to rotation and  $j$ , we compute the mass-weighted velocity dispersion,  $\sigma_m$  within the half-light isophote. As can be seen in Fig. 5, the first panel in the final row shows increasing preponderance of prolate systems with increasing dispersion.

This parameter is calculated using the LOS-velocity dispersion maps for each observation:

$$\sigma_m = \frac{\sum_i^N M_i \sigma_i}{\sum_i^N M_i}, \quad (11)$$

where here,  $\sigma_i$  is the LOS velocity dispersion and  $M_i$  is the stellar mass per pixel,  $i$ , contained within the half-light radius.

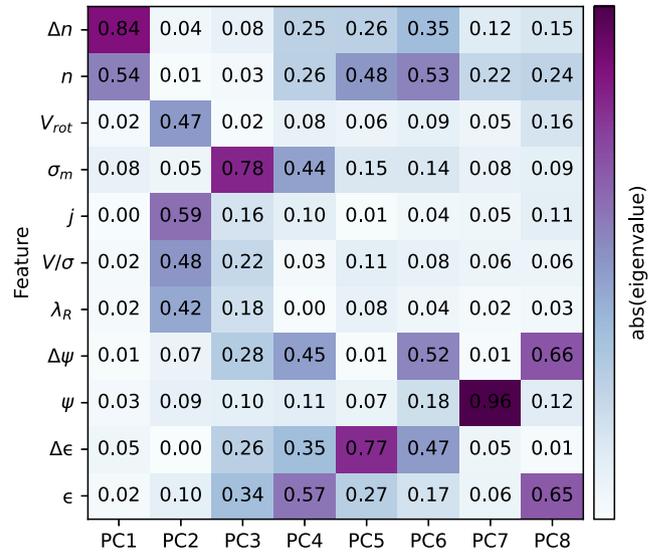
#### 2.4.8 Rotational velocity, $V_{\text{rot}}$

The final flavour of kinematic parameter used is the rotational velocity,  $V_{\text{rot}}$ . This is another way to parameterise the level of rotation support for a system and was commonly recovered for long-slit spectroscopy and radio interferometry measurements (Davies & Birkinshaw 1988; Bak & Statler 2000). To compute this parameter for our galaxies, we use the kinematic position angle returned for the computation of  $\psi$  and realign our galaxy velocity map such that the maximum gradient is horizontal within the image. The central row of pixels from the velocity image is then used to visualise the galaxy rotation curve, with the left and right of the curve normalised and averaged.  $V_{\text{rot}}$  is then quoted at the turnover of the rotation curve.

We can see from the final panel in Fig. 5 that oblate and triaxial systems often occupy these higher  $V_{\text{rot}}$  regions.

### 2.5 Selecting input features

From the initial set of 11 galaxy structural and kinematic parameters listed in Table 1, we further perform feature selection using principal component analysis (PCA; Pearson 1901) on the training data set. Before applying PCA, all parameters of the training data are scaled based on the median and interquartile range between the 25th and 75th quantiles as a preprocessing step. This PCA step is useful in determining which parameters hold the most constraining power and which sets may exhibit redundancies



**Figure 6.** Feature selection of galaxy kinematic parameters (see Table 1 for notation) using principal component analysis. Absolute eigenvalues of the associated principal component (PC) are labelled and coloured, where a value of 1 indicates the strongest possible contribution with darker gradient.

when tackling intrinsic shapes. The resulting eigenvalues from the PCA quantifies the relative contributions of the various parameters to the main principal components (PCs). The larger the absolute value of the corresponding eigenvalue, the more influence the parameter has on the PC.

Fig. 6 shows the distribution of PCs along with weightings of each parameter on the PC and corresponding eigenvalues. We select up to 8 PCs, which explain  $\sim 99\%$  of the variance in the training data. By performing this step, we can filter out redundant parameters and select those that are significant, which can lead to the overall improvement in the performance of the predictive model (see Appendix A for comparative result without performing feature selection). For each PC, the feature with the highest absolute eigenvalue is used as input feature to train the neural network model described in the next section. This corresponds to 8 features (*Used in MDN* column in Table 1), namely ellipticity  $\epsilon$ , ellipticity error  $\Delta\epsilon$ , kinematic misalignment angle  $\psi$ , kinematic misalignment angle error  $\Delta\psi$ , specific angular momentum  $j$ , mass-weighted average velocity dispersion  $\sigma_m$ , Sérsic index  $n$ , and Sérsic index error  $\Delta n$ .

## 3. Recover $p$ - $q$ distribution using machine learning

### 3.1 Building the mixture density network (MDN)

In order to recover  $p$  and  $q$  as probability distributions, we build a mixture density network (MDN; Bishop 1994) using KERAS (Chollet et al. 2015), an open-source PYTHON package for deep learning, with TENSORFLOW (Abadi et al. 2015) backend. MDN combines a deep or fully connected neural network with mixture of distributions. It is a feed-forward neural network that maps a set of input features,  $\vec{x}$ , to generate output,  $y$ , of mixture models (McLachlan & Basford 1988). Generally, the mixture distribution is represented by a Gaussian Mixture Model (GMM). The conditional probability density function can be expressed as:

$$p(y|\vec{x}) = \sum_{i=1}^m \alpha_i(\vec{x}) \mathcal{N}(y; \mu_i(\vec{x}), \sigma_i^2(\vec{x})), \quad (12)$$

where  $m$  is the number of mixture components, and  $\alpha_i(\vec{x})$ ,  $\mu_i(\vec{x})$  and  $\sigma_i^2(\vec{x})$  are the corresponding weight, mean and variance of the  $i$ -th component of the mixture.

The network architecture consists of several layers, starting with an input layer, followed by dense hidden layers characterised by a non-linear activation function, and lastly an output layer with GMM, as illustrated in Fig. 1. For the hidden layer, we use three dense layer networks with 128 neurons and rectified linear unit as activation function. Additionally, a dropout layer with 15% drop rate is inserted between each hidden layer to reduce over-fitting.

In the MDN output layer, we find that changing the number of mixture components does not significantly affect the accuracy and set it to be 5. The GMM parameters are derived from the network output vector,  $\vec{z}$  as follows:

$$\alpha_i(\vec{x}) = \frac{\exp(z_i^a)}{\sum_{j=1}^M \exp(z_j^a)}, \quad (13)$$

$$\sigma_i(\vec{x}) = \exp(z_i^s), \quad (14)$$

$$\mu_i(\vec{x}) = z_i^m. \quad (15)$$

The softmax activation function in Equation (13) ensures that the weights are positive and sum to one. An exponential activation is applied in Equation (14) such that the standard deviations are positive. For the means in Equation (15), a linear activation is used. During the training process, the GMM parameters are optimised by minimising the negative logarithmic likelihood of Equation (12) as the loss function using root mean squared propagation. We train the MDN model for 5 000 epochs using the training data set. Early stopping is imposed to terminate the run when there is no significant improvement in the model performance's on the validation set. Finally, the optimised GMM parameters are used to predict the distributions of  $p$  and  $q$  of the test dataset.

### 3.2 Computing evaluation metrics

To evaluate the performance of the model predictions, we compute the commonly used root mean squared error (RMSE) metric. The RMSE is the average of the squares of the difference between the actual value,  $y$ , and predicted value,  $\hat{y}$ , given by:

$$RMSE = \sqrt{\frac{1}{N_{gal}} \sum_{i=1}^{N_{gal}} (\hat{y}_i - y_i)^2}, \quad (16)$$

where  $N_{gal}$  is the number of galaxies in the dataset. The mean of the predicted  $p$  and  $q$  distributions are used to compare to the actual values of  $p$  and  $q$ .

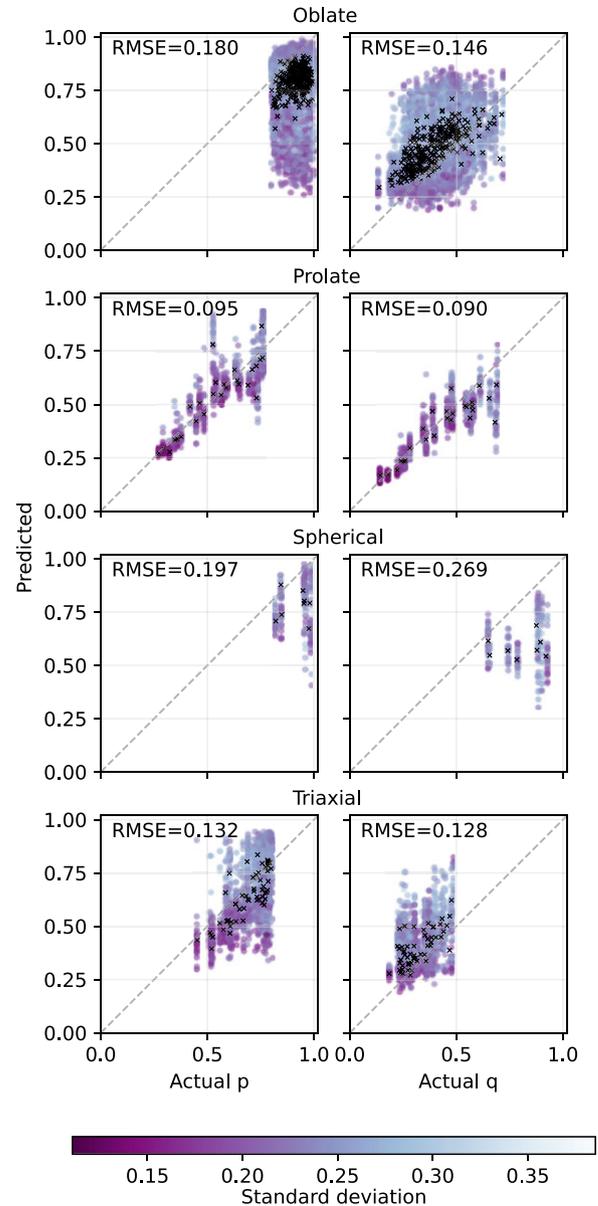
Additionally, we also assess the degree of success and contamination. For each predicted shape class, we compute the negative predicted value (NPV) defined as:

$$NPV = \frac{\text{Number of TRUE negatives}}{\text{Number of negative calls}}, \quad (17)$$

and the positive predicted value (PPV) or precision is defined as:

$$PPV = \frac{\text{Number of TRUE positives}}{\text{Number of positive calls}}. \quad (18)$$

As such, the NPV describes the number of systems that have correctly been identified as NOT a given shape as a fraction of all the systems identified as NOT that shape, i.e. larger negative predicted rates imply a smaller contamination of this shape into other shape

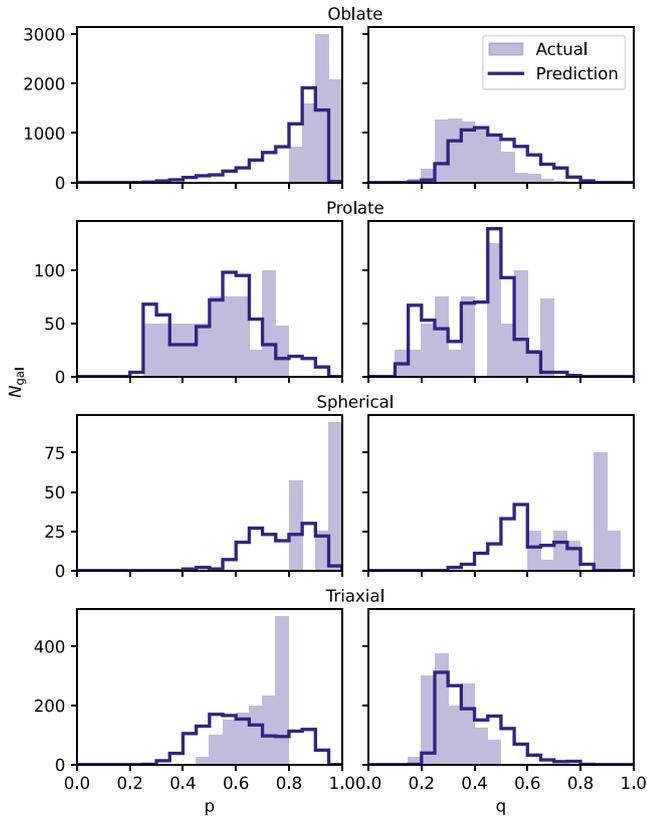


**Figure 7.** Predicted against actual  $p$  and  $q$  for each galaxy shape using mixture density network (MDN) for the test data set. The black crosses represent the average prediction, while circles represent projections of individual galaxies colour coded by the standard deviation from the MDN output. The darker the gradient, the more certain. The prediction error is evaluated by the root mean squared error (RMSE), where lower values represent better agreement. For reference, the identity is shown as a grey dashed line. Although there is a large variation in the standard deviation of individual prediction, in most of the systems, the average of the predictions are close to the actual value.

classes. By contrast, the PPV indicates the number of correctly identified objects of a given shape, i.e. the degree one can trust the shape given by the predicted  $p$  and  $q$  as a function of galaxy class.

## 4. Results

Fig. 7 presents the predicted mean of  $p$  and  $q$  for each galaxy system from the MDN model in comparison to the actual values. An indication of the uncertainty given by the standard deviation output from the Gaussian mixture parameter is shown in coloured points. The corresponding distributions are shown in Fig. 8.

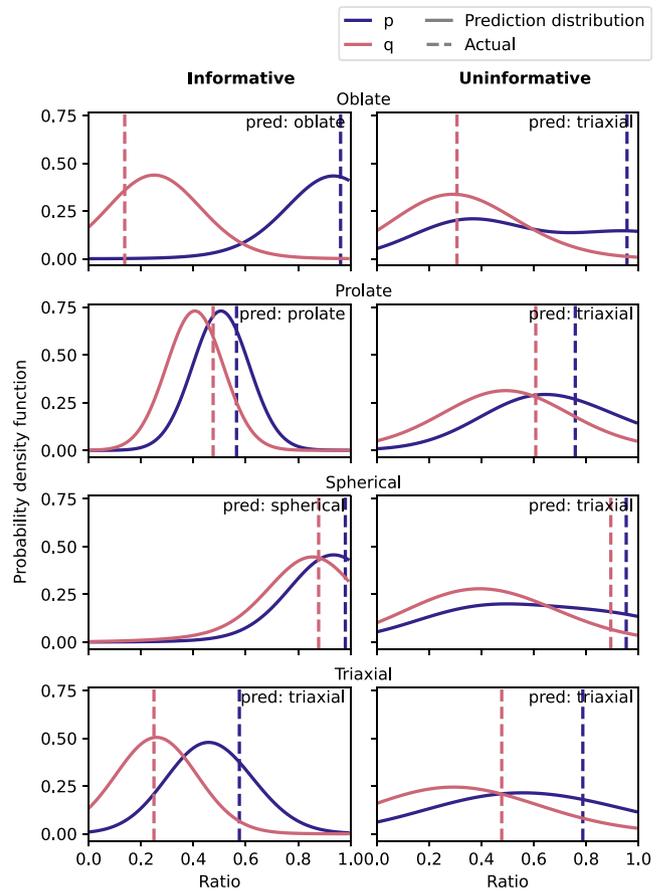


**Figure 8.** Distributions of  $p$  and  $q$  recovered from the mixture density network model for each galaxy shape compared to the actual for the test data set.

For oblate systems, which are the most abundant class objects, the predicted  $p$  values tend to be underestimated, while the  $q$  values tend to be overestimated. The distribution of  $p$  is highly skewed towards low values. Prolate galaxies are well constrained with the recovered  $p$  and  $q$  distributions closely matching those of the actual distributions. Spherical objects, on the other hand, are much harder to predict with both  $p$  and  $q$  predictions often lower than the actual values. Their recovered distributions are also poorly estimated. For triaxial objects,  $p$  is better predicted than  $q$ , whereby the latter is overestimated in most cases. It is also worth noting that there is a large scatter in the standard deviation, which can extend up to  $\sim 0.37$ . However, the predicted means (indicated by the crosses in Fig. 7) remain close to the true values.

Generally, the MDN is able to recover the underlying shape distributions of  $p$  and  $q$  for most of the systems, particularly prolate objects yield the best recovery with the lowest RMSE. This is followed by triaxial, oblate, and lastly spherical objects with the largest RMSE. Comparing the distributions between the actual and predicted for oblate and triaxial systems, even though their distributions are not exact, the MDN does manage to capture the overall shape of the distribution. For example, both actual and predicted  $p$  distributions for the oblate galaxies have similar negatively skewed shaped distribution. The  $q$  distributions for the triaxial galaxies appear to be positively skewed.

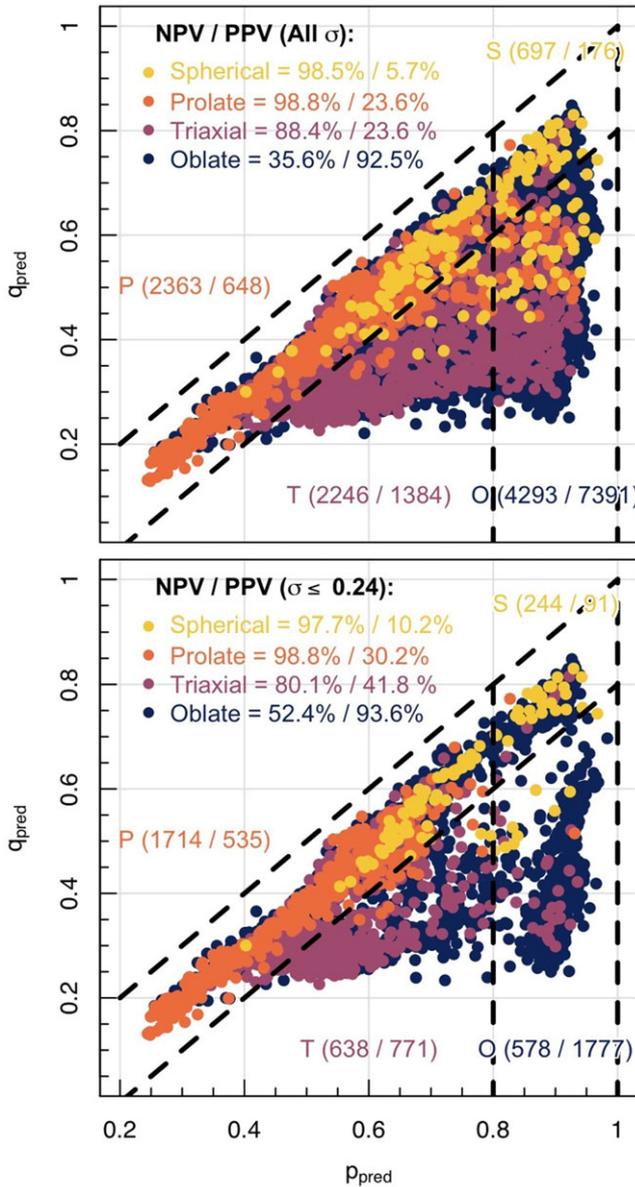
For each prediction, the MDN outputs a probability density function that gives a useful indication of the allowable range of  $p$  and  $q$  values and whose standard deviations serve as estimates of the uncertainties. Accounting for this, we further group them into ‘informative’ and ‘uninformative’ based on the  $p - q$  errors.



**Figure 9.** The recovered  $p - q$  shape probability density function within one standard deviation ( $\sigma$ ) from the mixture density network, showing examples of ‘informative’ (left column) and ‘uninformative’ (right column) predictions for each galaxy shape. ‘Informative’ fits are those with low standard deviation of  $\sigma \leq 0.24$ , while vice versa for ‘uninformative’ fits. The predicted shape is shown on the top right of each panel. The vertical dashed line shows the actual value.

Examples of the recovered shape distributions are portrayed in Fig. 9. For objects that are predicted with low uncertainties of less than the mean of the standard deviation at  $\sigma \leq 0.24$ , we labelled them as ‘informative’, as shown in the left column. The rest, that do not satisfy that criteria, are labelled as ‘uninformative’ in the right column. Generally, it can be seen that when the retrieved distribution of  $p$  is broad (i.e. high standard deviation), the retrieved distribution of  $q$  will also be broad, with Pearson correlation coefficient between the uncertainties of  $p$  and  $q$  of 0.796 ( $p$ -value  $\ll 0.01\%$ ). This suggests that both  $p$  and  $q$  predictions are less reliable if either one has large error.

In Fig. 10, we show the predicted  $p$  and  $q$  values of each observation in our testing set coloured by its true underlying 3D shape for all objects (top panel) and objects in the ‘informative’ group (bottom panel). Generally, the NPV and PPV improve as more certain (lower  $\sigma$ ) predictions are filtered (see Appendix B for trends in NPV/PPV with  $\sigma$ ). Visually, we see that predicted classes are correlated with their true underlying shapes, with contamination specifically from oblate systems being predicted as every other shape class. To quantify the degree of the success and contamination, we also compute the predictive values, namely NPV (Equation (17)) and PPV (Equation (18)), which are quoted as percentages in Table 2.



**Figure 10.** Predicted  $p$  and  $q$  for each galaxy shape using mixture density network for all results (above) and for objects with a certainty of  $\sigma \leq 0.24$  (below). Points are coloured by their true 3D shape class. The total number of predicted objects is indicated by the number in the round bracket in each shape region, followed by the number of true objects of that shape. The negative predicted values (NPV) and positive predicted values (PPV) are shown as a function of galaxy shape within the legend of each plot. We see the expected distributions broadly approaching the true as we select more certain predictions, though uncertain oblate systems (shown in blue) contaminate every other class in both plots. The reasons for this are discussed in Section 5.

Certain galaxy shape classes are harder to predict than the others. Some classes, like spherical galaxies, are rarer and constitute a smaller proportion or relative area (see Fig. 2) of the evenly sampled  $p - q$  space, and this is also true in our Universe. Based on the predictive values, we can be confident of the MDN predictions for oblate galaxies, which exhibit the highest PPV. We see that galaxies predicted to be oblate are very likely to truly be oblate, with  $> 92\%$  of all systems in this region truly being this shape intrinsically. On the other hand, oblate systems have the lowest NPV rate, at  $\sim 36\%$ ,

**Table 2.** The negative predicted values (NPV) and positive predicted values (PPV) for each underlying 3D shape given as percentages, where a higher value is better. In the left column of each metric, we show the values when all returned shapes are considered. In the right column of each metric, we consider only results with uncertainty less than 0.24.

Shape	NPV (all)	NPV ( $\sigma \leq 0.24$ )	PPV (all)	PPV ( $\sigma \leq 0.24$ )
Oblate	35.6%	52.4%	92.5%	93.6%
Prolate	98.8%	98.8%	23.6%	30.2%
Spherical	98.5%	97.7%	5.7%	10.2%
Triaxial	88.4%	80.1%	23.6%	41.8%

demonstrating that an object classified as not oblate in shape has a high chance of actually being an oblate system. Though further improvement could be seen, to  $\sim 52\%$ , by selecting samples with higher certainty. Despite contaminating the most of other regions in the predicted  $p - q$  parameter space, very few other shapes are falling into this region. By contrast, predominantly due to the contamination by oblate systems, we see the PPV values of all other shapes are significantly lower.

At the other extreme, spherical systems have the lowest PPV at  $< 10\%$  and thus lowest confidence. This implies that many of the predicted spherical galaxies will be false positives. Conversely, with a NPV of  $> 97\%$ , we can confidently rule out a galaxy that is identified as not spherical to be true negative. This is also the case for prolate and triaxial galaxies, which also present high NPV, but low PPV. The prolate systems are predicted with particularly high NPV rate, at  $\sim 99\%$ , showing that an object identified as not prolate is very likely not to be intrinsically prolate.

### 5. Discussion and future prospects

As demonstrated in Figs. 7–10, the machine learning approach to intrinsic shape recovery has a number of benefits over previous methods. Our approach returns a probability density function for every projection of individual galaxies. This approach thus provides an estimate of the intrinsic shape along with uncertainties for individual galaxies. Subsequently, this can be used to quantify our trust in the predictions. We can be certain of our prediction for objects with accurately recovered  $p$  and  $q$  values with low standard deviation. Those with large uncertainties are less reliable even for those with retrieved  $p - q$  close to the actual.

In comparison to the original work of Bassett & Foster (2019), we do not see a preferential recovery of  $p = 1$  in our predictions. In particular, Fig. 8 in this work compares favourably to Figure 7 from Bassett & Foster (2019). As suspected in that work, this improvement implies that  $\psi$  is not uniquely linked to intrinsic shape – the addition of further kinematic and photometric information is necessary for realistic  $p - q$  recovery. Considering the results of the PCA, we can see that  $\psi$  is important to the determination, but in combination with a number of other parameters (Fig. 6). From Fig. 5, we also see that no single feature provides a clear mapping to the intrinsic 3D shape.

The kinematic misalignment  $\psi$  does contribute to the absolute eigenvalues of the PCA, but only at the 7th component in Fig. 6. Interestingly, we see that parameters such as the mass-weighted dispersion  $\sigma_m$  and the specific angular momentum  $j$  provide greater constraints to the intrinsic shape recovery than their combined counter-parts,  $\lambda_R$  and  $V/\sigma$ . We note that the dynamical range of each parameter has been controlled for through

normalisation during the PCA, therefore differences in the parameter dynamical ranges cannot be driving these trends. This may be because a level of degeneracy is introduced by the division of parameters in  $\lambda_R$  and  $V/\sigma$  which may negate their shape describing power.

From Fig. 10, we see that, despite clear improvements on previous approaches, we do struggle to recover some regions of the  $p - q$  parameter space. This may in part be due to the formulation of the problem. Mathematically, we know that  $p > q$ , given  $a > b > c$  (and  $p = b/a$ ,  $q = c/a$ ) and that both must be less than or equal to 1. As the MDN returns a Gaussian distribution of predictions in  $p$  and  $q$ , it is down-weighting the possibility of shapes that occur at these boundaries, especially the spherical objects where both ratios are near the boundary  $p \sim q \sim 1$ . This is clearly seen in Fig. 10, where the predicted values fail to occupy any region close to the extremes of the parameter space.

The MDN uses a GMM, which may not be appropriate should the underlying distribution of shapes not be well characterised by Gaussians. As can be observed in Fig. 8, the expected distributions can vary between different systems. Oblate objects mainly have high  $p$  values. However, the recovered distribution has a long tail towards lower values, which causes contamination to other galaxy classes in the  $p - q$  space. For spherical objects, the data appear sparse and modelling with multiple Gaussian components might not be a good representation of their distribution. This has also been noted in Bassett & Foster (2019), particularly for the  $p - q$  recovery of spherical galaxies. Additionally, based on our investigation, using a deeper neural network model and more mixture components in the MDN will not significantly improve the predictions for individual galaxy types. A further exploration is to perform a detailed search of the MDN architecture and the best set and combinations of galaxy properties that can help in improving the machine learning model performance. The addition of other kinematic and photometric properties, especially those known to correlate with structural parameters, such as colour, ionised gas kinematic parameters and star formation, might also provide additional information to aid in the recovery of  $p$  and  $q$ .

This work presents a proof of concept that demonstrates the capability of MDN models to recover the  $p$  and  $q$  values to infer the shape of individual galaxy. Instead of relying on restrictive assumptions about the relationship between the intrinsic shape and parameters such as the kinematic misalignment,  $\psi$ , we use the MDN model to simultaneously train on various kinematic and photometric parameters that provide meaningful properties to characterise 3D galaxy shapes. Generally, the MDN model can recover the shape distribution for prolate objects, but performs more poorly for other shapes. The MDN can provide estimates of the  $p$  and  $q$  values as well as meaningful uncertainties through probability distributions of those parameters. We recommend the use of informative retrieved  $p$  and  $q$  values with low standard deviation, while cautioning against using those with large standard deviation.

The MDN pipeline requires low computational power to run, especially once it has been trained and large volumes of test data can be evaluated almost instantly. Future work will include deploying MDN on mock data sets from a varied set of hydrodynamical cosmological simulations, such as IllustrisTNG. This would provide important insights on how differently the model might perform on real data, how well it scales across different

data sets and also inform how different cosmological simulations compare with one another and with real galaxies.

## 6. Conclusion

Recovering the underlying 3D intrinsic shapes of individual galaxies is a challenging task since our observations of galaxies appear as 2D projection maps. As opposed to traditional methods that assume a relationship between the intrinsic kinematic misalignment and morphological axes, we employ a supervised machine learning method. For this purpose, we build a mock observational dataset from the EAGLE hydrodynamical cosmological simulation as our training and testing datasets. We then apply a MDN to predict the intrinsic axis ratios,  $p$  and  $q$ , that characterise the 3D galaxy shape.

As with other methods, our approach also finds that it is harder to recover the distribution of  $p$  than that of  $q$ , which can be retrieved reasonably well for the majority of systems. We demonstrate that there is no simple pair of parameters that is able to provide a unique link to the 3D intrinsic shape. We find instead that the inclusion and combination of both kinematic and photometric information does further improve the  $p - q$  recovery. We show that our MDN model has great potential in recovering the intrinsic shapes of the galaxies, along with subsequent uncertainties. Our MDN approach can be applied to any IFS data. These works will serve as the foundation for future investigations on the use of ML for galaxy intrinsic shape recovery.

**Data availability statement.** The EAGLE simulation is publicly available at <http://icc.dur.ac.uk/Eagle/>. The code to generate 2D mock IFS observations is publicly available at <https://github.com/kateharborne/SimSpin>. The codes used to perform the analysis presented in this work are available from the corresponding author upon reasonable request.

**Acknowledgement.** We thank the anonymous referee for valuable suggestions on the manuscript.

This project grew out of an Astronomy Data and Computing Services (ADACS) internship scholarship awarded to SY. ADACS is funded from the Astronomy National Collaborative Research Infrastructure Strategy (NCRIS) allocation provided by the Australian Government and managed by Astronomy Australia Limited (AAL). CF is the recipient of an Australian Research Council Future Fellowship (project number FT210100168) funded by the Australian Government. JTM and CF are the recipients of ARC Discovery Project DP210101945. Part of this research was conducted by the Australian Research Council Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D), through project number CE170100013.

This research was undertaken in part using the Setonix machine at the Pawsey Supercomputing Centre in Perth, Australia. We also acknowledge the Virgo Consortium for making their simulation data available. The EAGLE simulations were performed using the DiRAC-2 facility at Durham, managed by the ICC, and the PRACE facility Curie based in France at TGCC, CEA, Bruyères-le-Châtel.

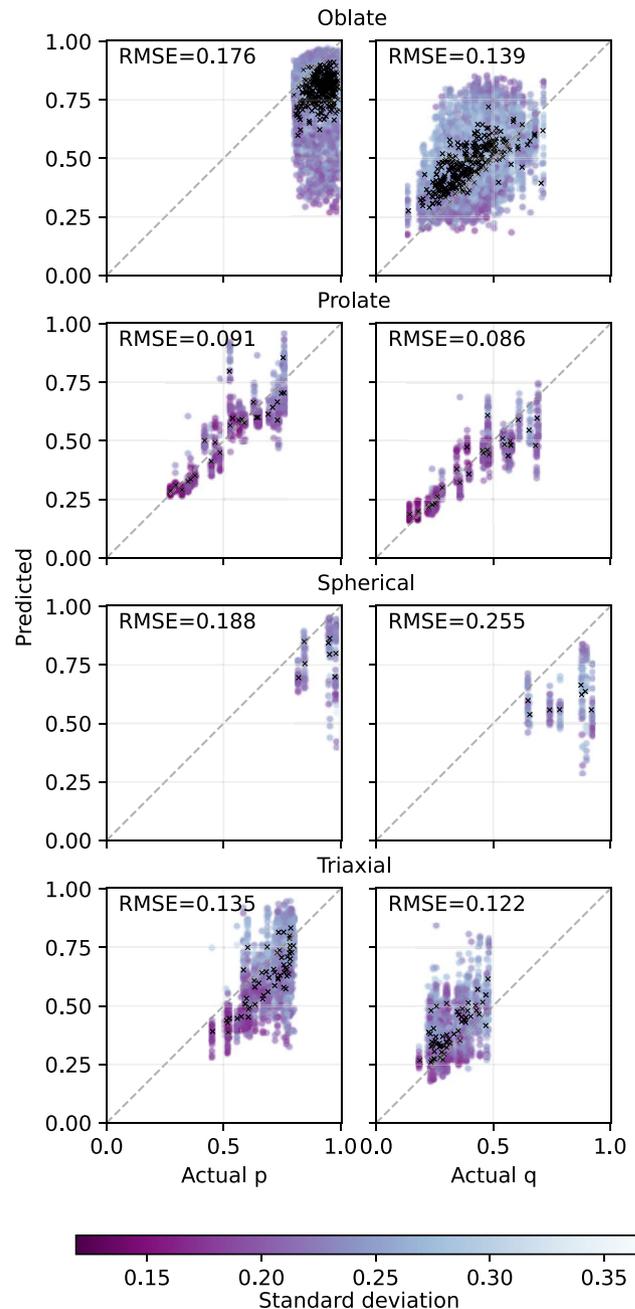
## References

- Abadi, M., et al. 2015, TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems, software available from [tensorflow.org](https://www.tensorflow.org)
- Allgood, B., et al. 2006, *MNRAS*, **367**, 1781
- Bacon, R., et al. 2001, *MNRAS*, **326**, 23
- Bak, J., & Statler, T. S. 2000, *AJ*, **120**, 110
- Bassett, R., & Foster, C. 2019, *MNRAS*, **487**, 2354
- Binggeli, B. 1980, *A&A*, **82**, 289
- Binney, J. 1978, *CA*, **8**, 27

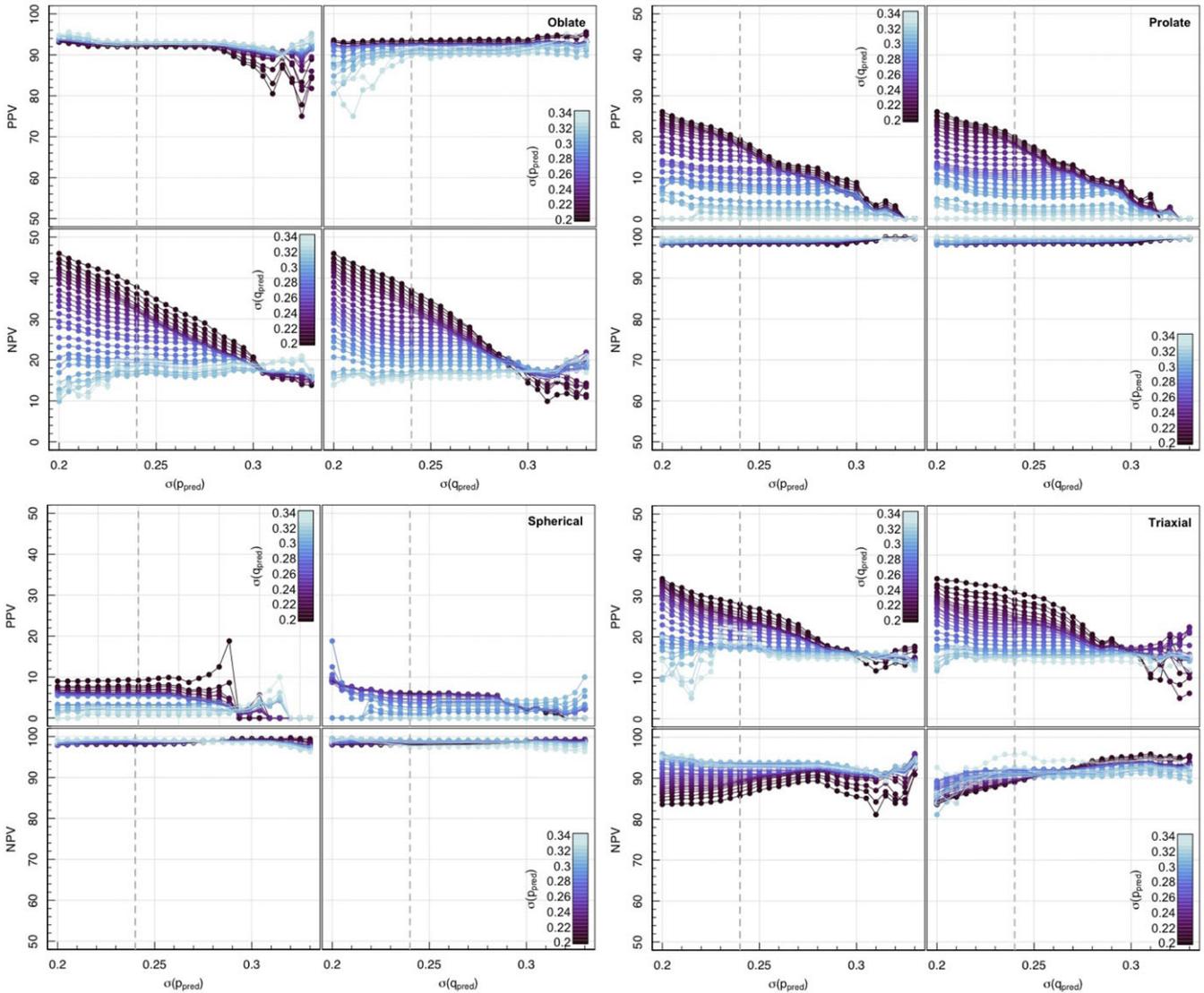
Binney, J. 1985, *MNRAS*, 212, 767  
 Binney, J. 2005, *MNRAS*, 363, 937  
 Bishop, C. M. 1994, Technical report, Aston University  
 Bottrell, C., Hani, M. H., Teimoorinia, H., Patton, D. R., & Ellison, S. L. 2022, *MNRAS*, 511, 100  
 Cappellari, M., et al. 2011, *MNRAS*, 413, 813  
 Cavanagh, M. K., Bekki, K., & Groves, B. A. 2021, *MNRAS*, 506, 659  
 Chollet, F., et al. 2015, Keras, <https://keras.io>  
 Contopoulos, G. 1956, *ApJ*, 124, 643  
 Crain, R. A., et al. 2015, *MNRAS*, 450, 1937  
 Croom, S. M., et al. 2012, *MNRAS*, 421, 872  
 Davies, R. L., & Birkinshaw, M. 1988, *ApJS*, 68, 409  
 Davies, R. L., Efstathiou, G., Fall, S. M., Illingworth, G., & Schechter, P. L. 1983, *ApJ*, 266, 41  
 D'Eugenio, F., Houghton, R. C. W., Davies, R. L., & Bontà, E. D. 2013, *MNRAS*, 429, 1258  
 Emsellem, E., et al. 2007, *MNRAS*, 379, 401  
 Ene, I., et al. 2018, *MNRAS*, 479, 2810  
 Fasano, G. 1991, *MNRAS*, 249, 208  
 Foster, C., et al. 2016, *MNRAS*, 457, 147  
 Foster, C., et al. 2017, *MNRAS*, 472, 966  
 Foster, C., et al. 2021, *PASA*, 38, e031  
 Franx, M., Illingworth, G., & de Zeeuw, T. 1991, *ApJ*, 383, 112  
 Graham, M. T., et al. 2018, *MNRAS*, 477, 4711  
 Green, A. W., et al. 2017, *MNRAS*, 475, 716  
 Greene, J. E., et al. 2018, *ApJ*, 852, 36  
 Harborne, K., et al. 2020, *MNRAS*, 497, 2018  
 Harborne, K. E., et al. 2023, *PASA*, 40, e048  
 Hubble, E. P. 1926, *ApJ*, 64, 321  
 Illingworth, G. 1977, *ApJ*, 218, L43  
 Kimm, T., & Yi, S. K. 2007, *ApJ*, 670, 1048  
 Krajnovic, D., Cappellari, M., Zeeuw, P. T. d., & Copin, Y. 2006, *MNRAS*, 366, 787  
 Lagos, C. d. P., et al. 2018, *MNRAS*, 476, 4327  
 Lambas, D. G., Maddox, S. J., & Loveday, J. 1992, *MNRAS*, 258, 404  
 Li, H., et al. 2018a, *ApJL*, 863, L19  
 Li, H., et al. 2018b, *MNRAS*, 473, 1489  
 Ludlow, A. D., Fall, S. M., Schaye, J., & Obreschkow, D. 2021, *MNRAS*, 508, 5114  
 Ludlow, A. D., Schaye, J., Schaller, M., & Richings, J. 2019, *MNRAS*, 488, L123  
 McAlpine, S., et al. 2016, *A&C*, 15, 72  
 McKay, M. D., Beckman, R. J., & Conover, W. J. 1979, *Technometrics*, 21, 239  
 McLachlan, G. J., & Basford, K. E. 1988, *Mixture Models. Inference and Applications to Clustering* (New York: M. Dekker)  
 Morales, G., et al. 2018, *A&A*, 614, A143  
 Padilla, N. D., & Strauss, M. A. 2008, *MNRAS*  
 Pearson, K. 1901, *LEDPMJS*, 2, 559  
 Robotham, A. S., et al. 2018, *MNRAS*, 476, 3137  
 Robotham, A. S. G., Taranu, D. S., Tobar, R., Moffett, A., & Driver, S. P. 2017, *MNRAS*, 466, 1513  
 Ryden, B. S. 2006, *ApJ*, 641, 773  
 Sanchez, S. F., et al. 2012, *A&A*, 538, 31  
 Sandage, A., et al. 1970, *ApJ*, 160, 831  
 Sande, J. v. d., et al. 2017, *MNRAS*, 472, 1272  
 Sande, J. v. d., et al. 2019, *MNRAS*, 484, 869  
 Schaye, J., et al. 2015, *MNRAS*, 446, 521  
 Sérsic, J. L. 1963, *BAAA*, 6, 41  
 Statler, T. S. 1994, *ApJ*, 425, 458  
 Stein, M. 1987, *Technometrics*, 29, 143  
 van de Sande, J., et al. 2017, *ApJ*, 835, 104  
 van de Sande, J., et al. 2018, *NatAs*, 483  
 van den Bosch, R. C. E., & van de Ven, G. 2009, *MNRAS*, 398, 1117  
 Villaescusa-Navarro, F., et al. 2021, *ApJ*, 915, 71  
 Vincent, R. A., & Ryden, B. S. 2005, *ApJ*, 623, 137  
 Weijmans, A. M., et al. 2014, *MNRAS*  
 Wilkinson, M. J., et al. 2023, *MNRAS*, 519, 5942

## Appendix A. Galaxy Shape Predictions without Feature Selection

Fig. A1 shows the predicted mean of  $p$  and  $q$  from the MDN model when no feature selection is applied, i.e. using all parameters listed in Table 1. In all cases except for  $p$  values for triaxial systems, the MDN performs slightly better with lower RMSE when feature



**Figure A1.** Predicted against actual  $p$  and  $q$  for each galaxy shape using mixture density network (MDN) without performing feature selection for the test data set. The black crosses represent the average prediction, while circles represent projections of individual galaxies colour coded by the standard deviation from the MDN output. The darker the gradient, the more certain. The prediction error is evaluated by the root mean squared error (RMSE), where lower values represent better agreement. For reference, the identity is shown as a grey dashed line. Compared to Fig. 7, the RMSE is marginally worse for all systems except the  $p$  values for triaxial when no feature selection is performed.



**Figure B1.** Trends in positive predicted values (PPV) and negative predicted values (NPV) with varying standard deviations from MDN output ( $\sigma$ ) for each galaxy shape. In each panel, the colour of the points shows the constant bin in the  $p$  or  $q$  value not on the axis, i.e.  $\sigma(p_{\text{pred}})$  in the plot of  $\sigma(q_{\text{pred}})$  and vice versa. The darker the gradient, the narrower the  $\sigma$  and more certain. The vertical grey dashed line shows the mean of the standard deviation at  $\sigma = 0.24$ .

selection is included (see Fig. 7). Incorporating feature selection will be particularly useful especially in the presence of large number of features, which we intend to explore in future work.

### Appendix B. Trends in Certainty of Galaxy Shape Predictions

Fig. B1 shows the trends in positive predicted values (PPV) and negative predicted values (NPV) with different cut-off selection for the standard deviation of the MDN output. The cut-off criterion for Figs. 9–10 and Table 2 of the main paper using the mean of the standard deviation at  $\sigma = 0.24$  is shown in vertical grey dashed line. Broadly, there are improvements in PPV and NPV as more certain objects are selected.

Notably for oblate systems, the NPV rate is refined by a factor of  $\sim 4$  improvement. The PPV for prolate and triaxial systems also improve by a factor of  $\sim 2$  with the removal of less confident predictions. The NPV results for triaxial systems behave opposite from expectation, yielding worse predictions as the standard deviation decreases. This may be due to the expected distribution not being well captured by the MDN model, particularly for  $p$  (see Fig. 8). The recovered  $p$  values for triaxial objects cover a broader range of values compared to the true underlying shapes. There is no improvement in the NPV for spherical systems due to the smaller sample size in that group. In general, predictions with lower  $\sigma$  are to be preferred to ensure the retrieved galaxy shapes are ‘informative’.