

**A PROBABILISTIC DEEP LEARNING MODEL TO
DISTINGUISH CUSPS AND CORED IN DWARF
GALAXIES**

Trabajo Fin de Máster

Author

Julen Expósito

Thesis supervisors

Arianna di Cintio

Marc Huertas-Company



Instituto de Astrofísica de Canarias
Universidad de La Laguna
Departamento de Astrofísica

Abstract

Numerical simulations within a cold dark matter (DM) cosmology form halos with a characteristic density profile with a logarithmic inner slope of -1. Various methods, such as Jeans and Schwarzschild modelling, have been used in an attempt to determine the inner density of observed dwarf galaxies, in order to test this theoretical prediction. Here, we develop a convolutional mixture density neural network (CMDNN) to derive a posterior distribution of the inner density slopes of DM halos. We train the CMDNN on a suite of simulated galaxies from the NIHAO and AURIGA projects, inputting line-of-sight velocities and 2D spatial information of the stars within simulated galaxies. The output of the CMDNN is a probability density function representing the posterior probability of a certain slope to be the correct one, thus producing accurate and complex information on the uncertainty of the predictions.

The model recovers accurately the correct inner slope of dwarfs: $\sim 82\%$ of the galaxies have a derived inner slope within ± 0.1 of their true value, while $\sim 98\%$ within ± 0.3 . We then apply our model to four Local Group dwarf spheroidal galaxies and find similar results to those obtained with the Jeans modelling based code GRAVSPHERE. Fornax dSph has a strong indication of possessing a central DM core, Carina and Sextans have cusps (although the latter with a large uncertainty), while Sculptor shows a double peaked PDF indicating that a cusp is preferred, but a core can not be ruled out.

Our results show that simulation-based inference with neural networks provide a innovative and complementary method for the determination of the inner DM profiles in galaxies.

Keywords: galaxies: dwarf - evolution - formation - haloes - dark matter - simulations - NIHAO - slope - NFW
machine learning: convolutional - neural - network - posterior - probability - distribution

Title: A probabilistic deep learning model to distinguish cusps and cores in dwarf galaxies
Author: Julen Expósito
Supervisors: Arianna di Cintio, Marc Huertas-Company
Study programme: Computational Astrophysics
Institution: Instituto de Astrofísica de Canarias
Year: 2022
Pages: 40

Summary in spanish

El modelo cosmológico Λ CDM ha producido resultados muy satisfactorios a la hora de explicar la estructura a gran escala del universo y las propiedades básicas de las galaxias que lo conforman (Guo et al. 2016; Planck Collaboration 2016). Sin embargo, las observaciones en sistemas a escala galáctica han mostrado varias discrepancias con las predicciones del modelo (Bullock and Boylan-Kolchin 2017). En la literatura, estos problemas se han denominado tradicionalmente los problemas a pequeña escala del modelo Λ CDM.

Uno de los principales y más antiguos de estos problemas es el llamado problema *cusp/core*. Las simulaciones numéricas dentro del modelo Λ CDM forman halos de materia oscura (en los que habitan las galaxias) con un perfil de densidad de masa característico que presenta una pendiente interna en escala logarítmica de $\gamma \sim -1$. Este perfil no concuerda con las observaciones de algunas galaxias con curvas de rotación bien medidas, que prefieren ajustes con pendientes internas de densidad menos fuertes, de $\gamma \sim -0.5 - 0$ (p.ej. J. I. Read, Iorio et al. 2017; Simon et al. 2005; de Blok et al. 2008; Moore 1994). Tradicionalmente se ha llamado *core* a la parte interna de los perfiles de densidad con pendientes internas más llanas, y *cusp* a la parte interna de los perfiles de densidad con las grandes pendientes predecidas por el modelo Λ CDM.

Se ha demostrado que los *cores* pueden explicarse en el paradigma Λ CDM al considerar el efecto que los bariones tienen sobre la materia oscura. Si una galaxia enana acreta lentamente gas y posteriormente este se elimina repentinamente a través de procesos como los vientos estelares o explosiones de supernovas, la materia oscura se expande, disminuyendo la densidad de la misma en la parte central del halo en la que se encuentra la galaxia y pudiendo llegar a formar un *core* si se dan ciertas condiciones razonables (p.ej. Navarro, Eke et al. 1996; Gnedin and Zhao 2002; J. I. Read and Gilmore 2005; Pontzen and Governato 2012). Este proceso ha sido exitosamente modelado en simulaciones hidrodinámicas de galaxias que toman en consideración este *feedback* bariónico.

Aún así, existe mucha controversia dentro de la investigación del problema *cusp/core*, principalmente debido a la dificultad de obtener mediciones precisas de la distribución de materia oscura del halo de las galaxias observadas, siendo habitual que las técnicas tradicionalmente utilizadas para inferir los perfiles de densidad, como la modelización de Jeans (p.ej. Marel 1994; Battaglia, Helmi et al. 2008; Collins et al. 2021; J. I. Read, Walker et al. 2019), no permitan constreñir el valor de la pendiente interna de manera que pueda descartarse o asegurarse la presencia de un *core*.

En este trabajo presentamos un modelo novedoso para determinar la pendiente interna del perfil de densidad de los halos de materia oscura con una cuantificación robusta de la incertidumbre, haciendo uso para ello de técnicas de *Machine Learning*. El objetivo de este trabajo es poder inferir las pendientes internas de las galaxias utilizando únicamente información posicional y espectroscópica de sus estrellas.

Nuestro método utiliza redes neuronales convolucionales de densidad mixta (*Convolutional Mixture Density Neural Networks*, CMDNNs) para modelar la subestructura compleja de las galaxias. Una red neuronal supervisada puede ser entrenada con datos de simulaciones de galaxias, de las cuales conocemos ya la pendiente interna de su perfil de densidad. Una vez entrenada, la red puede inferir la pendiente interna de datos observacionales, haciendo uso del aprendizaje automático que ha realizado al estudiar previamente las simulaciones.

Entrenamos nuestro modelo utilizando un gran conjunto de simulaciones hidrodinámicas de galaxias enanas con halos con masas de 10^9 a $10^{11.5} M_{\odot}$ procedentes de los proyectos NIHAO y AURIGA (Wang et al. 2015; A. A. Dutton et al. 2020; Grand et al. 2017), hasta un total de 183 galaxias diferentes. El uso de diferentes modelos y parámetros físicos empleados en dichas simulaciones nos permite tener

un amplio rango de perfiles de densidad en galaxias con masas similares, incluyendo tanto *cores* como *cusps*.

Hacemos únicamente uso de las velocidades y posiciones de las estrellas encontradas en las galaxias enanas simuladas para construir funciones de densidad de probabilidad (*Probability Density Function*, PDF) continuas de la distribución de dichas estrellas en dos espacios bidimensionales distintos:

- El espacio de fases de las distancias de las estrellas al centro de la galaxia y su velocidad en la línea de visión, que presenta información sobre la relación entre la distancia al centro gravitacional y la velocidad orbital asociada al perfil de densidad de materia oscura.
- El espacio de posiciones de las estrellas en el plano del cielo (x,y), que da cuenta de la forma y tamaño de la galaxia.

Estos PDF bidimensionales, con los cuales definimos las galaxias que queremos evaluar, son los *inputs* de nuestra red neuronal. La red va aplicando transformaciones a los PDF en diferentes capas sucesivas con el objetivo de ir encontrando características de ordenes cada vez superiores según se alcanzan profundidades mayores en la red. Finalmente se llega a una última capa que construye una distribución de probabilidad, dando pues una función que indica la probabilidad de que la galaxia evaluada tenga una u otra pendiente interna.

Dichas transformaciones son alteradas y desarrolladas iterativamente durante el entrenamiento de la red con las simulaciones, hasta llegar a un punto en el que se finaliza la construcción del modelo cuando la red es capaz de inferir correctamente la pendiente interna de las simulaciones con las que ha sido alimentada, dando fin así al entrenamiento. Dicho modelo es después capaz de evaluar datos nuevos que nunca haya visto e inferir la pendiente interna para los mismos.

La pendiente interna de las galaxias simuladas es predecida por nuestro modelo con una desviación media respecto al valor real de $\mu = 0.055$. El modelo predice correctamente el valor de la pendiente interna con un error por debajo de 0.1 (0.3) para el 82% (98%) de las galaxias del set de datos. Estas predicciones presentan, además, incertidumbres con una desviación estándar media de $\sigma = 0.108$. La precisión conseguida es suficientemente pequeña como para poder distinguir con claridad entre *cores* y *cusps* en más del 95% de los casos.

Hemos también aplicado el modelo a cuatro galaxias enanas esféricas, obteniendo valores de pendientes internas consistentes con los resultantes del uso del código de modelización de Jeans GRAVSPHERE. Concretamente, encontramos que Fornax dSph presenta un fuerte indicio de tener un *core*, Carina y Sextans tienen *cusps* (aunque esta última con una gran incertidumbre), y Sculptor muestra una distribución de probabilidad con un doble pico que otorga mayor probabilidad a la presencia de un *cusp*, pero sin poder descartar la posibilidad de un *core*. Estos resultados son consistentes con varias pendientes internas derivadas previamente para estas galaxias (p.ej. Goerdt, Moore, J. I. Read et al. 2006; Agnello and Evans 2012; C. B. Brook and Di Cintio 2015; Walker and Peñarrubia 2011; Richardson and Fairbairn 2014; J. I. Read, Walker et al. 2019).

A la luz de estos resultados, queda claro que nuestro nuevo modelo es una herramienta prometedora para el estudio de las propiedades de las galaxias enanas, y que la inferencia mediante el uso de redes neuronales entrenadas con datos simulacionales proporciona un método innovador y complementario a los ya existentes para la determinación de los perfiles de densidad de materia oscura en las galaxias.

Contents

1	Introduction	6
2	Data	9
3	Methodology	13
3.1	Construction of the Data Set	13
3.1.1	Defining the inner slope	13
3.1.2	Expanding the data set	14
3.2	Pre-processing	15
3.2.1	Kernel Density Estimation	15
3.2.2	Model inputs	15
3.3	The model	16
3.3.1	Neural networks	17
3.3.2	Architecture of the model	19
3.3.3	Training and evaluation	20
3.3.4	Representing uncertainties	22
4	Results	23
4.0.1	Predicting DM inner slopes	23
4.0.2	Uncertainty in the inference	25
4.0.3	Effect of viewing angle on the inference of DM slopes	26
4.1	Application to observed galaxies	28
4.1.1	Testing the similarity of training vs observational data	28
4.1.2	Deriving central DM density slopes of dSphs with CNNs	31
5	Conclusion	35
	References	37

Chapter 1

Introduction

The Λ CDM model in cosmology is extremely successful in explaining the large-scale structure of the universe and the basic properties of galaxies (Guo et al. 2016; Planck Collaboration 2016). In this model, cosmic structure originates from primordial adiabatic fluctuations and grows through gravitational instability in an expanding background. Galaxies are formed in dark matter halos: virialized concentration of dark matter formed after the collapse of overdense regions of the Universe.

Even though its success, it has shown some problems at scales for collapsed objects with virial masses below $M \sim 10^{11} M_{\odot}$, which correspond to halos with virial radius of $R \sim 150$ kpc. Observations of galaxies inhabiting these halos have shown several discrepancies with the predictions of the model (Bullock and Boylan-Kolchin 2017). In the literature, these problems have been traditionally named the small-scale problems of the Λ CDM model.

One of the major and older of these problems is the cusp/core problem. Pure DM structure formation simulations in Λ CDM predict haloes that have an NFW density profile (Navarro, Frenk et al. 1996):

$$\rho_{\text{NFW}} = \rho_0 \left(\frac{r}{r_s} \right)^{-1} \left(1 + \frac{r}{r_s} \right)^{-2}, \quad (1.1)$$

where ρ_0 is the central density and r_s the scale length. This implies an inner profile $\rho_{\text{DM}} \propto r^{-1}$. In general, Λ CDM simulations that include only dark matter find inner profiles $\rho_{\text{DM}} \propto r^{-\gamma}$ with $\gamma \sim 0.8 - 1.4$ over distances near the center of small galaxies (Navarro, Ludlow et al. 2010), which implies a steeply rise of density while approaching the center. This central profile disagrees with observations of some galaxies with well-measured rotation curves, which prefer fits with shallower inner density slopes of $\gamma \sim 0 - 0.5$ (e.g. J. I. Read, Iorio et al. 2017; Simon et al. 2005; de Blok et al. 2008; Moore 1994). We call cored profiles these density profiles with shallow inner slopes, and cuspy density profiles to step profiles similar to NFW.

The cusp/core problem has generated substantial interest over the past decades and there have been numerous attempts to tackle it by focusing on changing the collisionless cold dark matter typically assumed. Over the years several alternative dark matter models have been proposed that tackle the cusp/core problem, like self-interacting dark matter, that invokes a new force acting purely in the dark sector and could transform a dense cusp to a core through energy transfer between the DM particles (e.g. Spergel and Steinhardt 2000; Kaplinghat et al. 2016; Robles et al. 2017), or warm dark matter (e.g. Macciò et al. 2012; Schneider et al. 2017).

However, it has been shown that cores can be explained in the paradigm of Λ CDM with collisionless dark matter when considering the effect baryons have on dark matter. Navarro, Eke et al. 1996 showed

that, if gas is slowly accreted onto a dwarf galaxy and then suddenly removed through processes such as stellar winds or supernovae feedback, the DM expand, lowering its central density. This effect of DM heating is small in realistic conditions (Gnedin and Zhao 2002), but J. I. Read and Gilmore 2005 showed that if the effect repeats over several cycles of star formation, it accumulates leading to a complete core formation. This core can be permanent if the outflows are sufficiently rapid (Pontzen and Governato 2012). Cores can also be formed by dynamical friction, causing dense gas clumps to impart angular momentum to the inner density profile and flattening it (e.g. El-Zant et al. 2001; Goerdt, Moore, J. Read et al. 2010; Cole et al. 2011), although Governato, C. Brook et al. 2010 showed that stellar feedback is most effective at expanding low mass haloes.

Hydrodynamical simulations of dwarf galaxies that take into consideration baryonic feedback have produced the creation of cores in certain conditions (e.g. Governato, C. Brook et al. 2010; Zolotov et al. 2012; Tollet et al. 2016). Governato, Zolotov et al. 2012 showed that the inner slope flattens with increasing stellar mass as an effect from the increase of available energy from supernovae, setting a lower limit of stellar masses on the order of $10^7 M_\odot$ for galaxies to be able to form cores by baryonic feedback. On the other hand, Di Cintio, C. B. Brook, Macciò et al. 2014 showed that the deepened potential in the central region of the haloes resulting from the increase of stellar mass opposes the flattening process of the density profile, which prevents core formation in too massive galaxies. Di Cintio, C. B. Brook, Macciò et al. 2014 proposed that γ depends on the stellar-to-halo mass ratio of galaxies: at $M_{star}/M_{halo} \lesssim 10^{-4}$ the supernovae energy is not enough to produce expansion of the halo. At higher values, γ increases until reaching a maximum around $M_{star}/M_{halo} \sim 3 - 5 \cdot 10^{-3}$, to be reduced again with increasing stellar-to-halo mass ratio as the outflow process becomes more and more ineffective. In realistic galaxies, following the empirical relation between stellar and halo mass of galaxies (Moster, Somerville et al. 2010), this corresponds to $M_{star} \sim 10^{8.5} M_\odot$ and $M_{halo} \sim 10^{11} M_\odot$.

Nevertheless, the cusp/core problem is far from being completely solved. There is continuous discussion around the presence or not of cores in several observed dwarf galaxies (e.g. Evans et al. 2009; Lora et al. 2013; Genina et al. 2017), due to the difficulties of uncovering the underlying dark matter distribution in observed galaxies, which leads to flexible constraints that commonly allow for both cored and cuspy fits.

Historically, mass modelling techniques and Jeans analysis has been used to infer the dark matter density profile of galaxies using observational data (e.g. Marel 1994; Battaglia, Helmi et al. 2008; Collins et al. 2021; J. I. Read, Walker et al. 2019): in rotating galaxies, starting from the total HI gas rotation curve and subtracting the contribution of baryons it is possible to infer the dark matter density profile of the halo, whereas in dispersion supported galaxies, once a stellar velocity dispersion is measured, Jeans modelling allows to measure the density profile of galaxies.

Usually, a near-spherical stellar system is supposed, which is a reasonable approximation to spheroidal and elliptical galaxies. This allows the use of the spherically symmetric Jeans equation projected along the line of sight (J. I. Read and Steger 2017):

$$\sigma_{LOS}^2(R) = \frac{2}{\Sigma_{star}(R)} \int_R^\infty \left(1 - \beta(r) \frac{R^2}{r^2}\right) \frac{\nu(r) \sigma_r^2(r) r}{\sqrt{r^2 - R^2}} dr, \quad (1.2)$$

where r is the radius, $\nu(r)$ is the tracer density, which describes the radial density profile of a population of massless tracers moving in the gravitational potential of a spherical mass distribution, $\beta(r)$ is the velocity anisotropy, which describes the orbital structure of the stellar system where $\beta = 0$ corresponds to an isotropic distribution, $\beta = 1$ is a fully radial distribution and $\beta = -\infty$ is a fully tangential distribution. $\sigma_r(r)$ is the radial velocity dispersion, $\Sigma_{star}(R)$ denotes the tracer surface mass density at projected radius R and $\sigma_{LOS}(R)$ is the line-of-sight velocity dispersion of the tracers.

Equation 1.2 shows that the velocity anisotropy $\beta(r)$, which can't be directly determined, degenerates with $\sigma_r(r)$ and, therefore, with the cumulative mass distribution $M(< r)$. This is known as the $M - \beta$ degeneracy (e.g. Mathieu and Merrifield 2000; Lokas and Mamon 2003; Diakogiannis et al. 2014). Since the radial density profile $\rho(r)$ maps to the enclosed mass $M(< r)$, this is also named the $\rho - \beta$ degeneracy.

Several methods have been developed in the last decades to break the $\rho - \beta$ degeneracy (e.g. Merrifield and Kent 1990; van der Marel and Franx 1993; Saglia et al. 2000; Wojtak et al. 2009; Watkins et al. 2013), which present a wide dispersion of conflicting results in the literature. For example, for the Sculptor dSph, split populations methods (Saglia et al. 2000) favour a cored profile (e.g. Battaglia, Helmi et al. 2008; Agnello and Evans 2012), whereas a VSP method (Merrifield and Kent 1990) indicates the presence of a cusp (Richardson and Fairbairn 2014).

In this work we have developed an alternative method to recover properties of the dark matter density profile of dwarf galaxies from observables without the need to employ Jeans modelling, directly avoiding the problem of $\rho - \beta$ degeneracy. In this project, we suggest the use of Machine Learning methods to that end.

Thanks to the fast development of new methods and access to larger amounts of data, machine learning (ML) techniques have been gaining more and more presence in astronomy and astrophysics, from survey telescope scheduling through object detection and classification, to cleaning images and making large simulations smarter and quicker (Szabó, R. et al. 2022). Recently, a suite of ML algorithms has been used to reconstruct dynamical cluster masses. This class of methods often involves training a neural network on a large data set of simulations-generated mock observations to then produce inference of unlabeled observations. For example, Ntampaka et al. 2016 introduced an ML method to infer mass from the full LOS velocity distribution of cluster members, attempting to capture higher-order features of the velocity distribution using a support distribution machine. On the other hand, Ho et al. 2019 tried to achieve the same by applying convolutional neural networks with satisfactory results.

Influenced by these works, this project aims to test if similar methods could be successfully used to infer the inner slope of the dark matter density profile of halos from radial velocities and positions of stars, and, in a greater scope, whether ML methods can be employed efficiently to derive inner density profiles of real galaxies with small uncertainty.

I have built a neural network trained with hydrodynamic simulations of dwarf galaxies to determine the inner slopes. I use a deep learning methodology based on the use of convolutional neural networks (CNNs) to parameterize higher-order features from phase-space mappings of positional and dynamical distributions of stars. The training of the neural network is done with a suite of 183 dwarf galaxies from the NIHAO and AURIGA projects.

The work is organised as follows. In section 2 we describe the simulation data I have used for the training of the neural network. In section 3 we explain how this data is treated and the ML method in great detail. In section 4 we show and discuss the results of the trained model on a test sample, and then we apply the model to four observed dwarf spheroidal galaxies; Fornax, Sculptor, Sextans and Carina and we compare the result with those obtained with the Jeans modelling code GravSphere (J. I. Read and Steger 2017).

Chapter 2

Data

We aim to train a neural network on a data set of simulated dwarf galaxies with known properties to then produce inference on observations. Independently on the methods and architecture of the network, the model will only be able to make correct inferences if the data it has been trained on corresponds to and gives adequate information about the systems on which the inference is going to be done.

This means that any difference between the physics found in the simulations and that present in the real galaxies will diminish the network's ability to analyse those galaxies, as any features in the observations that are not found in the set of simulations with which the network has been trained will only constitute noise that will diminish the accuracy and correctness of its inferences. For the same reason, it is necessary to have a large dataset covering as wide a variety of dwarf galaxies as possible. The more examples we have, the better the network's ability to "understand" the physical relationship between the input data and what we want it to infer. For example, if all the galaxies in the neural network had a cuspy density profile, the neural network would not be able to make a proper inference about the inner density profile of a cored galaxy, as it has not been exposed to that type of system. As a general rule, increasing the number of simulated galaxies with different characteristics will improve the predictive ability of the model.

In summary, we need a large number of high-resolution dwarf galaxy simulations that are as realistic and complete as possible to train our network.

We firstly include dwarf galaxies with the fiducial NIHAO (Numerical Investigation of a Hundred Astrophysical Objects) model, ranging in halo mass from an order of $10^9 M_\odot$ to $10^{11.5} M_\odot$ and stellar mass from an order of $10^5 M_\odot$ to $10^{10} M_\odot$. NIHAO is a set of 100 cosmological zoom-in hydrodynamical simulations performed using the GASOLINE code (around 60 of them being dwarf galaxies), with an improved implementation of the SPH algorithm (Wang et al. 2015), which stands out for offering simulations of high spatial resolution over a wide range of halo masses. They are fully cosmological simulations of galaxy formation run in a flat Λ CDM cosmology with parameters from the Planck Collaboration (Colaboration 2014).

The masses and force softenings of the dark matter particles of NIHAO simulations are chosen to resolve the mass profile at $\lesssim 1$ per cent of the virial radius, which results in $\approx 10^6$ DM particles inside the virial radius of all main haloes at $z = 0$. The corresponding masses and force softenings for the gas particles are a factor of $\Omega_b/\Omega_{dm} = 0.182$ and $\sqrt{\Omega_b/\Omega_{dm}} = 0.427$ lower (A. A. Dutton et al. 2020).

For the purposes of this work is very important to have simulations that handle as best as possible the baryonic feedback effects. In GASOLINE, stars form from gas that is both cool ($T < 15000$ K) and dense ($\rho > n$, being n the star formation threshold). In the fiducial NIHAO simulations the star formation threshold is $n = 10$ particles per cm^{-3} , which is, for all the simulations with different resolutions

levels, roughly the maximum density that can be resolved:

$$n_{\max} \approx 50m_{\text{gas}}/\epsilon_{\text{gas}}^3, \quad (2.1)$$

where m_{gas} is the gas particle mass, ϵ_{gas} is the gas gravitational force softening and 50 is the number of SPH particles in the smoothing kernel (A. A. Dutton et al. 2020).

Thus, thermal feedback is employed in two epochs as described in Stinson et al. 2013. The first epoch models the energy input from stellar winds and photoionization from bright young stars before supernovae explode. The second epoch models the energy input from supernovae and starts 4 Myr after the formation of the star. The parameters of the fiducial NIHAO simulations for feedback efficiency in these two epochs are, respectively, $\epsilon_{\text{ESF}} = 0.13$ and $\epsilon_{\text{SN}} = 1.0$, which were calibrated against the evolution of the stellar mass versus halo mass relation from halo abundance matching from Moster, Naab et al. 2013 for a $z = 0$ Milky Way mass halo.

NIHAO galaxies have been proven consistent with a wide range of galaxy properties. They form the right amount of stars (Wang et al. 2015), the masses and half-light sizes of the cold gas are consistent with observations (Macciò et al. 2016), they follow the Tully-Fisher relations (A. Dutton et al. 2017), they reconcile the conflict between the steep halo velocity function of Λ CDM and the shallow H1 linewidth velocity function observed in the nearby universe (Macciò et al. 2016) and they emulate the Milky Ways central stellar bar (Buck et al. 2019), among other successes. Given all of this success, they are a reliable representation of real galaxies.

Even so, 60 galaxies are not enough to build a complete set of examples for network training. Furthermore, the fiducial NIHAO galaxies have a density profile that depends on mass (Di Cintio, C. B. Brook, Macciò et al. 2014), which would allow the machine learning code to infer information about the inner density profile based on any indicator of total mass, rather than by the details of the stellar dynamics.

For this reason, we also use simulations from A. A. Dutton et al. 2020 that use the same model, but for different values of star formation thresholds n from 0.1 to 100 particles per cm^{-3} . The $n > 10$ cases have smaller force softenings than the fiducial NIHAO simulations to allow the gas to clump on smaller scales and this reach higher densities, as can be inferred from equation 2.1. Each value of the star formation threshold has been recalibrated, changing the early stellar feedback efficiency ϵ_{ESF} and/or the efficiency of star formation. Some examples of cored and cuspy galaxies from this expanded NIHAO dataset can be seen in 2.1, along with their respective DM density profiles. For a more detailed explanation of the simulation parameters and characteristics, see A. A. Dutton et al. 2020.

To increase the variety of galaxies in the training dataset and decrease the possibility that the neural network learns characteristics of the simulations purely due to the particularities of the simulation code, we have also included 10 simulated dwarf galaxies without baryonic feedback from the comparison set of no-feedback simulations of NIHAO and 12 simulations from AURIGA project. AURIGA is another suite of fully cosmological magneto-hydrodynamical zoom simulations, carried with a different code than NIHAO; AREPO, a moving mesh code that utilizes a dynamic unstructured mesh that can be transformed through a mesh reconstruction at any time-step, which is not the case for standard grid-based methods like GASOLINEs (Grand et al. 2017).

These contributions add up to 183 simulated dwarf galaxies: 60 simulations from the fiducial NIHAO suite Wang et al. 2015, 101 simulations from A. A. Dutton et al. 2020 with varying density thresholds and varying density profile, 10 simulations without stellar feedback and 12 simulations from Grand et al. 2017. We end up with a range of density profiles at every mass, including simulations that have the same initial conditions but different input physics, which results in different profiles for galaxies

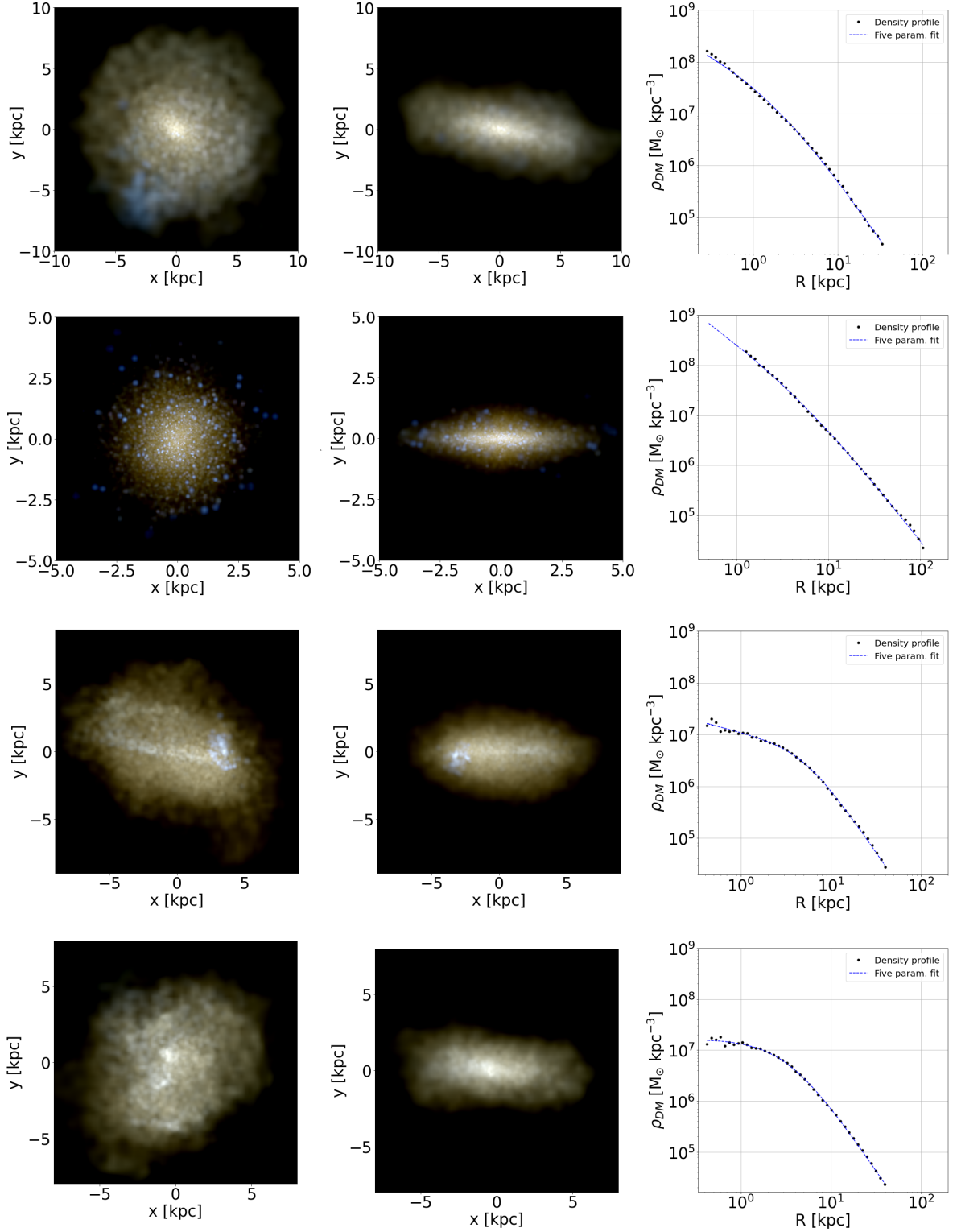


Figure 2.1: Several cored and cuspy galaxies from the simulation dataset. Left: 3-color image of the stars in a face on angle. Each row is a different galaxy. Center: 3-color image of the stars in a side on angle. Right: DM density profile and the fit to a double-power law.

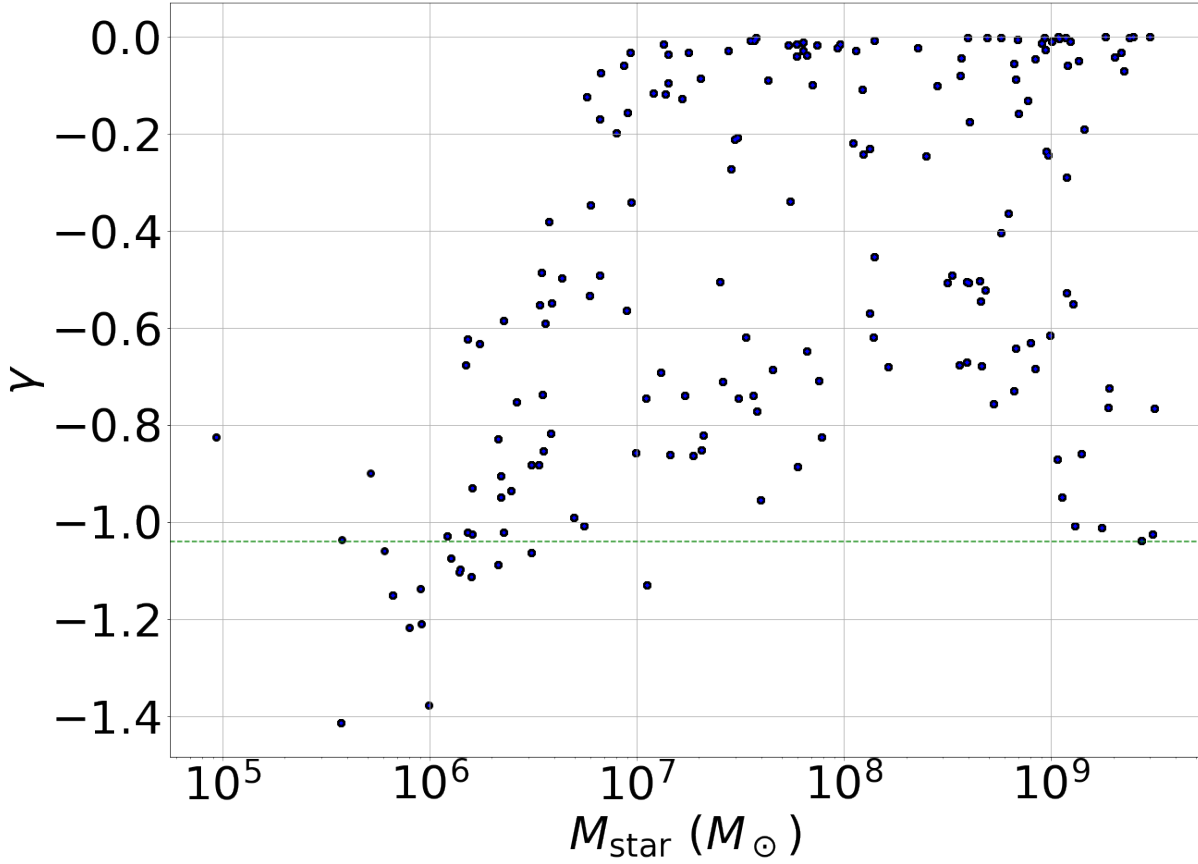


Figure 2.2: Relationship between the stellar-to-halo-mass ratio and the inner slope of the DM density profile (defined as the logarithmic slope at 150 pc) of the galaxies in the dataset. The green horizontal line marks the value of γ for a NFW profile: -1.04.

with similar masses. The relationship between the stellar mass ratio and the inner slope of the profiles for our dataset can be seen in Figure 2.2.

To increase even more the size of our simulation suite, we also use 3 different output timesteps: $z=0$, $z=0.112$, and $z=0.226$. Each of the simulations is virialized at all these times, so this does not change significantly the range of density profiles, but does change the position and velocities of the stars, effectively increasing the sample of stars for each galaxy. This creates a final sample of 549 galaxy snapshots.

Chapter 3

Methodology

In this chapter, we present the deep learning methodology used to acquire information about the inner slope of the dark matter density profiles of galaxies from positional and kinematic information of samples of stars. Our model uses the distribution of line-of-sight velocities and star positions projected on arbitrary planes to train a deep neural network and make it able to produce probabilistic estimations of the inner slope of a given galaxy when being fed with its spectroscopic samples.

3.1 Construction of the Data Set

3.1.1 Defining the inner slope

The dark matter density profiles of the galaxies have been realised taking 50 bins logarithmically spaced between three times the softening length ϵ and the virial radius of the dark matter halo R_{vir} of each galaxy. Being $\rho_{\text{DM}}(r)$ the density at a radial distance r from the halo centre, we define the slope s of the density profile between points r_1 and r_2 as the slope of the straight line joining points $\rho_{\text{DM}}(r_1)$ and $\rho_{\text{DM}}(r_2)$ on the logarithmic scale:

$$s = \frac{\log_{10} \left(\frac{\rho_{\text{DM}}(r_2)}{\rho_{\text{DM}}(r_1)} \right)}{\log_{10} \left(\frac{r_2}{r_1} \right)}. \quad (3.1)$$

We then define the inner slope to be the slope when $r_1 = 150$ pc and $r_2 = 151$ pc. To avoid the noise effect of the calculated density profile in the innermost regions, we have fitted the density profiles $\rho_{\text{DM}}(r)$ to an analytical density profiles $\bar{\rho}_{\text{DM}}(r)$. This profile follows a double-power law model (Di Cintio, C. B. Brook, A. A. Dutton et al. 2014).

$$\bar{\rho}_{\text{DM}}(r) = \frac{\rho_s}{\left(\frac{r}{r_s} \right)^\gamma \left[1 + \left(\frac{r}{r_s} \right)^\alpha \right]^{(\beta-\gamma)/\alpha}}, \quad (3.2)$$

where r_s is the scale radius, ρ_s the scale density and (α, β, γ) free parameters of the model. r_s and ρ_s are values related to the mass and the time of halo formation (references), $-\gamma$ and $-\beta$ are, respectively, the logarithmic slope of the density profile in the inner and outer region of the halo and α is related to the smoothness with which the transition is made between the two (the higher the α , the greater the sharpness in the transition).

To make the fit we have calculated the coefficient of determination R^2 , defined as the square of the Pearson correlation coefficient (i.e. the covariance between the points of the analytic profile and the simulated profile divided by the product of their standard deviations), where the range of possible values goes from $-\infty$ to 1, with 1 corresponding to an exact correlation. The correlation coefficient R^2 is greater than 0.995 for 98.4% of the galaxies and for none of them it is less than 0.99. We conclude, therefore, that the $\bar{\rho}_{\text{DM}}(r)$ profiles appropriately represent the dark matter density profile of all the galaxies of the sample.

3.1.2 Expanding the data set

One of the most important aspects of training a neural network is the number of input elements with which it will be trained and the way in which these elements are defined.

From the positions and velocities of the stars in our 183 galaxies, each of which has a defined inner slope value, we wanted to create a data set that

1. represents the galaxies with motion information of an arbitrary number of stars, to be able to apply the model to any galaxy for whose stars we have motion information, regardless of their number.
2. is made up of elements that we can construct from observational data.
3. has a uniform distribution of inner slopes, in order not to introduce a bias in the training process.

Currently, the number of stars for which motion information is available for the local group galaxies is of the order of 10^3 , but 78% of our observed galaxies have more than 10^4 star particles and 44% have more than 10^5 . It is therefore possible to multiply our data set by constructing subsets of galaxies composed of $N_i < N_t$ number of stars, being N_t the total number of star particles available in each randomly chosen simulation, so that the spatial distribution of the star particles is maintained. In this way, the data set is expanded with N_t/N_i subsets of a given galaxy G_i . One third of the subsets for a given galaxy is done on the simulation snapshot with redshift $z = 0$. The other two thirds are done on the simulation data with redshifts $z = 0.112$ and $z = 0.226$ to add more variety to the data from each galaxy. At this redshifts the galaxy is already virialized.

Each subset is composed of a different number of stars depending on the resolution of the original simulation, so that condition 1 is fulfilled.

When observing galaxies, the only positional information we can obtain about the stars is their projected position in the plane of the sky, and the only kinematic information we can obtain from a spectroscopic data is their velocity in the line of sight, so to fulfill condition 2 our input elements must be constructed from a sample of stars with three associated values: the two-dimensional coordinates of their projected position in a plane and their projected velocity in the normal axis of that plane. In recent years, accurate information on the proper motion of observed stars has been achieved in projects such as GAIA Collaboration 2021, which constitutes a potential new source of exploitable information in the future. However, the model built in this work will use only the line-of-sight velocity of the stars.

To take into account the different angles from which we can observe a given galaxy, we have made projections in arbitrary planes of the different subsets. In this step we have fulfilled condition 3 by oversampling the least represented slopes (making multiple projections to the same subset) and under-sampling the most represented slopes (ejecting some of the subsets made from the dataset).

3.2 Pre-processing

The inputs of our deep neural network model are continuous 2D probability density functions (PDFs) of the distribution of stars in projected spaces, constructed with bivariate kernel density estimations (KDEs). The mapping generated with KDEs allow to encapsulate the features of the original discrete distributions in the same form even if each galaxy subset is represented by a different number of stars.

3.2.1 Kernel Density Estimation

The reason behind using KDEs to make smoothed and continuous PDFs to represent the distribution of the stars in a defined space is the capacity they have to encapsulate the features of the original discrete distribution in the same form even if each galaxy subset is represented by a different number of stars. The PDFs resulting from applying a KDE to different sets of stars belonging to the same galaxy (each with a different number of stars) will have similar characteristics to each other thanks to smoothing and can be represented by matrices of the same dimension.

Let $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$ denote a sample of size n from a random variable with density f , each variable being a two-dimensional vector for the case of a bivariate KDE. The kernel density estimate of f at the point \mathbf{x} is given by

$$f_h(\mathbf{x}) = \frac{1}{n|\mathbf{H}|^{1/2}} \sum_{i=1}^n K[\mathbf{H}^{-1/2}(\mathbf{x} - \mathbf{X}_i)], \quad (3.3)$$

where K is a kernel function and \mathbf{H} is a 2x2 bandwidth matrix. The KDE sums up the density contributions from the collection of data points at the evaluation point \mathbf{x} , so that data points close to \mathbf{x} contribute significantly to the total density, while data points further away from \mathbf{x} contribute less. The shape of those contributions is determined by K , and their width and orientation by \mathbf{H} .

Usually the kernel function K is chosen to be a probability density symmetric about zero (Sheather 2004). In this work we use a 2D Gaussian kernel:

$$K(\mathbf{u}) = (2\pi)^{-3/2} |\mathbf{H}|^{1/2} \exp\left(-\frac{1}{2} \mathbf{u}^T \mathbf{H}^{-1} \mathbf{u}\right), \quad (3.4)$$

where $\mathbf{u} = \mathbf{x} - \mathbf{X}_i$. For the bandwidth matrix, a scaling factor κ is multiplied by the covariance matrix of the data. For the selection of κ we use Scott's Rule (Scott 1992), which, for equally weighted points and two dimensions is

$$\kappa = n^{-\frac{1}{6}}, \quad (3.5)$$

where n is the number of data points. This leads to a fairly strong smoothing, which interested us to reduce the relevance of the number of stars and strengthen the overall evaluation of the data as opposed to individual stars.

3.2.2 Model inputs

From the projected information (positions in the x-y plane and v_{LOS}) of the sample of stars representing each galaxy we have made two maps:

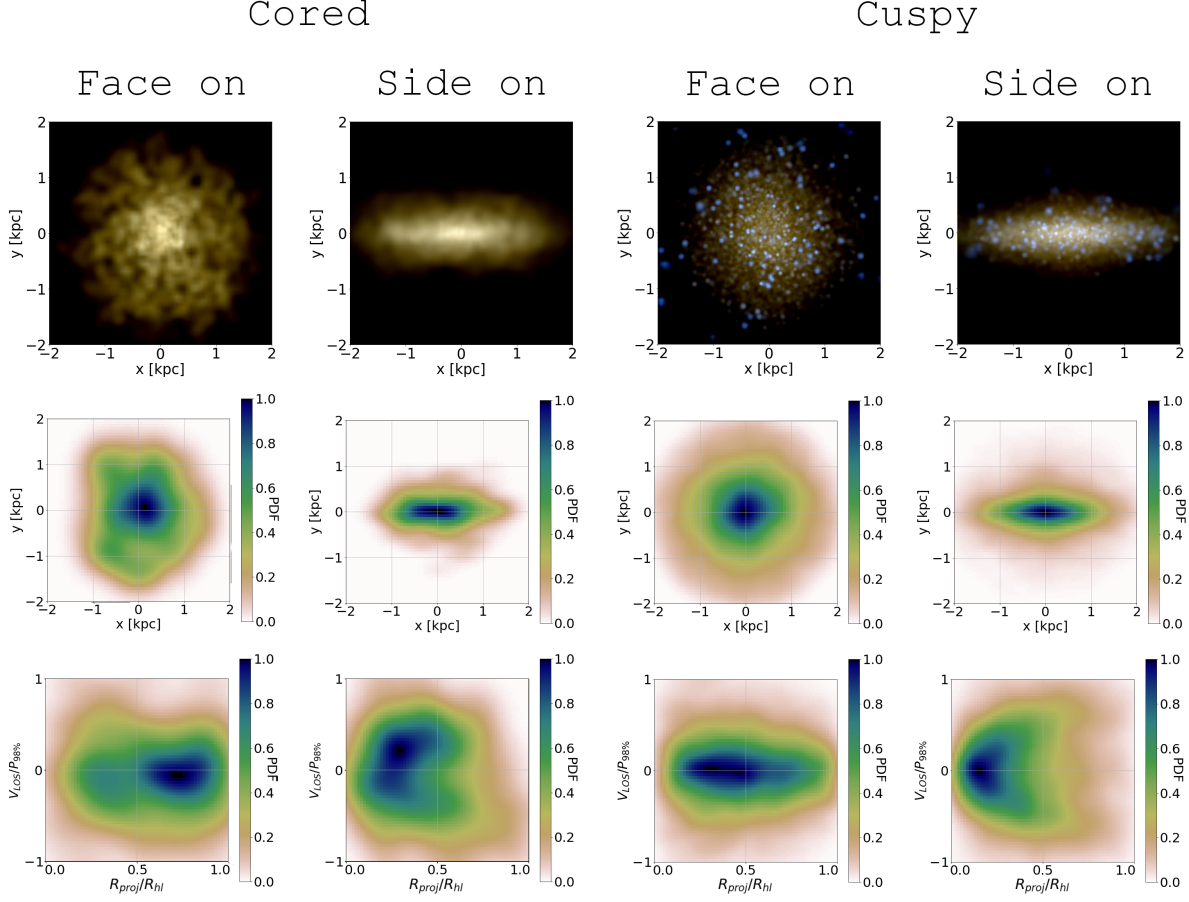


Figure 3.1: Model inputs for a cored and a cuspy galaxy, each one represented face-on and edge-on. The logarithmic slope at 150 pc is $\gamma = -0.20$ for the cored galaxy and $\gamma = -1.32$ for the cuspy galaxy. From top to bottom: A 3-color image of the stars in the galaxy; the PDF in the $\{x,y\}$ space; and the PDF in the $\{\bar{R}_{\text{proj}}, \bar{v}_{\text{LOS}}\}$ phase space.

- A PDF sampled at 64×64 points with the distribution of stars in the $\{x,y\}$ space, between -2 kpc and 2 kpc in each coordinate, in the reference system where $(x,y) = (0,0)$ is the center of the galaxy.
- A PDF sampled at 64×64 points with the distribution of stars in the $\{\bar{R}_{\text{proj}}, \bar{v}_{\text{LOS}}\}$ phase space, where $\bar{R}_{\text{proj}} = \sqrt{x^2 + y^2}/R_{\text{hlr}}$ is the radial position normalized with the half-light radius R_{hlr} and $\bar{v}_{\text{LOS}} = v_{\text{LOS}}/P_{98\%}$ is the line-of-sight velocity normalized with the 98% percentile of the absolute value of v_{LOS} of all the stars of the sample $P_{98\%}$. \bar{R}_{proj} range covers from 0 to 1, and \bar{v}_{LOS} covers from -1 to 1.

So, the inputs for the neural networks are, for each galaxy subset, these two PDFs (normalized by dividing all their elements by the maximum value of each PDF) and the value of its inner slope. An example of the resulting PDFs for a cored and a cuspy galaxy can be seen in Figure 3.1.

3.3 The model

A neural network can be formally described as a trainable and flexible approximation of a model $\mathbb{M} : d \rightarrow t$. The network maps an input data d to a prediction \hat{t} of the target t . This network is parameterized

by a set of trainable weights and a set of hyperparameters. The weights are iteratively optimized during training to minimize a particular loss function, which provides a measure of how close the network prediction \bar{t} is to the target t .

In this work, we use convolutional neural networks (CNNs), a class of deep neural networks (DNNs), to construct a model in which the input data d are the two PDFs described in Section 3.2.2, and the targets t are the inner slopes of the galaxy subsets associated to those two PDFs. We then make a convolutional mixture density neural network (CMDNN) by embedding a Gaussian density layer within the CNN as the last layer.

3.3.1 Neural networks

Deep neural networks

Any neural network is conformed by a set of neuron layers, defined by the following function:

$$f(\mathbf{x}) = g(\mathbf{W} \cdot \mathbf{x} + \mathbf{b}), \quad (3.6)$$

where \mathbf{x} is the input of the layer, \mathbf{W} the weight matrix (which each element being the weight of each element of the vector \mathbf{x}) and \mathbf{b} is a vector called the bias parameter of the layer. $g(\mathbf{z})$ is known as the activation function, which purpose is to break the linearity between the input and the output of the neuron.

A DNN is a neural network conformed by more than one neuron layer. The layers between the input layer (the layer that takes as inputs the input data of the neural network) and the output layer (the layer that gives as output the outputs of the neural network) are called hidden layers.

A feed-forward DNN is a DNN where the neuron layers are evaluated in sequence, passing information from layer to layer without recurrence, which means we can describe the output $\mathbf{h}^{(l)}$ of the l -th layer as

$$\mathbf{h}^{(l)} = g(\mathbf{W}^{(l)} \cdot \mathbf{h}^{(l-1)} + \mathbf{b}^{(l)}). \quad (3.7)$$

The training of the model is done by optimizing the weight matrices $\mathbf{W}^{(l)}$. A model is trained on a set of input data d for which the targets t are known iteratively. In each iteration, the network performance (the similarity between the outputs \bar{t} and the targets t) is evaluated using a loss function, and the weights are actualized to minimize that function by an optimization algorithm. When the loss function stops decreasing and converges to a certain value, the network is said to be optimized. The performance evaluation is done, then, on a set of independent data the model has not seen during training.

Convolutional neural networks

CNNs are a particular type of DNNs especially suited for problems where spatially correlated information is crucial. The main feature of a CNN is the presence of convolutional layers, which are constructed in a way that restrict neurons in one layer to receive information only from within a small neighborhood of the previous layer. This allows neurons to extract simple features from subsets of the previous layer, forming higher-order features in subsequent layers.

A convolutional layer is designed as follows: A convolutional kernel, commonly referred to as a filter, of a given size, encoding a set of neurons, is applied to each pixel (in the case of 2D images

as inputs) of the input image and its vicinity as it scans through the whole region. A given pixel in a specific layer is only a function of the pixels in the preceding layer which are enclosed within the window defined by the kernel, known as the receptive field of the layer. This yields a feature map which encodes high values in the pixels which match the pattern encoded in the weights and biases of the corresponding neurons in the convolutional kernel, which are optimized during training (Kodi Ramanah, Wojtak and Arendse 2020).

A convolutional layer may be described as a linear operation with the discrete convolution implemented via matrix multiplication. In terms of equation 3.7:

$$\mathbf{h}_j^{(l)} = g \left(\sum_{i \in M_j} \mathbf{h}^{(l-1)} \times k_{ij}^{(l)} + \mathbf{b}_j^{(l)} \right), \quad (3.8)$$

where k is the convolutional kernel (the filter) and M_j is the receptive field of the neuron j . One convolutional layer could have multiple filters, which repeat this operation with different kernels, constructing many feature maps per layer, known as channels.

The receptive field is defined by the dimensions of the filter, the stride and the existence or not of padding. The application of the filter can be described as a process of sliding it over the input image of the convolutional layer. We call *stride* to the number and direction of pixels you move the filter at each step, and *padding* to the addition of empty pixels around the edges with the purpose of alleviating information loss around the edges.

Usually, and limiting ourselves to simple architectures, a CNN is a series of convolutional layers followed by a pooling layer as a subsampling or dimensionality reduction step, a process which will reduce the initial input image to a compact representation of features. Then, that representation is reshaped as a vector, which is subsequently passed to a sequence of dense layers (LeCun et al. 2015). This design allows the neural network to autonomously extract meaningful spatial features from the input image. The stack of several convolutional layers builds an internal hierarchical representation of features encoding the most relevant information from the input image. Stacking subsequent convolutional layers naturally strengthens the sensitivity of the most internal layers to features on increasingly larger scales, because the size of the receptive field becomes larger as we go deeper in the CNN.

Mixture density neural networks

A MDNN is a network with layers whose outputs follow a multi-dimension probability distribution, called mixture density layers. This layers take as inputs n nodes, with n being the number of parameters in the desired distribution, transform their values to respect the parameter constraints of the distribution and interpret them as those parameters to construct it. When used as the last layer of the network, it allows joint optimization of the features from the DNN together with a bayesian posterior backend, combining the advantages of deep feature extraction with probabilistic representation of the results.

In this work we use a double gaussian 1D distribution:

$$p_g(x|\theta) = \sum_{j=1}^2 \phi_j \mathbf{N}(x, \mu_j, \sigma_j), \quad (3.9)$$

where $\mathbf{N}(x, \mu_j, \sigma_j)$ is the j Gaussian with mean μ_j and standard deviation σ_j , ϕ_j is the weight of the j Gaussian, so that $\sum_{j=1}^2 \phi_j = 1$, and θ represents the vector of parameters of the distribution.

3.3.2 Architecture of the model

The convolutional sequences are constructed using pairs of convolutional and pooling layers, followed by a dropout layer and a flattening of its output.

The pooling layers downsample their input along its spatial dimension (in this case: height and width). This could be done by several algorithms. In this work, we use the Max Pooling method, which takes the maximum value over a certain input window for each channel. A max pooling layer is defined, as convolutional layers, by the shape of the input window (known as the pooling size), the stride and the padding.

The dropout layer randomly sets input units to 0 with a certain frequency, and scales the rest such that the sum over all inputs is unchanged (Chollet et al. 2015). This is done to prevent overfitting during training; i.e. prevent the model from learning from details that can not be generalized and noise in the training data, negatively impacting the performance on different data.

The final flattening transforms the output into a vector of neurons. Then, the vectors from each convolutional sequence are concatenated and used as input to the joint sequence. The joint sequence is constructed as follows: a normalization layer, two dense layers, a dropout layer, and a final dense layer which elements that represent the parameters the output layer uses to construct the posterior distribution.

The normalization layer applies batch normalization to the input. This normalization maintains the mean of the output close to 0 and its standard deviation close to 1. It works differently during training and during inference:

During training the layer normalizes the output using the mean and standard deviation of the current inputs by the next function:

$$\mathbf{h}^{(l)} = \frac{\gamma (\mathbf{h}^{(l-1)} - \mu_{\mathbf{h}^{(l-1)}})}{\sqrt{\sigma_{\mathbf{h}^{(l-1)}} + \epsilon}} + \beta, \quad (3.10)$$

where ϵ is a small constant, γ a learned scaling factor initialized as 1, β a learned offset factor initialized as 0, and $\mu_{\mathbf{h}^{(l-1)}}$ and $\sigma_{\mathbf{h}^{(l-1)}}$ are the mean and standard deviation of the neurons of the layer $\mathbf{h}^{(l-1)}$, respectively.

During inference (i.e. while using the trained model to makes prediction in some data) the layer normalizes the input using a moving average of the mean $\bar{\mu}$ and standard deviation $\bar{\sigma}$ of the inputs it has seen during training. In the first iteration of the training, $\bar{\mu} = \mu_{\mathbf{h}^{(l-1)}}$ and $\bar{\sigma} = \sigma_{\mathbf{h}^{(l-1)}}$. In subsequent iterations they are updated in the following way:

$$\begin{aligned} \bar{\mu}_{n+1} &= \bar{\mu}_n \cdot m + \mu_{\mathbf{h}^{(l-1)}}(1 - m) \\ \bar{\sigma}_{n+1} &= \bar{\sigma}_n \cdot m + \sigma_{\mathbf{h}^{(l-1)}}(1 - m). \end{aligned} \quad (3.11)$$

The use of the moving averages of the training during inference means that the layer will only normalize its input if the evaluated data is similar to the inference data.

The output layer, which constructs the posterior, is a Gaussian density layer that outputs a mixture distribution where all components are from different parameterizations of a normal distribution.

All the convolutional and dense layers of the model use a non-linear activation function, except the final dense layer of the joint sequence, which passes its neuron parameters directly to the output layer as the parameters of the multivariate Gaussian distribution. The rest of the layers (pooling, dropout...) have a linear activation, which doesn't apply any transformation: $f(z) = z$.

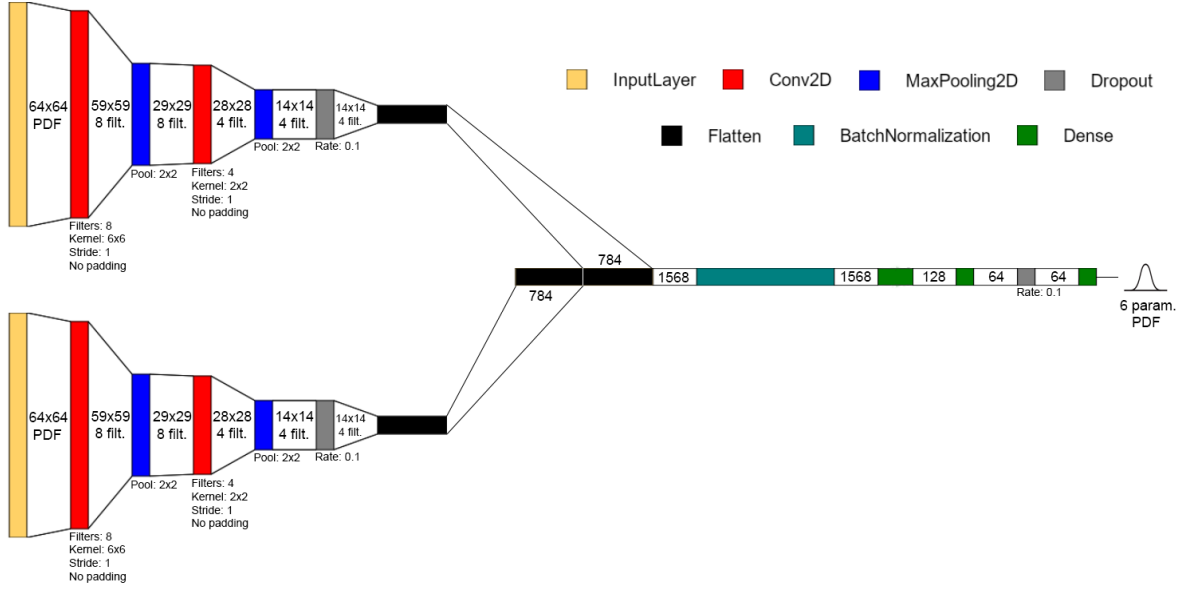


Figure 3.2: Schematic representation of our double channel CMDNN architecture to infer inner slope of the DM profiles (slope at 150 pc) of galaxies from their 2D phase-space mappings of positional and dynamical distributions of stars. The CMDNN extracts the spatial features from the phase-space mappings and gradually compresses into high-order features until describing the input with only 6 parameters, which are used as parameters of a double Gaussian corresponding to the probability density distribution for the inner slope.

The non-linear activation functions allow the network to learn complex structures in the data without the restraints of purely linearly connected neurons. In this work, we use a rectified linear unit (ReLU) in all the layers with a non-linear activation function. The ReLU function is given by $f(z) = \max(z, 0)$, which is linear for values equal or greater than zero, but returns zero if the input is negative. This function is able to introduce non-linearity in the network while maintaining properties of a linear activation function that are desirable in gradient-based methods of optimization during training (see section 1.3.4), and it has been shown to not saturate as much as other popular activation functions (Glorot et al. 2010).

A schematic view of the architecture used in this work can be seen in Figure 3.2.

3.3.3 Training and evaluation

The training is done over a training set consisting on 10273 galaxy subsets with their respected inner slopes, which function as targets. The loss function to minimize during the training is the negative logarithmic likelihood of the training sample, defined as

$$L = -\ln L^* = -\sum_{i=1}^N \ln [p_g(t_i|\theta)], \quad (3.12)$$

where t_i is the inner slope of the galaxy subset i and θ the set of parameters of the gaussian distribution p_g . Note that, for a certain galaxy subset, the likelihood is the value of the probability density function of the multivariate Gaussian in its real inner slope value; i.e the probability the model predicts for the inner slope of the galaxy to be its correct value. Taking the logarithm allows to calculate the joint value by taking sums instead of products, avoids having to calculate exponentials in the Gaussians, and eliminates the potential problem of going below the floating point precision if the negative likelihood becomes

too small during training. It is, so, more computationally efficient than using the joint likelihood directly, while maintaining the capacity to function as the loss function, because $x_1 < x_2 \rightarrow \ln x_1 < \ln x_2$.

The minimization of the loss function is done with the adaptative moment estimation (adam) optimizer, an algorithm for optimization that uses the gradient descent iterative technique. These techniques take randomly initialized weights and make small steps in the direction of the slopes at each iteration (epoch, from now on), looking for the global minimum of the loss function.

Adam updates the weights as

$$w_{i+1} = w_i - \frac{\alpha}{\sqrt{\hat{v}_i} + \epsilon} \hat{m}_i, \quad (3.13)$$

where i is the current epoch, ϵ a small constant to prevent division by zero, α the learning rate and

$$\begin{aligned} \hat{m}_i &= \frac{m_i}{1 - \beta_1}, \quad m_i = \beta_1 m_{i-1} + (1 - \beta_1) \left[\frac{\partial L}{\partial w_i} \right], \\ \hat{v}_i &= \frac{v_i}{1 - \beta_2}, \quad v_i = \beta_2 v_{i-1} + (1 - \beta_2) \left[\frac{\partial L}{\partial w_i} \right]^2, \end{aligned}$$

where β_1 and β_2 are called the decay rates. In this algorithm, m_i and v_i are estimates of the mean and the uncentered variance of the gradients, which are initialized as vectors of zeroes. This initialization produces a bias towards zero in their values, which is handled in the bias-corrected estimates \hat{m}_i and \hat{v}_i . Note that the terms m_i produce larger updates for larger gradients. Its effect is accelerating the optimization process towards the relevant direction, gaining faster convergence by giving momentum to the weights: weights on a steep slope of L will gain more speed towards the minimum, while weights near the minimum will be updated in the same iteration by a much smaller amount. The v_i term, on the other hand, produces an adaptative learning rate to each weight $\alpha/(\sqrt{\hat{v}_i} + \epsilon)$ because the difference in the magnitudes of the gradients between different layers makes the appropriate learning rate vary widely between different weights.

In this work we use the default values proposed by the designers of Adam: $\alpha = 0.001$, $\beta_1 = 0.9$, $\beta_2 = 0.999$ and $\epsilon = 10^{-8}$. Between the popular learning-method algorithms, Adam is shown to compare favorably in performance and computational cost (Kingma and Ba 2014).

The weights are updated several times in each epoch, only using a subset of the training dataset for the calculation of the loss function in each update. For the epoch i , subsets of $S < N$ elements, where N is the number of elements in the training dataset and S is called the batch size, are randomly selected from the complete training dataset, separating the training dataset into N/S batches. In one epoch, these batches are passed sequentially through the model, which calculates the loss function and updates the weights after each batch has passed. Training with different subsets in each epoch reduces the risk of getting stuck at a local minimum, granting a more robust convergence while significantly speeding up the training process.

In each epoch, after updating the weights, the model is evaluated on a validation dataset, composed of galaxy subsets that are not in the training dataset. We consider the model has converged when the loss function for the validation dataset hasn't improved in the last 10 epochs.

Usually, the test dataset for the final evaluation of the converged model is constructed by randomly taking a sufficient number of elements from the complete dataset to correctly represent all feature variety in the data. In this work, due to the limited number of galaxies available, removing too many galaxies with varying characteristics from the training dataset is expected to worsen the performance

of the model, since we do not have many different examples of galaxies with similar characteristics to each other. To properly evaluate the model, we have performed multiple complete training runs using the projections of only 10 galaxies as validation and test datasets in each one, changing the galaxies that would come out of the training dataset in each of the training runs to evaluate the network in several projections of every galaxy. This allows us to analyse the consistency of the model training and its performance in a large number of galaxies without compromising the train dataset.

3.3.4 Representing uncertainties

The output posterior distribution represents the random or aleatoric uncertainty in the slope prediction of the final model, but it does not represent the uncertainty due to the stochastic nature of the weight determination while training the neural network (epistemic uncertainty), which can lead to different models for the same training conditions when dealing with limited data. We use the Monte Carlo dropout method (MC-Dropout) (Gal and Ghahramani 2015) to approximate the epistemic uncertainty which is based on the repeated evaluation of the same input, randomly setting to 0 the weights on some layers while doing each inference, to construct a final evaluation with statistical information about the epistemic uncertainty. Gal and Ghahramani 2015 showed that applying dropout during inference is equivalent to an approximation to a probabilistic Deep Gaussian process. It means we can measure the epistemic uncertainty by applying the dropout layer during inference for a statistically relevant number of them, acquiring a predictive mean and variance for each point of the posterior distribution. The constructed final posterior for each galaxy projection is the normalized mean of 100 multivariate Gaussian posteriors inferred by the model with active dropout layers.

Chapter 4

Results

The goal of our work is to infer the slope of the mass density profile in the central region of a galaxy (from now on: inner slope) from spectroscopic data of a random sample of its stars.

To do so, all simulated galaxies and their subsets of stars are randomly projected in several sky planes, to simulate several viewing angles, and the neural network is trained to be able to determine the inner slope of the galaxy from the positions and line-of-sight velocities of its stars. For each galaxy the neural network outputs a probability density function which approximates the posterior probability of the inner slope of the galaxy, given the observables and under the prior distribution of slopes with which the neural network has been trained.

4.0.1 Predicting DM inner slopes

We define two different methods to construct the predicted slope value γ from the posteriors:

- by using the mode of the posterior distribution (i.e. the maximum of the PDF): $\gamma_{\text{Pred,mode}}$.
- by using the mean of the normalized posterior distribution: $\gamma_{\text{Pred,mean}}$.

The error ϵ of a prediction is defined as $\epsilon_i = \gamma_{\text{Real}} - \gamma_{\text{Pred,i}}$, where γ_{Real} is the real slope at 150 pc of the DM profile of a galaxy simulation. The results for the mode method can be seen in Fig. 4.1, which shows the difference between the real and predicted slopes of our simulated dwarf galaxies, $\gamma_{\text{Real}} - \gamma_{\text{Pred}}$, as a function of the real slope. Each point represents the mean error for every projection of each individual galaxy, while the error bars indicate the minimum and the maximum value amongst every possible projection of each galaxy. Shaded colored horizontal areas represent increasing uncertainty ranges, from ± 0.05 to ± 0.4 .

The mean global error on the predicted inner slope, for all the galaxies in our set, is of $\mu_\epsilon = 0.055$ for the mode method and of $\mu_\epsilon = 0.068$ for the second method. Note that while cuspy and ‘in between’ galaxies are scattered around $\gamma_{\text{Real}} - \gamma_{\text{Pred}} = 0$, cored galaxies tending towards $\gamma = 0$ are necessarily only scattered at $\gamma_{\text{Real}} - \gamma_{\text{Pred}} \geq 0$, since by construction the maximum possible inner slope is 0.

In Table 4.1 we can see the percentages of correctly predicted inner slopes, taking into account all the projections of every galaxy (middle column) and each galaxy individually (right column), for our complete test dataset, within several uncertainty ranges. Roughly 82% of the galaxies recover the correct, real inner slope within ± 0.1 , while 98% of them lie within $|\gamma_{\text{Real}} - \gamma_{\text{Pred}}| \leq 0.3$. These ranges are clearly small enough to shed light on the discussion regarding the presence or not of cores in dwarf galaxies.

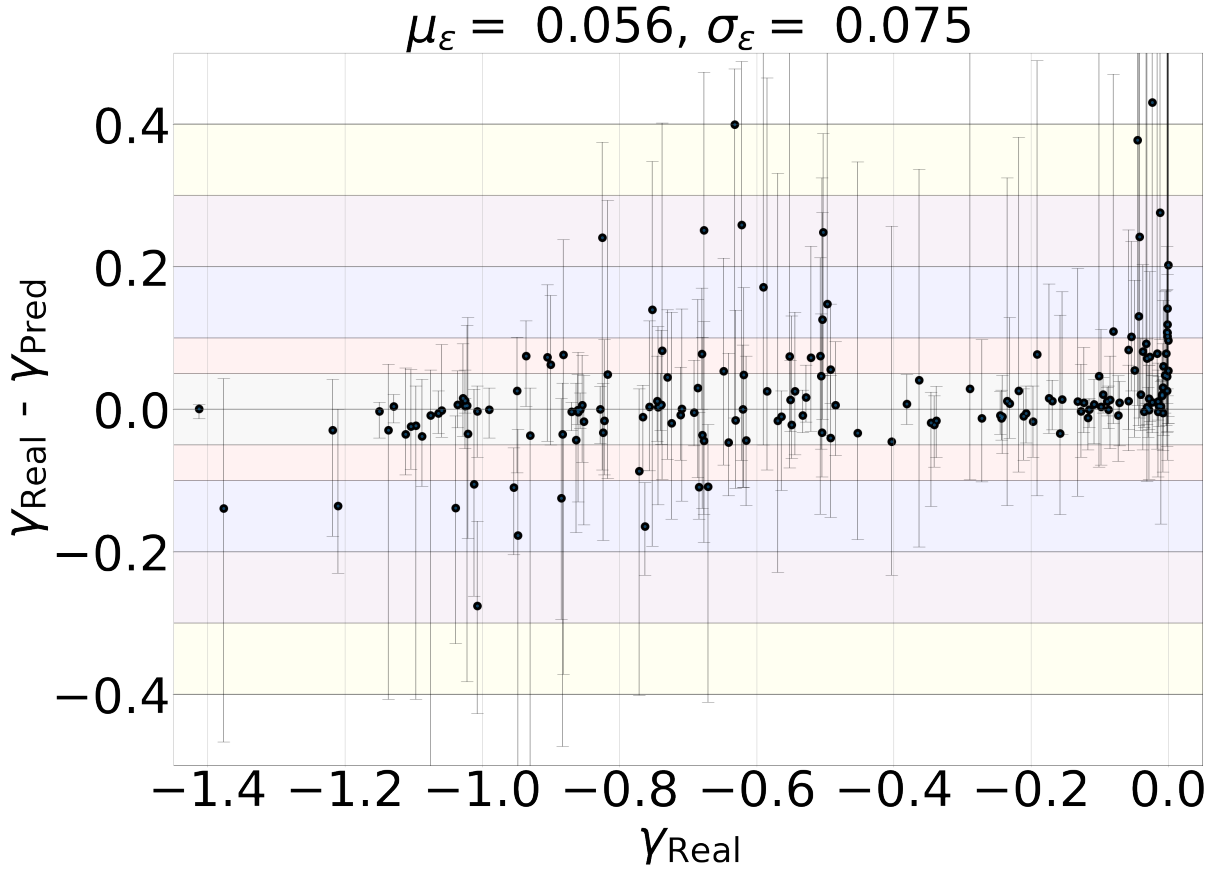


Figure 4.1: Real minus predicted value of DM profiles inner slopes (defined at 150 pc) vs real inner slope, for the simulated galaxies used in this work, defining the predicted value as the mode of the posterior distribution.. Each point represents the mean $\gamma_{\text{Real}} - \gamma_{\text{Pred}}$ for all the projections of each individual galaxy, while error bars span the range between the minimum and the maximum error amongst every possible projection of each galaxy. Colored areas represent increasing error ranges, from 0.05 to 0.4.

Error range ($\pm X$)	% of projections with $ \epsilon \leq X$	% of galaxies with $ \epsilon \leq X$
0.05	66.67	67.80
0.1	80.79	81.92
0.2	94.35	94.35
0.3	98.31	98.31
0.4	99.44	98.87

Table 4.1: Percentage of all the projections (central column) and of individual galaxies (right column) whose predicted inner slope lies within a given error range X , i.e. $|\epsilon| = |\gamma_{\text{Real}} - \gamma_{\text{Pred,mode}}| \leq X$. Here, we used the mode of the posteriors method to derive the inner slopes.

Finally, a histogram of the error distribution for every projection of each galaxy (i.e. 10273 in total) can be seen in Fig. 4.2, indicating that the values of $\gamma_{\text{Real}} - \gamma_{\text{Pred}}$ are peaked at and symmetrically distributed around 0.0, except very cored galaxies that have by definition $\gamma_{\text{Real}} - \gamma_{\text{Pred}} \geq 0$, as already stated. We showed that our method predicts accurately the expected inner slope of galaxies regardless of their actual real slope, with a mostly uniform scatter of $\sigma_\epsilon = 0.075$. This also explains the asymmetry found in Fig. 4.2 which shows a higher degree of errors overestimating the slope of galaxies versus underestimating the slope of galaxies.

The predicted values do not follow a Gaussian distribution around the real value: roughly 75% of the points can be found within $1 - \sigma$, which is compensated by a decay in the probability of finding a prediction at greater distances from the correct value less than would correspond to a Gaussian distribution of errors for larger values of ϵ .

4.0.2 Uncertainty in the inference

In Fig. 4.3 we show the standard deviation σ_{pos} of each posterior from every galaxy in the test dataset, defined as the square root of the variance of the normalized posterior:

$$\sigma_{\text{pos}}^2 = \int_{-\infty}^{\infty} (\gamma - \mu_{\text{pos}})^2 P(\gamma) d\gamma, \quad (4.1)$$

where $P(\gamma)$ is the normalized posterior distribution and μ_{pos} is the mean of the distribution:

$$\mu_{\text{pos}} = \int_{-\infty}^{\infty} \gamma P(\gamma) d\gamma. \quad (4.2)$$

The mean of all the σ_{pos} of the dataset, $\mu_{\sigma_{\text{pos}}}$, is around 0.1 and only 8.99% of the projections have values of σ_{pos} greater than 0.2, uncertainties that are small enough to clearly distinguish between cores and cusps in the vast majority of cases. Fig. 4.3 shows that the standard deviation σ_{pos} of each posterior PDF is uniform across the inner slopes values, i.e. the width of the PDFs does not depend on the inner slope of galaxies, such that the model is not biased towards recovering with higher accuracy either cusps or cores. The uncertainty of a single projection is dependent on the angle of observation. Greater differences are found in very cored galaxies, but in general this dependence doesn't depend strongly on the real slope. These differences between projections of a galaxy compared to the differences between individual galaxies imply that most of the angles of observation yield small uncertainties for the majority of the galaxies, but some angles tend to increase uncertainty significantly.

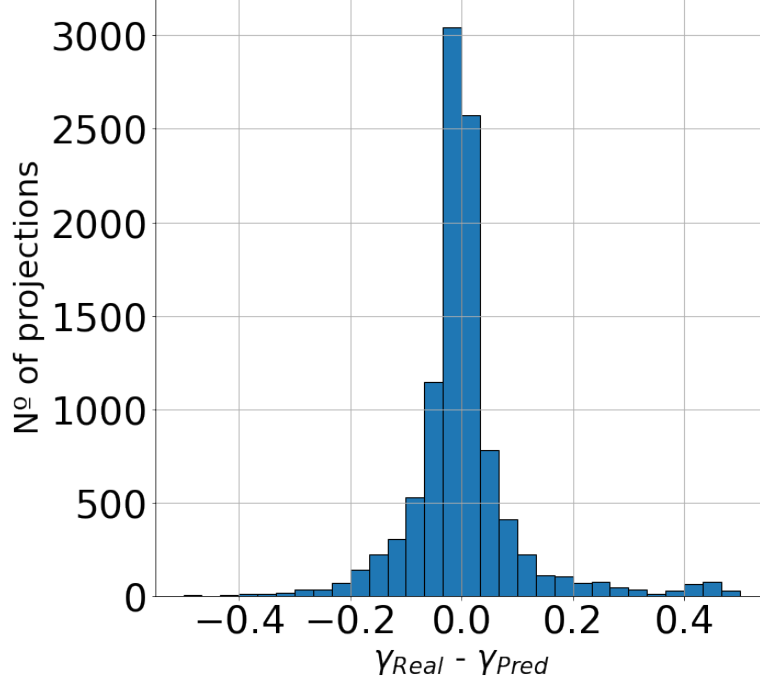


Figure 4.2: Distribution of $\gamma_{\text{Real}} - \gamma_{\text{Pred,mode}}$ for every projection of each galaxy in our training set. The mean of this distribution is $\mu_{\epsilon} = 0.055$ and its scatter $\sigma_{\epsilon} = 0.075$.

Region	% of projections for which γ_{Real} is within region
$1 - \sigma_{\text{pos}}$	86.29
$2 - \sigma_{\text{pos}}$	97.57
$3 - \sigma_{\text{pos}}$	99.73

Table 4.2: Percentage of predictions within increasing σ_{pos} ranges X , defined as $|\gamma_{\text{Real}} - \gamma_{\text{Pred}}| \leq X$.

Table 4.2 shows the percentage of the test dataset projections where the true value of the parameter lies within different multiples of σ_{pos} . If we approximate the posteriors to single Gaussians (which is a proper approximation for roughly 90% of the projections), a well calibrated uncertainty should provide around 68% of the outputs within a confidence level of $1 - \sigma_{\text{pos}}$. Our greater percentage ($\sim 86\%$) of projections within the confidence level of $1\sigma_{\text{pos}}$ indicates that the model is over-predicting the uncertainties σ_{pos} , yielding broader posteriors than it should. This can be an effect of a too high dropout rate during training, which has been shown to have such a outcome on the results of probabilistic neural network models (Ghosh et al. 2022). As it is, our model should be interpreted as conservative, since a future, better calibrated CNN would provide even tighter uncertainties in recovering the true inner slope of a galaxy.

4.0.3 Effect of viewing angle on the inference of DM slopes

Most of the posteriors for the different projections have an approximately normal distribution (the second Gaussian disappearing or constituting a skewness correction to the main Gaussian), but several of them have two distinct peaks. Specifically, around 30% of the galaxies have double peaks in more

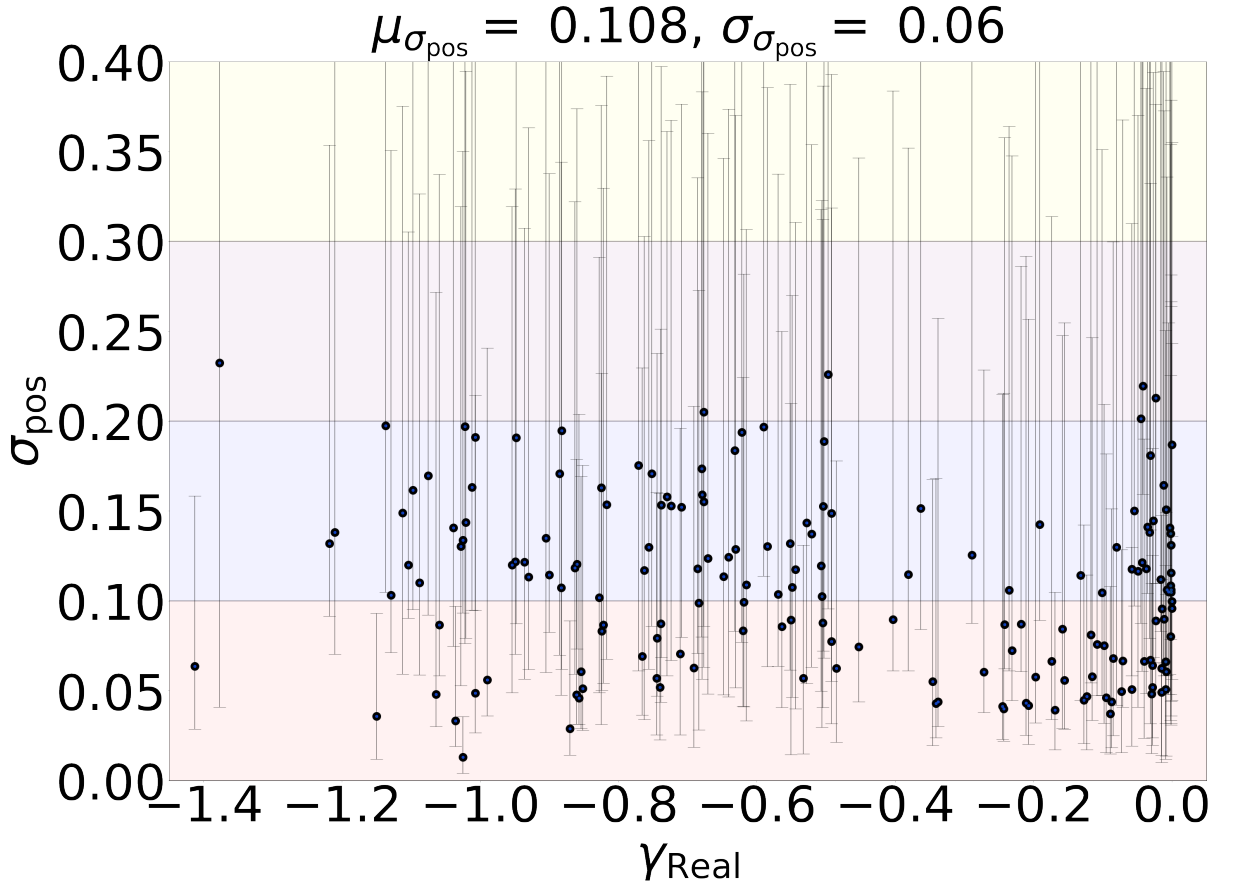


Figure 4.3: Standard deviation σ_{pos} of each posterior PDF vs γ_{Real} , for the simulated galaxies used in this work. Each point represents the mean standard deviation of each posterior, for every projection of an individual galaxy. The error bars range between the minimum and maximum standard deviation value of the posteriors of all the projections of that galaxy.

than 10% of their posteriors. 54% of these galaxies are cored while 46% are cuspy, indicating that the appearance of double peaks in the PDFs arises in both scenarios (here, we define as cored galaxies those with inner slope $-0.6 < \gamma < 0$, and cuspy any galaxy with $\gamma < -0.6$).

In Fig. 4.4 and 4.5 we show the PDFs and posteriors of two galaxies at different observation angles, spanning the range between a face-on and a edge-on view. Strikingly, these images show that the width of the PDFs as well as the appearance of double peaks are strongly related to the viewing angle of the galaxy. This indicates that the appearance of double peaks is a consequence of the fact that some information on the underlying DM profiles is hidden when viewing the galaxy at some particular angle, while it is released and efficiently passed to the network when looking at the galaxy from other angles. For example, in Fig 4.5 we observe that the double peaks in the posterior distribution disappear when the galaxy is seen edge-on, while a face-on configuration provides a second peak that mimics the presence of a cusp. However, this is just an example, and we have several cases of galaxies in which the double peaks appear in edge-on view and disappear in face-on, so that the appearance of these multiple peaks is not related to a specific edge-on or face-on configuration.

4.1 Application to observed galaxies

We proceed to test our model with real observed galaxies, in order to ensure the applicability of the model and to verify that the neural network is not detecting features of simulated galaxies that do not correspond to any real physical system.

We selected a sample of four dSphs for which detailed spectroscopic samples of stellar-kinematic data have been published. At this stage, we adopt only the catalogs by Walker, Mateo et al. 2009 to directly compare our results with those obtained using the code GRAVSPHERE in J. I. Read, Walker et al. 2019. These galaxies are Carina, Sextans, Fornax and Sculptor, for which we further use the center position, velocity, ellipticity and half-light radius (circularized as $R'_{\text{hlr}} = R_{\text{hlr}}\sqrt{1 - \text{ell}}$, where R_{hlr} and ell are the half-light radius and ellipticity) as compiled in Battaglia, Taibi et al. 2022.

To build our input PDFs, we considered only those stars with a 90% or higher probability of being part of the galaxy and we take the mean value of the line-of-sight velocity for those stars with multiple measurements. In total, we considered 460 stars for Carina, 1353 for Fornax, 809 for Sculptor and 327 for Sextans, and for each of these stars we used projected x-y positions and line-of-sight velocities.

4.1.1 Testing the similarity of training vs observational data

When training a neural network with simulations to then perform inference on real data, there is always the risk that the network will detect and learn from specific features of the simulation code that do not correspond to reality, and this would cause issues when interpreting observational data, as they have different qualities than those used in the training set. We can test the degree to which our network sees observational data as equivalent to the data it has been trained on by observing the parameter space of the test dataset, defined as the set of all combinations of the six parameters corresponding to each element of such dataset. Namely, our outputs are defined by the mean, standard deviation and weight of two Gaussians: this 6D parameter space will have regions populated with points and regions completely empty, corresponding to the combinations of parameters that do not parameterise the characteristics of any physical system found in the dataset. If the neural network does not see differences in the input with respect to the data it has been trained on, the resulting parameters, coming from the evaluation of observational data with our model, will fall within the populated regions of the parameter space of the simulation dataset.

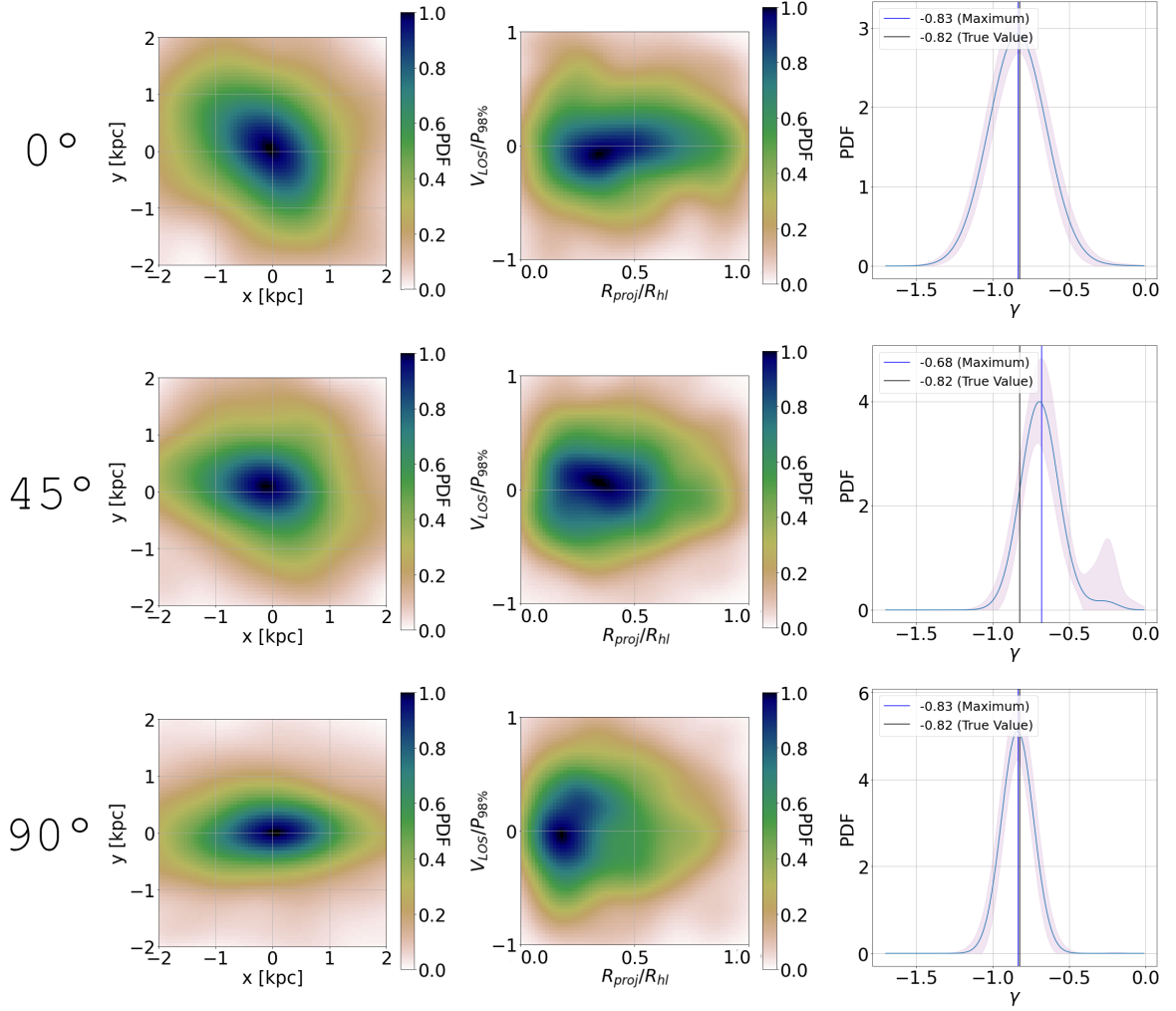


Figure 4.4: Probability density distributions used by the neural network as input in the case of one simulated galaxy subset seen at 0° (face-on), 45° and 90° (side-on), alongside with the Bayesian posteriors predicted by the model. Left columns: PDFs in the $\{x, y\}$ space. Central columns: PDFs in the $\{\hat{R}_{\text{proj}}, \hat{v}_{\text{LOS}}\}$ phase space. Right columns: predicted Bayesian posterior in the space of inner slope of the DM profile (slope at 150 pc); shaded regions represent the standard deviation of the posterior values for the MC-Dropout inferences at each slope point, while the blue vertical line shows the mode (maximum) of the posterior distribution and the black one the true value of the inner DM slope.

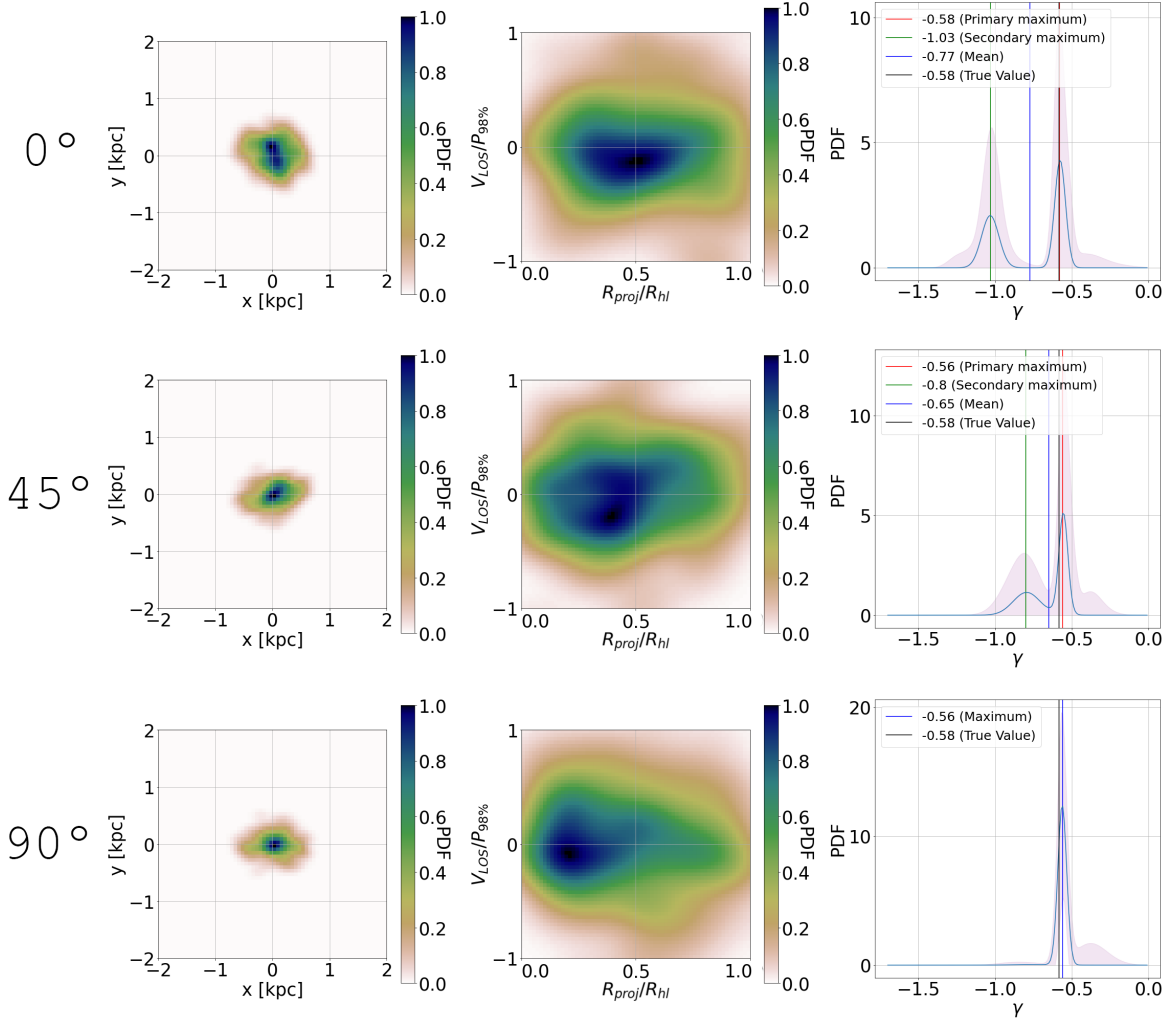


Figure 4.5: Probability density distributions used by the neural network as input in the case of one simulated galaxy subset seen at 0° (face-on), 45° and 90° (side-on), alongside with the Bayesian posteriors predicted by the model. Left columns: PDFs in the $\{x, y\}$ space. Central columns: PDFs in the $\{R_{\text{proj}}, \hat{v}_{\text{LOS}}\}$ phase space. Right columns: predicted Bayesian posterior in the space of inner slope of the DM profile (slope at 150 pc); shaded regions represent the standard deviation of the posterior values for the MC-Dropout inferences at each slope point. The red vertical line shows the primary maximum of the posterior distribution (the mode), the green one the secondary maximum and the black one the true value of the inner DM slope. As a blue line, the mean between primary and secondary maximum is shown, when two peaks exist (in the bottom panel, instead, the blue line represents the unique maximum). This example shows how the appearance of double peaks in the posterior distributions is strongly related to the viewing angle.

To be able to visualize the 6D parameter space and test if this is the case, we use the Uniform Manifold Approximation and Projection for Dimension Reduction (UMAP) technique from McInnes et al. 2018 to perform a dimension reduction from 6D to 2D, thus mapping each combination of means, standard deviations and weights to only two adimensional parameters representing such ‘contraction’, preserving the global structure of the original parameter space. This allows to visualize the parameter space in 2D. The result of the dimension reduction process from the complete test sample can be seen in Fig. 4.6, alongside with the position of the four observed dwarf galaxies shown in the same parameter space, each indicated as colored star.

As expected, the spatial location of the points in the parameter space is strongly linked to the value of their inner slope: points with a similar inner slope cluster together, showing that the network is properly parameterizing the inner slope of galaxies during training. Interestingly, the four observed galaxies fall into the regions occupied by the simulated ones, which indicates that the model is considering them as data of equivalent nature as the test data. However, the fact that all four are close to the edges of the simulation input parameters could indicate the presence of some features that the model has not found in the simulations.

4.1.2 Deriving central DM density slopes of dSphs with CNNs

We now infer the inner slope of the observed dwarfs. Fig. 4.7 shows the posterior distributions constructed by the model for each observed galaxy. Fornax presents a very narrow peak around $\gamma = -0.38$, indicating that this galaxy has a large central DM core, while a secondary peak would give a 12% probability that the inner slope is around $\gamma = -0.81$. This is consistent with several previous works that predict a cored profile for Fornax see Goerdt, Moore, J. I. Read et al. 2006; Walker and Peñarrubia 2011; C. B. Brook and Di Cintio 2015; Pascale et al. 2018, amongst others. For the other three galaxies, a cusp is predicted with varying degrees of certainty. The model has a clear peak around $\gamma = -1.06$ for Carina, which roughly corresponds to the slope of a NFW profile at 150 pc.

Sextans presents a relatively large uncertainty in the inner slope value, as depicted by the quite broad PDFs, with a broad peak around $\gamma = -1.25$ and a strong right wing that does not fall below 10% of the peak value until it reaches $\gamma = -0.68$. Finally, Sculptor peaks at $\gamma = -1.08$, but it has a wide secondary peak, predicting a 18% probability of having a mild core with $\gamma = -0.75$. A small core was derived for Sculptor by using kinematical data and a mass-dependent profile fit in C. B. Brook and Di Cintio 2015, in agreement with the Walker and Peñarrubia 2011 and Agnello and Evans 2012 methods that, employing multiple stellar populations within a galaxy, also predicted a core in such dwarf (see also Zhu et al. 2016). Other studies, however, surprisingly predict a cusp for Sculptor after all (Richardson and Fairbairn 2014), highlighting the importance of deriving the DM density of this dSphs with several different methods.

Our derived posterior distributions offer great versatility in interpreting the results, allowing for a more complex analysis compared to models that only allow for uncertainty ranges around the inferred value.

We compare the results of our model with the inner slopes inferred for these same galaxies at 150 pc using GRAVSPHERE, a non-parametric spherical Jeans analysis code, which make use of photometric and kinematic data from the galaxies (J. I. Read, Walker et al. 2019). The inferred values, along with their 68 per cent confidence intervals (in our case, taking the primary maximum as reference), are listed in Table 4.3. The derived values are consistent between the two models, within their respective uncertainty ranges, indicating that our neural network model is making predictions similar to those obtained by Jeans analysis.

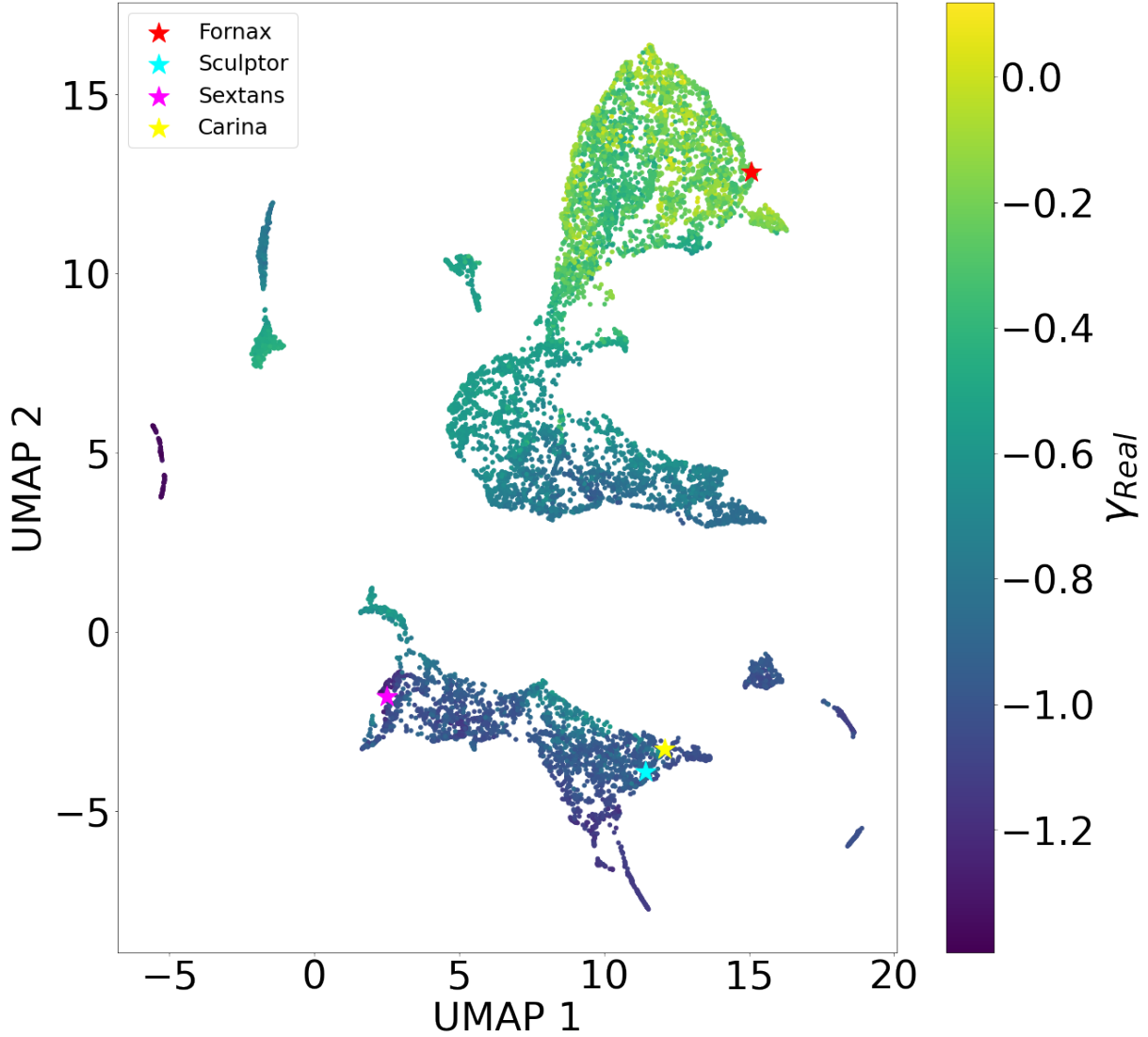


Figure 4.6: Representation of the parameter space for the test data from simulated galaxies reduced to 2-dimension with a Uniform Manifold Approximation and Projection (UMAP), color coded by the real expected inner DM slope. Plotted as colored stars are the locations in the reduced parameter space of Carina, Sextans, Sculptor and Fornax galaxies. Note that different inner slopes occupy different areas of the plot and, importantly, observed dwarf galaxies fall well within the simulation region, indicating that the neural network model is not seeing relevant differences between the simulated data with which we have fed it and the observational data.

	γ_{GS}	γ_{NN}
Carina	$-1.23^{+0.39}_{-0.35}$	$-1.06^{+0.05}_{-0.04}$
Sextans	$-0.95^{+0.25}_{-0.25}$	$-1.25^{+0.25}_{-0.09}$
Fornax	$-0.30^{+0.21}_{-0.28}$	$-0.38^{+0.01}_{-0.02}$
Sculptor	$-0.83^{+0.30}_{-0.25}$	$-1.08^{+0.08}_{-0.04}$

Table 4.3: Inner slope of the DM profile (at 150 pc) for Carina, Sextans, Sculptor and Fornax galaxies predicted by GRAVSPHERE (γ_{GS}) and our Neural Network (γ_{NN}), with their 68 per cent confidence intervals (for the Neural Network posterior, taking the primary maximum as reference).

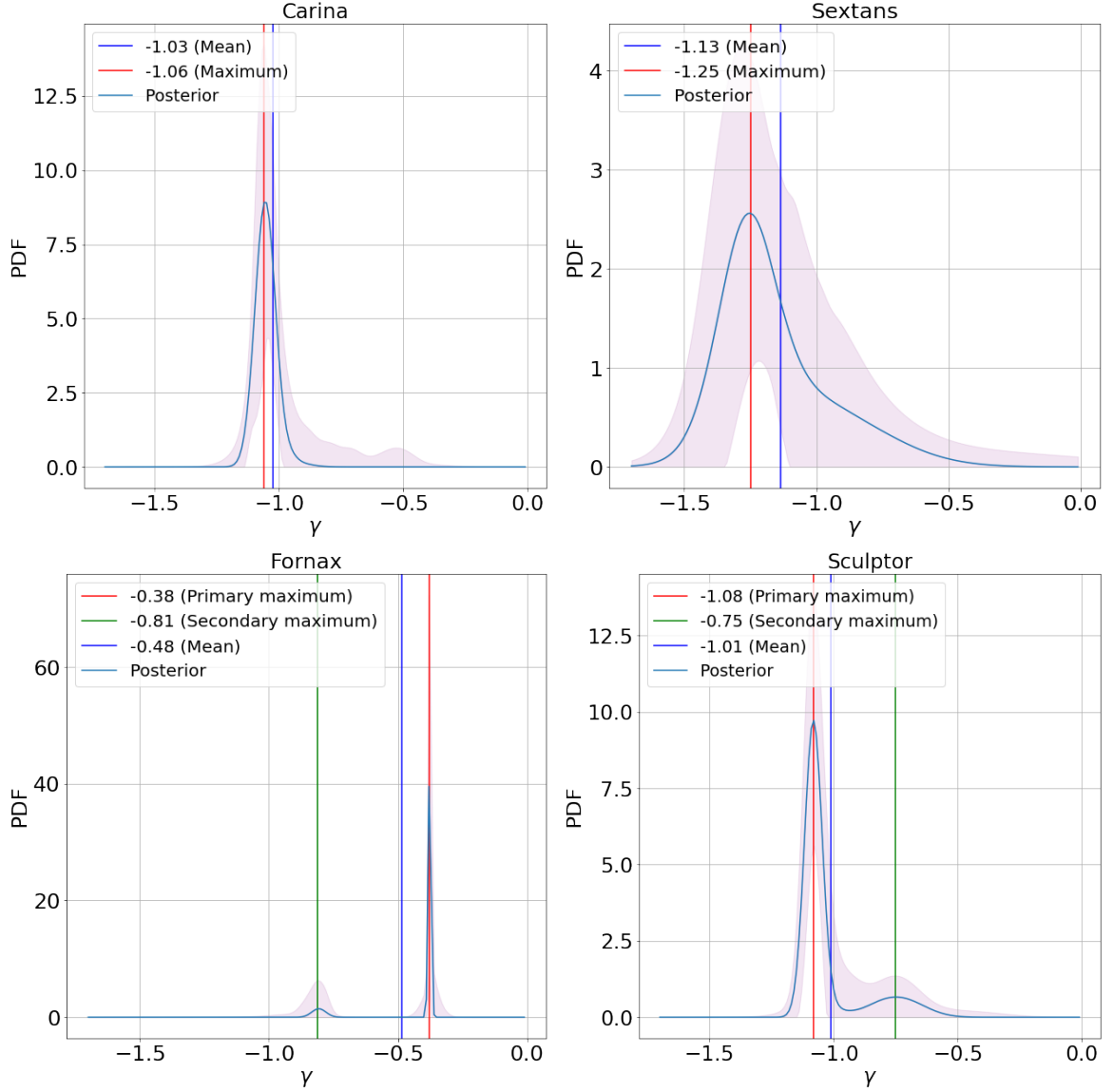


Figure 4.7: Bayesian posterior distributions in the space of inner slope of DM profiles (slope at 150 pc) predicted by our neural network model for the observed dSphs Carina, Sextans, Sculptor and Fornax. Shaded regions represents the standard deviation of the posterior values for the MC-Dropout inferences at each slope point. In each panel, the global maximum of the posterior distribution as well as the mean value are indicated, together with primary and secondary peaks when they exist. Fornax has the strongest signature of a central DM density core, while Carina has the strongest signature of having a NFW profile. Sextans is cuspy, though with a large uncertainty, while Sculptor is cuspy with a secondary peak indicating a mild core.

Compared to GRAVSPHERE and similar codes, the neural network approach is significantly faster. In a modern laptop, GRAVSPHERE will need about half a day to run an analysis of one of these galaxies, whereas the neural network can be trained with the amount of data used in this work in less than half an hour on a standard GPU. Furthermore, the training and the evaluation are independent calculation in a neural network model, which means that, once the model has been trained, its application to any input data to construct the posterior distribution is nearly instantaneous. This feature will not change no matter how much the model is expanded and complexified to perform more complete analyses of the galaxy of interest.

Chapter 5

Conclusion

We present a novel model for the determination of the inner profile of dark matter halos with robust uncertainty quantification using machine learning techniques. The goal of this work is to be able to infer central density slopes (γ) of the dark matter density profile of galaxies by just using positional and spectroscopic data of its stars. Our method uses convolutional mixture density neural networks (CMDNNs) with a mixture density layer backend to model complex galaxy substructure. We use line-of-sight velocities and positions of stars within simulated dwarf galaxies, employing Kernel Density Estimations (KDEs) to construct continuous 2D probability density functions (PDFs) of the distribution of such stars in $\{\hat{R}_{\text{proj}}, \hat{v}_{\text{LOS}}\}$ and $\{x, y\}$ spaces, which serve as input to our neural network using a double channel architecture (Fig. 3.1 and 3.2).

We train and evaluate our model using a large set of fully cosmological simulations of dwarf galaxies with halo masses of 10^9 to $10^{11.5} M_{\odot}$, and stellar masses of 10^5 to $10^{9.5} M_{\odot}$ from the NIHAO and AURIGA projects (Wang et al. 2015; A. A. Dutton et al. 2020; Grand et al. 2017). The use of different physical models employed in such simulations allows us to have a diverse range of density profiles at each particular galaxy mass, including both cores and cusps (Fig. 2.2). All simulated galaxies and their subsets of stars are randomly projected in several sky planes, to simulate several viewing angles.

The loss function to minimize during the training is the negative logarithmic likelihood of the training sample, defined as a multivariate Gaussian probability distribution, which is the output of our Gaussian density layer backend. This allows a flexible probabilistic representation of the results, which yields accurate and statistically consistent uncertainties. For each galaxy the neural network outputs a PDF which gives the posterior probability of a certain slope to be the inner slope of the galaxy. The main results of this work are listed here:

- The inner slope of simulated galaxies is predicted with a mean error of $\mu_{\epsilon} = 0.055$ (where the error is defined as $\epsilon = \gamma_{\text{Real}} - \gamma_{\text{Pred}}$, and the predicted inner slope, γ_{Pred} , is obtained from the mode of the PDFs) and a standard deviation of $\sigma_{\epsilon} = 0.075$ for the whole sample (Fig. 4.1 and 4.2);
- 82% (98%) of the galaxies have their inner slope correctly determined within ± 0.1 (0.3) of their true value (Table 4.1);
- The posteriors PDFs have a mean standard deviation of $\sigma_{\text{pos}} = 0.108$, showing no bias towards more accuracy for cuspy or cored galaxies (Fig. 4.3);
- While in most cases the output of the model is a single peaked PDF, in $\sim 30\%$ of the galaxies some of their projections show a double peak: we demonstrated that this is associated to a particular viewing angle, indicating the importance of properly determining the inclination of galaxies (Fig. 4.4 and 4.5);

- When applied to a set of four observed dSphs, our model recovers their inner slopes yielding values consistent with those obtained with the Jeans modelling codes GRAVSPHERE (Table 4.3);
- We found that Fornax dSph has a strong indication of having a central DM core, Carina and Sextans have cusps (although the latter with a large uncertainty), and Sculptor shows a double peaked PDF indicating that a cusp is preferred, but a core can not be ruled out (Fig. 4.7). These results are in agreement with several previously derived inner slopes for these galaxies (Goerd, Moore, J. I. Read et al. 2006; Agnello and Evans 2012; C. B. Brook and Di Cintio 2015; Walker and Peñarrubia 2011; Richardson and Fairbairn 2014; J. I. Read, Walker et al. 2019).

The current architecture could be used as a basis for building models that provide a more complete output, such as a prediction of the full density profile of galaxies. The nature of the neural network allows it to be constantly extended and improved. While we have implemented a network of relatively low complexity, there are a series of interesting possibilities with a further level of sophistication that are worth exploring. For example, the use of normalizing flows may yield to more robust results (Kodi Ramanah, Wojtak, Ansari et al. 2020) while the use of a 3D convolutional network applied to PDFs defined in the $\{x, y, \hat{v}_{\text{LOS}}\}$ phase space has given good results in galaxy cluster masses inference (Kodi Ramanah, Wojtak and Arendse 2020).

In the future, the architecture of this model could also be expanded with more input data, such as proper motion of the stars from missions like *GAIA* (Gaia Collaboration 2021). Furthermore, the inclusion of other spectroscopic samples present in the literature, as well as of those soon to be acquired with upcoming facilities, will certainly be beneficial for this analysis.

Adapting the architecture and introducing more information may enable the network to improve accuracy and reduce the range of variability of the results with respect to the angle of observation.

This newly developed neural network method is a promising tool for the study of properties of dwarf galaxies. We have shown that deep learning techniques provide a innovative method for the determination of the inner DM profiles in dwarf galaxies, complementary to the use of Jeans modelling.

Acknowledgements

I am grateful to Chris Brook for his feedback, ideas and work as basically an unofficial supervisor, as well as to Giuseppina Battaglia for her help with the processing of the observational data. I also thank Andrea Maccio, Rob Grant and the entire NIHAO and AURIGA teams for providing the simulations used in this work.

References

- 1 A. AGNELLO and N. W. EVANS:
'A VIRIAL CORE IN THE SCULPTOR DWARF SPHEROIDAL GALAXY',
[The Astrophysical Journal](#) **754**, L39 (2012).
- 2 G. BATTAGLIA, A. HELMI, E. TOLSTOY, M. IRWIN, V. HILL and P. JABLONKA:
'The Kinematic Status and Mass Content of the Sculptor Dwarf Spheroidal Galaxy',
[MNRAS](#) **381**, L13 (2008).
- 3 G. BATTAGLIA, S. TAIBI, G. F. THOMAS and T. K. FRITZ:
'Gaia early DR3 systemic motions of Local Group dwarf galaxies and orbital properties with a massive Large Magellanic Cloud',
[MNRAS](#) **517**, A54, A54 (2022).
- 4 C. B. BROOK and A. DI CINTIO:
'Expanded haloes, abundance matching and too-big-to-fail in the Local Group',
[MNRAS](#) **450**, 3920–3934 (2015).
- 5 T. BUCK, M. NESS, A. OBREJA, A. V. MACCIÒ and A. A. DUTTON:
'Stars behind bars II: a cosmological formation scenario for the milky way's central stellar structure',
[The Astrophysical Journal](#) **874**, 67 (2019).
- 6 J. S. BULLOCK and M. BOYLAN-KOLCHIN:
'Small-Scale Challenges to the Λ CDM Paradigm',
[MNRAS](#) **471**, 343–387 (2017).
- 7 F. CHOLLET ET AL.:
Keras,
(2015) <https://github.com/fchollet/keras>.
- 8 P. COLABORATION:
'*planck* 2013 results. XVI. cosmological parameters',
[Astronomy & Astrophysics](#) **571**, A16 (2014).
- 9 D. R. COLE, W. DEHNEN and M. I. WILKINSON:
'Weakening dark matter cusps by clumpy baryonic infall',
[Monthly Notices of the Royal Astronomical Society](#) **416**, 1118–1134 (2011).
- 10 G. COLLABORATION:
'*gaia* early data release 3',
[Astronomy & Astrophysics](#) **650**, C3 (2021).
- 11 M. L. M. COLLINS, J. I. READ, R. A. IBATA, R. M. RICH, N. F. MARTIN, J. PEÑARRUBIA, S. C. CHAPMAN, E. J. TOLLERUD and D. R. WEISZ:
'Andromeda XXI - a dwarf galaxy in a low-density dark matter halo',
[MNRAS](#) **505**, 5686–5701 (2021).
- 12 W. J. G. DE BLOK, F. WALTER, E. BRINKS, C. TRACHTERNACH, S.-H. OH and R. C. KENNICUTT JR.:
'High-Resolution Rotation Curves and Galaxy Mass Models from THINGS',
[MNRAS](#) **386**, 2648, 2648–2719 (2008).
- 13 A. DI CINTIO, C. B. BROOK, A. A. DUTTON, A. V. MACCIÒ, G. S. STINSON and A. KNEBE:
'A mass-dependent density profile for dark matter haloes including the influence of galaxy formation',
[MNRAS](#) **441**, 2986–2995 (2014).
- 14 A. DI CINTIO, C. B. BROOK, A. V. MACCIÒ, G. S. STINSON, A. KNEBE, A. A. DUTTON and J. WADSLEY:
'The dependence of dark matter profiles on the stellar-to-halo mass ratio: a prediction for cusps versus cores',
[MNRAS](#) **437**, 415–423 (2014).
- 15 F. I. DIAKOIANNIS, G. F. LEWIS and R. A. IBATA:
'Resolving the mass–anisotropy degeneracy of the spherically symmetric jeans equation – i. theoretical foundation',
[Monthly Notices of the Royal Astronomical Society](#) **443**, 598–609 (2014).
- 16 A. A. DUTTON, T. BUCK, A. V. MACCIÒ, K. L. DIXON, M. BLANK and A. OBREJA:
'NIHAO - XXV. Convergence in the cusp-core transformation of cold dark matter haloes at high star formation thresholds',
[MNRAS](#) **499**, 2648–2661 (2020).
- 17 A. DUTTON, A. OBREJA, L. WANG, T. GUTCKE, T. BUCK, S. UDRESCU, J. FRINGS, G. STINSON, X. KANG and A. MACCIÒ:
'Nihao xii: galactic uniformity in a Λ CDM universe',
[Monthly Notices of the Royal Astronomical Society](#) **467**, 4937–4950 (2017).
- 18 N. W. EVANS, J. AN and M. G. WALKER:
'Cores and cusps in the dwarf spheroidals',
[Monthly Notices of the Royal Astronomical Society: Letters](#) **393**, L50–L54 (2009).
- 19 GAIA COLLABORATION:
'Gaia Early Data Release 3. Summary of the contents and survey properties',
[MNRAS](#) **507**, A1, A1 (2021).
- 20 Y. GAL and Z. GHAHRAMANI:
Dropout as a bayesian approximation: representing model uncertainty in deep learning,
2015.

- 21 A. GENINA, A. BENÍ TEZ-LLAMBAY, C. S. FRENK, S. COLE, A. FATAHI, J. F. NAVARRO, K. A. OMAN, T. SAWALA and T. THEUNS:
'The core-cusp problem: a matter of perspective', [Monthly Notices of the Royal Astronomical Society](#) **474**, 1398–1411 (2017).
- 22 A. GHOSH, C. M. URRY, A. RAU, L. PERREAULT-LEVASSEUR, M. CRANMER, K. SCHAWINSKI, D. STARK, C. TIAN, R. OFMAN, T. T. ANANNA, C. AUJE, N. CAPPELLUTI, D. B. SANDERS and E. TREISTER:
Gampen: a machine learning framework for estimating bayesian posteriors of galaxy morphological parameters, 2022.
- 23 X. GLOROT, A. BORDES and Y. BENGIO:
'Deep sparse rectifier neural networks', in, Vol. 15 (2010).
- 24 O. Y. GNEDIN and H. ZHAO:
'Maximum feedback and dark matter profiles of dwarf galaxies', [Monthly Notices of the Royal Astronomical Society](#) **333**, 299–306 (2002).
- 25 T. GOERDT, B. MOORE, J. READ and J. STADEL:
'Core creation in galaxies and halos via sinking massive objects', [Astrophysical Journal - ASTROPHYS J](#) **725**, 1707–1716 (2010).
- 26 T. GOERDT, B. MOORE, J. I. READ, J. STADEL and M. ZEMP:
'Does the Fornax dwarf spheroidal have a central cusp or core?', [Monthly Notices of the Royal Astronomical Society](#) **368**, 1073–1077 (2006).
- 27 F. GOVERNATO, C. BROOK, L. MAYER, A. BROOKS, G. RHEE, J. WADSLEY, P. JONSSON, B. WILLMAN, G. STINSON, T. QUINN and P. MADAU:
'Bulgeless dwarf galaxies and dark matter cores from supernova-driven outflows', [Nature](#) **463**, 203–6 (2010).
- 28 F. GOVERNATO, A. ZOLOTOV, A. PONTZEN, C. CHRISTENSEN, S. H. OH, A. M. BROOKS, T. QUINN, S. SHEN and J. WADSLEY:
'Cuspy no more: how outflows affect the central dark matter and baryon distribution in Λ cold dark matter galaxies', **422**, 1231–1240 (2012).
- 29 R. J. J. GRAND, F. A. GÓMEZ, F. MARINACCI, R. PAKMOR, V. SPRINGEL, D. J. R. CAMPBELL, C. S. FRENK, A. JENKINS and S. D. M. WHITE:
'The Auriga Project: the properties and formation mechanisms of disc galaxies across cosmic time', [Monthly Notices of the Royal Astronomical Society](#) **467**, 179–207 (2017).
- 30 H. GUO, Z. ZHENG, P. S. BEHROOZI, I. ZEHAVID, C.-H. CHUANG, J. COMPARAT, G. FAVOLE, S. GOTTLÖBER, A. KLYPIN, F. PRADA, S. A. RODRÍGUEZ-TORRES, D. H. WEINBERG and G. YEPES:
'Modelling galaxy clustering: halo occupation distribution versus subhalo matching', **459**, 3040–3058 (2016).
- 31 M. HO, M. M. RAU, M. NTAMPAKA, A. FARAHI, H. TRAC and B. PÓCZOS:
'A robust and efficient deep learning method for dynamical mass measurements of galaxy clusters', [The Astrophysical Journal](#) **887**, 25 (2019).
- 32 M. KAPLINGHAT, S. TULIN and H.-B. YU:
'Dark matter halos as particle colliders: unified solution to small-scale structure puzzles from dwarfs to clusters', [Phys. Rev. Lett.](#) **116**, 041302 (2016).
- 33 D. P. KINGMA and J. BA:
Adam: a method for stochastic optimization, 2014.
- 34 D. KODI RAMANAH, R. WOJTAK, Z. ANSARI, C. GALL and J. HJORTH:
'Dynamical mass inference of galaxy clusters with neural flows', [Monthly Notices of the Royal Astronomical Society](#) **499**, 1985–1997 (2020).
- 35 D. KODI RAMANAH, R. WOJTAK and N. ARENDSE:
'Simulation-based inference of dynamical galaxy cluster masses with 3D convolutional neural networks', [Monthly Notices of the Royal Astronomical Society](#) **501**, 4080–4091 (2020).
- 36 Y. LECUN, Y. BENGIO and G. HINTON:
'Deep learning', [nature](#) **521**, 436–444 (2015).
- 37 E. L. ŁOKAS and G. A. MAMON:
'Dark matter distribution in the coma cluster from galaxy kinematics: breaking the mass-anisotropy degeneracy', [Monthly Notices of the Royal Astronomical Society](#) **343**, 401–412 (2003).
- 38 V. LORA, E. K. GREBEL, F. J. SÁ NCHEZ-SALCEDO and A. JUST:
'SEXTANS' COLD SUBSTRUCTURES AS a DYNAMICAL JUDGE: CORE, CUSP, OR MOND?', [The Astrophysical Journal](#) **777**, 65 (2013).
- 39 A. V. MACCIÒ, S. PADUROIU, D. ANDERHALDEN, A. SCHNEIDER and B. MOORE:
'Cores in warm dark matter haloes: a Catch 22 problem', **424**, 1105–1112 (2012).
- 40 A. V. MACCIÒ, S. M. UDRESCU, A. A. DUTTON, A. OBREJA, L. WANG, G. R. STINSON and X. KANG:
'NIHAO X: reconciling the local galaxy velocity function with cold dark matter via mock H I observations', [Monthly Notices of the Royal Astronomical Society: Letters](#) **463**, L69–L73 (2016).
- 41 R. P. VAN DER MAREL:
'Velocity profiles of galaxies with claimed black holes – III. Observations and models for M87', [Monthly Notices of the Royal Astronomical Society](#) **270**, 271–297 (1994).

- 42 A. MATHIEU and M. R. MERRIFIELD:
'From kinematics to dynamics in thin galactic discs',
[318](#), 40–46 (2000).
- 43 L. MCINNES, J. HEALY and J. MELVILLE:
Umap: uniform manifold approximation and projection for dimension reduction,
2018.
- 44 M. R. MERRIFIELD and S. M. KENT:
'Fourth Moments and the Dynamics of Spherical Systems',
[99](#), 1548 (1990).
- 45 B. MOORE:
'Evidence against dissipation-less dark matter from observations of galaxy haloes',
[370](#), 629–631 (1994).
- 46 B. P. MOSTER, T. NAAB and S. D. M. WHITE:
'Galactic star formation and accretion histories from matching galaxies to dark matter haloes',
[428](#), 3121–3138 (2013).
- 47 B. P. MOSTER, R. S. SOMERVILLE, C. MAULBETSCH, F. C. VAN DEN BOSCH, A. V. MACCIÒ, T. NAAB and L. OSER:
'Constraints on the Relationship between Stellar Mass and Halo Mass at Low and High Redshift',
[710](#), 903–923 (2010).
- 48 J. F. NAVARRO, V. R. EKE and C. S. FRENK:
'The cores of dwarf galaxy haloes',
[283](#), L72–L78 (1996).
- 49 J. F. NAVARRO, C. S. FRENK and S. D. M. WHITE:
'The structure of cold dark matter halos',
[The Astrophysical Journal](#) **462**, 563 (1996).
- 50 J. F. NAVARRO, A. LUDLOW, V. SPRINGEL, J. WANG, M. VOGELSBERGER, S. D. M. WHITE, A. JENKINS, C. S. FRENK and A. HELMI:
'The diversity and similarity of simulated cold dark matter haloes',
[402](#), 21–34 (2010).
- 51 M. NTAMPAKA, H. TRAC, D. J. SUTHERLAND, S. FROMENTEAU, B. PÓCZOS and J. SCHNEIDER:
'Dynamical Mass Measurements of Contaminated Galaxy Clusters Using Machine Learning',
[831](#), 135, 135 (2016).
- 52 R. PASCALE, L. POSTI, C. NIPOTI and J. BINNEY:
'Action-based dynamical models of dwarf spheroidal galaxies: application to Fornax',
[480](#), 927–946 (2018).
- 53 PLANCK COLLABORATION:
'Planck 2015 results - xiii. cosmological parameters',
[A&A](#) **594**, A13 (2016).
- 54 A. PONTZEN and F. GOVERNATO:
'How supernova feedback turns dark matter cusps into cores',
[421](#), 3464–3471 (2012).
- 55 J. I. READ, M. G. WALKER and P. STEGER:
'Dark matter heats up in dwarf galaxies',
[Monthly Notices of the Royal Astronomical Society](#) **484**, 1401–1420 (2019).
- 56 J. I. READ and G. GILMORE:
'Mass loss from dwarf spheroidal galaxies: the origins of shallow dark matter cores and exponential surface brightness profiles',
[356](#), 107–124 (2005).
- 57 J. I. READ, G. IORIO, O. AGERTZ and F. FRATERNALI:
'The stellar mass–halo mass relation of isolated field dwarfs: a critical test of Λ CDM at the edge of galaxy formation',
[Monthly Notices of the Royal Astronomical Society](#) **467**, 2019–2038 (2017).
- 58 J. I. READ and P. STEGER:
'How to break the density-anisotropy degeneracy in spherical stellar systems',
[Monthly Notices of the Royal Astronomical Society](#) **471**, 4541–4558 (2017).
- 59 T. RICHARDSON and M. FAIRBAIRN:
'On the dark matter profile in Sculptor: breaking the β degeneracy with Virial shape parameters',
[441](#), 1584–1600 (2014).
- 60 V. H. ROBLES, J. S. BULLOCK, O. D. ELBERT, A. FITTS, A. GONZÁLEZ-SAMANIEGO, M. BOYLAN-KOLCHIN, P. F. HOPKINS, C.-A. FAUCHER-GIGUÈRE, D. KEREŠ and C. C. HAYWARD:
'SIDM on fire: hydrodynamical self-interacting dark matter simulations of low-mass dwarf galaxies',
[Monthly Notices of the Royal Astronomical Society](#) **472**, 2945–2954 (2017).
- 61 R. P. SAGLIA, A. KRONAWITTER, O. GERHARD and R. BENDER:
'The orbital structure and potential of NGC 1399',
[The Astronomical Journal](#) **119**, 153–161 (2000).
- 62 A. SCHNEIDER, S. TRUJILLO-GOMEZ, E. PAPASTERGIS, D. REED and G. LAKE:
'Hints against the cold and collisionless nature of dark matter from the galaxy velocity function', English,
[Monthly Notices of the Royal Astronomical Society](#) **470**, 1542–1558 (2017).
- 63 D. SCOTT:
Multivariate density estimation: theory, practice, and visualization,
(John Wiley Sons, Inc., 1992).
- 64 S. J. SHEATHER:
'Density estimation',
[Statistical Science](#) **19**, 588–597 (2004).
- 65 J. D. SIMON, A. D. BOLATTO, A. LEROY, L. BLITZ and E. L. GATES:
'High-resolution measurements of the halos of four dark matter-dominated galaxies: deviations from a universal density profile',
[The Astrophysical Journal](#) **621**, 757–776 (2005).
- 66 D. N. SPERGEL and P. J. STEINHARDT:
'Observational evidence for self-interacting cold dark matter',
[Phys. Rev. Lett.](#) **84**, 3760–3763 (2000).

- 67 STINSON, BROOK, MACCIÒ, WADSLEY, QUINN and COUCHMAN:
'Making galaxies in a cosmological context: the need for early stellar feedback',
[Monthly Notices of the Royal Astronomical Society](#) **428**, 129–140 (2013).
- 68 SZABÓ, R., SZKLENÁR, T. and BÓDI, A.:
'Machine learning in present day astrophysics',
[Europhysics News](#) **53**, 22–25 (2022).
- 69 E. TOLLET, A. V. MACCIÒ, A. A. DUTTON, G. S. STINSON, L. WANG, C. PENZO, T. A. GUTCKE, T. BUCK, X. KANG, C. BROOK, A. D. CINTIO, B. W. KELLER and J. WADSLEY:
'NIHAO – IV: core creation and destruction in dark matter density profiles across cosmic time',
[Monthly Notices of the Royal Astronomical Society](#) **456**, 3542–3552 (2016).
- 70 R. P. VAN DER MAREL and M. FRANX:
'A New Method for the Identification of Non-Gaussian Line Profiles in Elliptical Galaxies',
[MNRAS](#) **407**, 525 (1993).
- 71 M. G. WALKER and J. PEÑARRUBIA:
'A Method for Measuring (Slopes of) the Mass Profiles of Dwarf Spheroidal Galaxies',
[MNRAS](#) **742**, 20, 20 (2011).
- 72 M. G. WALKER, M. MATEO and E. W. OLSZEWSKI:
'Stellar Velocities in the Carina, Fornax, Sculptor, and Sextans dSph Galaxies: Data From the Magellan/MMFS Survey',
[MNRAS](#) **137**, 3100–3108 (2009).
- 73 L. WANG, A. A. DUTTON, G. S. STINSON, A. V. MACCIÒ, C. PENZO, X. KANG, B. W. KELLER and J. WADSLEY:
'NIHAO project - I. Reproducing the inefficiency of galaxy formation across cosmic time with a large sample of cosmological hydrodynamical simulations',
[MNRAS](#) **454**, 83–94 (2015).
- 74 L. L. WATKINS, G. VAN DE VEN, M. DEN BROK and R. C. E. VAN DEN BOSCH:
'Discrete dynamical models of ω Centauri',
[MNRAS](#) **436**, 2598–2615 (2013).
- 75 R. WOJTAK, E. L. ŁOKAS, G. A. MAMON and S. GOTTLÖBER:
'The mass and anisotropy profiles of galaxy clusters from the projected phase-space density: testing the method on simulated data',
[Monthly Notices of the Royal Astronomical Society](#) **399**, 812–821 (2009).
- 76 A. EL-ZANT, I. SHLOSMAN and Y. HOFFMAN:
'Dark halos: the flattening of the density cusp by dynamical friction',
[The Astrophysical Journal](#) **560**, 636–643 (2001).
- 77 L. ZHU, G. VAN DE VEN, L. L. WATKINS and L. POSTI:
'A discrete chemo-dynamical model of the dwarf spheroidal galaxy Sculptor: mass profile, velocity anisotropy and internal rotation',
[MNRAS](#) **463**, 1117–1135 (2016).
- 78 A. ZOLOTOV, A. M. BROOKS, B. WILLMAN, F. GOVERNATO, A. PONTZEN, C. CHRISTENSEN, A. DEKEL, T. QUINN, S. SHEN and J. WADSLEY:
'BARYONS MATTER: WHY LUMINOUS SATELLITE GALAXIES HAVE REDUCED CENTRAL MASSES',
[The Astrophysical Journal](#) **761**, 71 (2012).