

Euclid preparation

LXVIII. Extracting physical parameters from galaxies with machine learning

Euclid Collaboration: I. Kovačić¹, M. Baes¹, A. Nersesian^{2,1}, N. Andreadis¹, L. Nemani³, Abdurro'uf⁴, L. Bisigello^{5,6}, M. Bolzonella⁷, C. Tortora⁸, A. van der Wel¹, S. Cuvuoti^{8,9}, C. J. Conselice¹⁰, A. Enia^{11,7}, L. K. Hunt¹², P. Iglesias-Navarro^{13,14}, E. Iodice⁸, J. H. Knapen^{13,14}, F. R. Marleau¹⁵, O. Müller¹⁶, R. F. Peletier¹⁷, J. Román^{14,13}, R. Ragusa⁸, P. Salucci^{18,19}, T. Saifollahi²⁰, M. Scodeggio²¹, M. Siudek^{22,23}, T. De Waele¹, A. Amara²⁴, S. Andreon²⁵, N. Auricchio⁷, C. Baccigalupi^{26,27,19,18}, M. Baldi^{11,7,28}, S. Bardelli⁷, P. Battaglia⁷, R. Bender^{29,30}, C. Bodendorf²⁹, D. Bonino³¹, W. Bon³², E. Branchini^{33,34,25}, M. Brescia^{35,8,9}, J. Brinchmann^{36,37}, S. Camera^{38,39,31}, V. Capobianco³¹, C. Carbone²¹, J. Carretero^{40,41}, S. Casas⁴², F. J. Castander^{23,43}, M. Castellano³, G. Castignani⁷, A. Cimatti⁴⁴, C. Colodro-Conde¹³, G. Congedo⁴⁵, L. Conversi^{46,47}, Y. Copin⁴⁸, F. Courbin^{16,49,50}, H. M. Courtois⁵¹, A. Da Silva^{52,53}, H. Degaudenzi⁵⁴, G. De Lucia²⁷, A. M. Di Giorgio⁵⁵, J. Dinis^{52,53}, M. Douspis⁵⁶, F. Dubath⁵⁴, X. Dupac⁴⁷, S. Dusini⁵⁷, A. Ealet⁴⁸, M. Farina⁵⁵, S. Farrens⁵⁸, F. Faustini^{59,3}, S. Ferriol⁴⁸, P. Fosalba^{43,23}, M. Frailis²⁷, E. Franceschi⁷, S. Galeotta²⁷, B. Gillis⁴⁵, C. Giocoli^{7,60}, A. Grazian⁶¹, F. Grupp^{29,30}, L. Guzzo^{62,25}, S. V. H. Haugan⁶³, W. Holmes⁶⁴, I. Hook⁶⁵, F. Hormuth⁶⁶, A. Hornstrup^{67,68}, K. Jahnke⁶⁹, M. Jhabvala⁷⁰, B. Joachimi⁷¹, E. Keihänen⁷², S. Kermiche⁷³, A. Kiessling⁶⁴, M. Kilbinger⁵⁸, B. Kubik⁴⁸, K. Kuijken⁷⁴, M. Kümmel³⁰, M. Kunz⁷⁵, H. Kurki-Suonio^{76,77}, S. Ligori³¹, P. B. Lilje⁶³, V. Lindholm^{76,77}, I. Lloro⁷⁸, D. Maino^{62,21,79}, E. Maiorano⁷, O. Mansutti²⁷, S. Marcin⁸⁰, O. Marggraf⁸¹, K. Markovic⁶⁴, M. Martinelli^{3,82}, N. Martinet³², F. Marulli^{83,7,28}, R. Massey⁸⁴, E. Medinaceli⁷, S. Mei⁸⁵, M. Melchior⁸⁰, Y. Mellier^{86,87}, M. Meneghetti^{7,28}, E. Merlin³, G. Meylan¹⁶, M. Moresco^{83,7}, L. Moscardini^{83,7,28}, S.-M. Niemi⁸⁸, J. W. Nightingale⁸⁹, C. Padilla⁹⁰, S. Paltani⁵⁴, F. Pasian²⁷, K. Pedersen⁹¹, V. Pettorino⁸⁸, S. Pires⁵⁸, G. Polenta⁵⁹, M. Poncet⁹², L. A. Popa⁹³, L. Pozzetti⁷, F. Raison²⁹, R. Rebolo^{13,14,94}, A. Renzi^{6,57}, J. Rhodes⁶⁴, G. Riccio⁸, E. Romelli²⁷, M. Roncarelli⁷, E. Rossetti¹¹, R. Saglia^{30,29}, Z. Sakr^{95,96,97}, A. G. Sánchez²⁹, D. Sapone⁹⁸, B. Sartoris^{30,27}, M. Schirmer⁶⁹, P. Schneider⁸¹, T. Schrabback¹⁵, A. Secroun⁷³, G. Seidel⁶⁹, S. Serrano^{43,99,23}, C. Sirignano^{6,57}, G. Sirri²⁸, L. Stanco⁵⁷, J. Steinwagner²⁹, P. Tallada-Crespí^{40,41}, D. Tavagnacco²⁷, A. N. Taylor⁴⁵, H. I. Teplitz¹⁰⁰, I. Tereno^{52,101}, R. Toledo-Moreo¹⁰², F. Torradeflot^{41,40}, I. Tutusaus⁹⁶, L. Valenziano^{7,103}, T. Vassallo^{30,27}, G. Verdoes Kleijn¹⁷, A. Veropalumbo^{25,34,104}, Y. Wang¹⁰⁰, J. Weller^{30,29}, A. Zacchei^{27,26}, G. Zamorani⁷, E. Zucca⁷, A. Biviano^{27,26}, E. Bozzo⁵⁴, C. Burigana^{5,103}, M. Calabrese^{105,21}, D. Di Ferdinando²⁸, J. A. Escartin Vigo²⁹, F. Finelli^{7,103}, J. Gracia-Carpio²⁹, S. Matthew⁴⁵, N. Mauri^{44,28}, M. Pöntinen⁷⁶, V. Scottez^{86,106}, M. Tenti²⁸, M. Viel^{26,27,18,19,107}, M. Wiesmann⁶³, Y. Akrami^{108,109}, V. Allevato⁸, S. Alvi¹¹⁰, S. Anselmi^{57,6,111}, M. Archidiacono^{62,79}, F. Atrio-Barandela¹¹², M. Ballardini^{110,7,113}, M. Bethermin^{20,32}, L. Blot^{114,111}, S. Borgani^{115,26,27,19}, S. Bruton¹¹⁶, R. Cabanac⁹⁶, A. Calabro³, B. Camacho Quevedo^{43,23}, G. Cañas-Herrera^{88,117}, A. Cappi^{7,118}, F. Caro³, C. S. Carvalho¹⁰¹, T. Castro^{27,19,26,107}, K. C. Chambers¹¹⁹, T. Contini⁹⁶, A. R. Cooray¹²⁰, O. Cucciati⁷, G. Desprez¹²¹, A. Díaz-Sánchez¹²², J. J. Diaz²², S. Di Domizio^{33,34}, H. Dole⁵⁶, S. Escoffier⁷³, A. G. Ferrari^{44,28}, P. G. Ferreira¹²³, I. Ferrero⁶³, A. Finoguenov⁷⁶, A. Fontana³, F. Fornari¹⁰³, L. Gabarra¹²³, K. Ganga⁸⁵, J. García-Bellido¹⁰⁸, T. Gasparetto²⁷, V. Gautard¹²⁴, E. Gaztanaga^{23,43,125}, F. Giacomini²⁸, F. Gianotti⁷, G. Gozalias^{126,76}, C. M. Gutierrez¹²⁷, A. Hall⁴⁵, S. Hemmati¹²⁸, H. Hildebrandt¹²⁹, J. Hjorth⁹¹, A. Jimenez Muñoz¹³⁰, J. J. E. Kajava^{131,132}, V. Kansal^{133,134}, D. Karagiannis^{135,136}, C. C. Kirkpatrick⁷², A. M. C. Le Brun¹¹¹, J. Le Graet⁷³, J. Lesgourgues⁴², T. I. Liaudat¹³⁷, A. Loureiro^{138,139}, J. Macias-Perez¹³⁰, G. Maggio²⁷, M. Magliocchetti⁵⁵, F. Mannucci¹², R. Maoli^{140,3}, J. Martín-Fleitas¹⁴¹, C. J. A. P. Martins^{142,36}, L. Maurin⁵⁶, R. B. Metcalfe^{83,7}, M. Miluzio^{47,143}, P. Monaco^{115,27,19,26}, A. Montoro^{23,43}, A. Mora¹⁴¹, C. Moretti^{18,107,27,26,19}, G. Morgante⁷, Nicholas A. Walton¹⁴⁴, L. Patrizii²⁸, V. Popa⁹³, D. Potter¹⁴⁵, I. Risso¹⁴⁶, P.-F. Rocci⁵⁶, M. Sahlén¹⁴⁷, E. Sarpa^{18,107,19}, C. Scarlata¹¹⁶, A. Schneider¹⁴⁵, M. Sereno^{7,28}, F. Shankar¹⁴⁸, P. Simon⁸¹, A. Spurio Mancini^{149,150}, J. Stadel¹⁴⁵, S. A. Stanford¹⁵¹, K. Tanidis¹²³, C. Tao⁷³, G. Testera³⁴, R. Teyssier¹⁵², S. Toft^{153,154}, S. Tosi^{33,34}, A. Troja^{6,57}, M. Tucci⁵⁴, C. Valieri²⁸, J. Valiviita^{76,77}, D. Vergani⁷, G. Verza^{155,156}, and P. Vielzeuf⁷³

(Affiliations can be found after the references)

ABSTRACT

The *Euclid* mission is generating a vast amount of imaging data in four broadband filters at a high angular resolution. This data will allow for the detailed study of mass, metallicity, and stellar populations across galaxies that will constrain their formation and evolutionary pathways. Transforming the *Euclid* imaging for large samples of galaxies into maps of physical parameters in an efficient and reliable manner is an outstanding challenge. Here, we investigate the power and reliability of machine learning techniques to extract the distribution of physical parameters within well-resolved galaxies. We focus on estimating stellar mass surface density, mass-averaged stellar metallicity, and age. We generated noise-free synthetic high-resolution ($100\text{ pc} \times 100\text{ pc}$) imaging data in the *Euclid* photometric bands for a set of 1154 galaxies from the TNG50 cosmological simulation. The images were generated with the SKIRT radiative transfer code, taking into account the complex 3D distribution of stellar populations and interstellar dust attenuation. We used a machine learning framework to map the idealised mock observational data to the physical parameters on a pixel-by-pixel basis. We find that stellar mass surface density can be accurately recovered with a ≤ 0.130 dex scatter. Conversely, stellar metallicity and age estimates are, as expected, less robust, but they still contain significant information that originates from underlying correlations at a sub-kiloparsec scales between stellar mass surface density and stellar population properties. As a corollary, we show that TNG50 follows a spatially resolved mass-metallicity relation that is consistent with observations. Due to its relatively low computational and time requirements, which has a time-frame of minutes without dedicated high performance computing infrastructure once it has been trained, our method allows for fast and robust estimates of the stellar mass surface density distributions of nearby galaxies from four-filter *Euclid* imaging data. Equivalent estimates of stellar population properties (stellar metallicity and age) are less robust but still hold value as first-order approximations across large samples.

Key words. galaxies: photometry, galaxies: general, methods: statistics

1. Introduction

Understanding the underlying processes that drive galactic evolution through cosmic time has been one of the burning issues in astrophysics for the past decades. An important class of constraints for galaxy evolution theories is scaling relations between the physical properties of galaxies. For example, galaxy stellar mass, considered one of the fundamental parameters of galaxies, has been found to be tightly correlated with star-formation rate (SFR; Brinchmann et al. 2004; Whitaker et al. 2012; Speagle et al. 2014; Watkins et al. 2022; Chamba et al. 2022; Popesso et al. 2023) as well as with gas phase (Lequeux et al. 1979; Tremonti et al. 2004; Ly et al. 2016; Zahid et al. 2017) and stellar metallicity (Kauffmann et al. 2003; Gallazzi et al. 2005). These correlations are well reproduced by observations of large data sets of galaxies (for a detailed review, see Maiolino & Mannucci 2019, and references within).

Due to observational constraints, early scaling relations were formulated by treating each galaxy as a single observational point (e.g. Schmidt 1959; Faber 1973; Lequeux et al. 1979). However, galaxies themselves are complex systems, and taking into account the powerful interplay between stars, gas, and dust in a galactic disc, the existence of scaling relations on sub-kiloparsec scales (Rosales-Ortega et al. 2012; Sánchez et al. 2013; Gao et al. 2018) becomes an even more powerful diagnostic of galaxy evolution.

A number of large imaging and integral-field spectroscopic surveys have become available in the recent years, and others will soon become operational, including Mapping Nearby Galaxies at Apache Point Observatory (MaNGA; Bundy et al. 2015; Abdurro'uf et al. 2022a), the *Vera C. Rubin* Observatory's Legacy Survey of Space and Time (LSST; Ivezić et al. 2019), and the Euclid Wide Survey (EWS; Laureijs et al. 2011; Euclid Collaboration: Scaramella et al. 2022). These surveys will provide detailed insight into thousands of nearby galaxies. Therefore, developing methods to accurately determine spatially resolved galaxy properties, such as stellar mass and surface density, is of great importance. Recent studies have shown that the structural properties of the stellar components in galaxies are also related to those of the surrounding dark matter halos (e.g. Salucci 2019). This entanglement is thought to be due to tuned baryonic feedback or due to specific properties of the dark particles. This link

between the luminous and the dark components of galaxies could make the availability of these observational properties a primary task in observational cosmology.

Traditionally, the physical properties of galaxies are estimated by fitting template spectral energy distributions (SEDs) to the observations (Walcher et al. 2011; Conroy 2013; Pacifici et al. 2023). SED fitting is usually applied to the global integrated SEDs of galaxies, and the result is an estimate of the main global physical properties, such as stellar mass, SFR, stellar metallicity, star-formation history (SFH), and level of dust attenuation (e.g. Leja et al. 2017, 2019; Nersesian et al. 2019). More recently, SED fitting has also been applied at resolved scales, that is, to individual pixels, yielding distributions of these physical properties at kiloparsec or sub-kiloparsec scales (e.g. Zibetti et al. 2009; Viaene et al. 2014; Sorba & Sawicki 2015; Abdurro'uf et al. 2021, 2022b, 2023).

Several issues arise due to the assumptions that have to be made when deriving the physical properties of galaxies. For example, as the main component of light produced within the galaxy, evolutionary tracks of stars can be used to fit the SEDs of stellar populations (so-called isochrone synthesis, e.g. Charlot & Bruzual 1991; Bruzual & Charlot 1993, 2003). However, complications occur from, among other issues, the incomplete libraries of stellar spectra (e.g. the lack of rest-frame UV libraries), which could be used to obtain accurate galactic parameters for high-redshift galaxies (Pellerin & Finkelstein 2010). Another common issue arises from the limitations of our understanding of the complex history in the interaction between the interstellar medium (ISM) and stars (e.g. Matteucci 2008) as well as modelling of the short and complex evolutionary tracks of the short-lived and difficult to observe high-mass stars.

Another complicating factor is the effect of interstellar dust. The contribution of interstellar dust to galaxy SEDs consists of two effects: attenuation of the intrinsic UV to near-IR (NIR) emission and thermal emission in the mid-IR (MIR) to millimetre regime. In SED fitting, the former is typically accounted for using a parameterised attenuation curve (e.g. Calzetti et al. 2000; Charlot & Fall 2000; Lo Faro et al. 2017; Salim et al. 2018; Decleir et al. 2019), whereas the latter effect is taken into account using empirical or radiative transfer-based template SEDs (e.g. Dale & Helou 2002; Lagache et al. 2004; Siebenmorgen & Krügel 2007; Law et al. 2018). However, our understanding of the complex dust contribution in the ISM is very incomplete, leaving a number of open questions, such as the details of the

* e-mail: inja.kovacic@ugent.be

shape, composition, size distribution, fluffiness, and formation and destruction processes of interstellar dust (e.g. [Draine 2003](#); [Galliano et al. 2021](#)). The attenuation curve and dust emission libraries are limited by the parameter space they were derived from, and linking dust attenuation and dust emission is complex due to non-linear radiative transfer effects ([Baes & Dejonghe 2001](#); [Steinacker et al. 2013](#)). Even with the number of large-scale panchromatic and spectroscopic surveys increasing the understanding of the complex dust grain physics in recent years, there is still a great diversity both in extinction and attenuation curves that contributes to the uncertainty of the properties derived via SED fitting ([Salim & Narayanan 2020](#)).

On top of that, SED fitting is usually computationally demanding, especially when investigating a large parameter space. Scaling up these methods becomes unfeasible if one is to effectively analyse the vast amount of data that will become available with the upcoming surveys. This points to the necessity of exploring alternative methods, and one such method that has been employed in recent years is machine learning (ML). In this work, we focus on supervised ML, which learns the relationship between a set of input parameters and a target via a set of examples (for a review on different ML algorithms and their applications in astronomy, we recommend [Baron 2019](#), and references within).

One of the early issues in astronomy addressed by ML has been measuring the photometric redshift of galaxies due to the difficulty of gaining true values of other galactic parameters from observations ([Tagliaferri et al. 2003](#); [Collister & Lahav 2004](#); [D’Abrusco et al. 2007](#); [Cariles et al. 2010](#); [Hildebrandt et al. 2010](#); [Abdalla et al. 2011](#); [Cavuoti et al. 2012](#); [Brescia et al. 2013](#); [Sadeh et al. 2016](#); [D’Isanto & Polsterer 2018](#); [Leistedt et al. 2023](#); [Alsing et al. 2023, 2024](#)), with some attempts to infer SFRs from photometric data (e.g. [Stensbo-Smidt et al. 2017](#); [Delli Veneri et al. 2019](#)). In recent years, a number of authors have focused on data from cosmological simulations ([Lovell et al. 2019](#); [Gilda et al. 2021](#); [Simet et al. 2021](#)) and observational data ([Acquaviva 2015](#); [Surana et al. 2020](#); [Iglesias-Navarro et al. 2024](#); [Alsing et al. 2024](#)), demonstrating that ML techniques can accurately recover a number of galactic parameters even on limited data sets and with much smaller computational and time requirements. One interesting application of ML is to infer hard to measure observations (for example far-IR [Luhman et al. 2003](#)), thus bridging the gap in the galactic spectral libraries on both global ([Dobbels et al. 2020](#)) and resolved ([Dobbels & Baes 2021](#)) scales. Interestingly, in [Dobbels & Baes \(2021\)](#) the authors find no significant dependence of the prediction by the ML algorithm on the pixel scale of the observations in their work.

These recent successful developments in deploying ML to solve astronomical problems have a potential to be very significant for the *Euclid* mission ([Euclid Collaboration: Mellier et al. 2024](#)). Successfully launched on 1 July 2023, it has started to generate a vast amount of imaging data in four broadband filters. In particular, the Euclid Wide Survey ([Laureijs et al. 2011](#)) is expected to detect $\sim 2 \times 10^9$ galaxies up to $z \sim 3$, with the majority of galaxies at $z \sim 1$ ([Euclid Collaboration: Scaramella et al. 2022](#)). Mapping the distribution of the physical properties of these galaxies as a function of cosmic time, and the correlation between them, offers an additional constraint to the scaling relations for theories of galaxy formation and evolution. At the same time, the *Euclid* surveys will spatially resolve tens of thousands of nearby galaxies to kiloparsec or sub-kiloparsec scales. Generating physically resolved maps at kiloparsec scales of the most important physical properties of galaxies offers valuable input to galaxy evolution theories. The sheer amount of data makes ML

techniques a necessary tool to efficiently generate physical maps at high spatial resolution.

There have been a number of attempts to test the feasibility of ML methods on simulated *Euclid* data, for example, modelling photometric redshift ([Euclid Collaboration: Desprez et al. 2020](#)), galaxy classification ([Euclid Collaboration: Humphrey et al. 2023](#)), identification of galaxy-galaxy strong lensing events ([Euclid Collaboration: Leuzzi et al. 2024](#)), deriving galaxy properties ([Euclid Collaboration: Bretonnière et al. 2022, 2023](#); [Euclid Collaboration: Bisigello et al. 2023](#); [Euclid Collaboration: Aussel et al. 2024](#)), and inferring global physical parameters of galaxies ([Euclid Collaboration: Enia et al. 2024](#)). The studies demonstrate that ML could be a viable alternative to traditional methods; however, caution should be taken into account due to the high impact of the parameter space on the algorithm’s predictive power and the difficulty of interpreting the results, both of which are common pitfalls with ML.

In this work, we expand on the approach explored by [Euclid Collaboration: Bisigello et al. \(2023\)](#), who attempted to infer galaxy physical properties with deep ML from mock *Euclid* magnitudes and *H*-band images. We focus on the spatially resolved distribution of physical properties within nearby galaxies. More specifically, we investigate to which degree we can infer physical property maps from sets of synthetic *Euclid* images of galaxies extracted from the TNG50 cosmological hydrodynamical simulation ([Pillepich et al. 2019](#)). We employ two ML methods, random forest (RF) and a deep neural network (DNN), and we compare them to an ordinary least squares linear regression (LR). We focus on stellar mass surface density (Σ_*), stellar-mass-averaged stellar metallicity (Z_*), and stellar-mass-averaged stellar age (t_*). The workflow of the entire ML process described in this work can be seen in Fig. 1. In this work, we focus on noise-free data; however, we plan to include noise and test this relationship on more realistic simulated data as well as on actual *Euclid* observations.

This paper is organised as follows: In Sect. 2, we introduce the synthetic images and describe the steps undertaken to prepare them for the ML method. In Sect. 3, we describe the ML algorithms considered in this work and the metrics used to analyse the viability of this method. Section 4 contains the estimates of galaxy properties, while Sect. 5 focuses on the underlying processes that drive these results, including a focused discussion on the effects of the underlying resolved mass-metallicity relation. The main findings and the future outlook are summarised in Sect. 6.

2. Data

2.1. The TNG50 simulation

IllustrisTNG ([Springel et al. 2017](#); [Pillepich et al. 2017](#); [Pillepich et al. 2018](#); [Naiman et al. 2018](#); [Marinacci et al. 2018](#), hereafter TNG) is a publicly available state-of-the-art large-volume magnetohydrodynamics cosmological simulation that uses the moving mesh code AREPO ([Springel 2010](#)) to model the formation and evolution of galaxies in a Λ CDM universe. As a successor of the Illustris project ([Vogelsberger et al. 2014](#); [Genel et al. 2014](#)), it offers an upgrade and refinement in its treatment of cosmic magnetism, supermassive black holes, galactic winds, and stellar evolution ([Weinberger et al. 2016](#)). The TNG project consists of three runs: TNG50, TNG100, and TNG300. They are defined by the size of their physical simulation box with side lengths of 50, 100, and 300 Mpc, respectively ([Nelson et al. 2019b](#)). Out of the three, the highest-resolution TNG50 ([Pillepich et al. 2019](#);

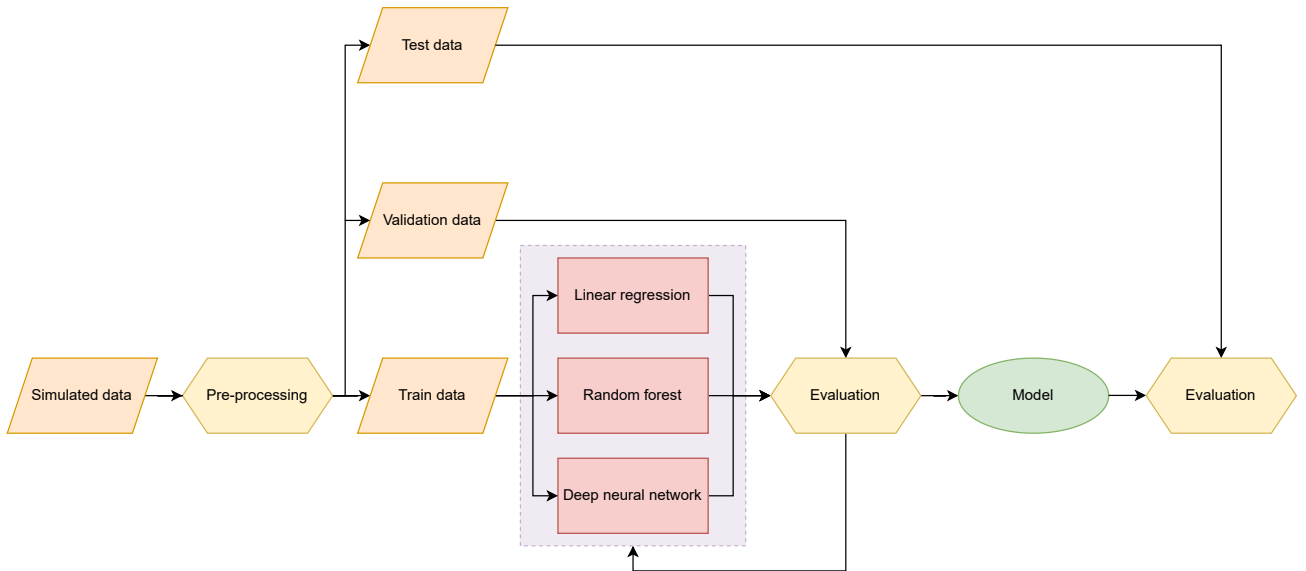


Fig. 1. Workflow of the entire ML process described in this work.

Nelson et al. 2019a) is best suited for the purpose of our work. It follows the evolution of a cubic volume of 51.7 comoving Mpc on the side, with cosmological parameters based on the *Planck* mission (Planck Collaboration: Ade et al. 2016): $\Omega_m = 0.3089$, $\Omega_b = 0.0486$, $\Omega_\Lambda = 0.6911$, $H_0 = 100 h \text{ km s}^{-1} \text{ Mpc}^{-1}$ with $h = 0.6774$, and a Chabrier (2003) initial mass function (IMF). It reaches a baryonic mass resolution of $8.5 \times 10^4 M_\odot$ and the average cell size of 70–140 pc in the star-forming regions of the galaxies. For more details on this simulation, we refer to Pillepich et al. (2019).

2.2. Synthetic images

Our data set consists of 1154 galaxies at $z = 0$ with a total mass range between $10^{9.8} M_\odot$, and $10^{12} M_\odot$ extracted from the TNG50 simulation. Based on this set of galaxies, Baes et al. (2024) recently released an image atlas, the TNG50-SKIRT Atlas, which contains synthetic high-resolution images in 18 broadband filters from UV to NIR wavelengths. The images are generated with the SKIRT radiative transfer code (Baes et al. 2011; Camps et al. 2015; Camps & Baes 2020) and take into account different stellar populations, absorption and scattering by ISM dust, and localised dust attenuation in star-forming regions. The atlas contains images for each galaxy at five different projections, where the angle of the first projection is arbitrarily determined, and the other four are spread over a unit sphere in a way to maximise the angular separation between them (Tammes 1930). Beside the dust-obscured images, the atlas also contains dust-free images in all 18 bands, as well as matching maps of the stellar mass surface density, dust mass surface density, and mass-weighted stellar age and metallicity. These physical parameter maps were generated by projecting the 3D physical fields such as the stellar mass density on the observer’s plane of the sky, using the ‘probe’ functionality in the SKIRT code (Camps & Baes 2020). The sample contains 869 star-forming galaxies and 285 quiescent galaxies (for a characterisation of the sample, see Fig. 1 in Baes et al. 2024).

We extend the TNG50-SKIRT Atlas image database presented in Baes et al. (2024) with additional synthetic images in the four *Euclid* broadband filters, that is, the optical I_E band of

the VIS instrument (Cropper et al. 2018; Euclid Collaboration: Mellier et al. 2024) and the Y_E , J_E , and H_E bands of the NIR NISP instrument (Euclid Collaboration: Schirmer et al. 2022; Euclid Collaboration: Jahnke et al. 2024). The images are generated in a similar way as for the original atlas: we use SKIRT to generate both dust-obscured and dust-free images, but now with the *Euclid* transmission curves. Apart from that, all other settings, such as the observer positions, the spatial resolution, and the field of view, remain identical as for the original images (see Baes et al. 2024, Table 2). The additional SKIRT simulations were run on the supercomputer facilities of the Vlaams Supercomputer Centre (VSC)¹.

All maps and images have a pixel scale of 100 pc and a field of view of 160 kpc on the side, corresponding to 1600×1600 pixels in size, thus ensuring the outer regions of even the most extended galaxies are covered. Except for the presence of Monte Carlo noise in dust-obscured images, the images are noise-free.

For this work, we opt only for galaxies belonging to one randomly chosen projection (O5), since the first projection angle is arbitrary, thus ensuring a similar randomness in the inclination of the observations. Of course, we can assume that the projection angle affects the accuracy of the result, but our choice is motivated by the real world scenario of observing a large number of galaxies. Additionally, we decide to infer the galaxy parameters on a pixel-by-pixel basis which, due to the high resolution of the images, gives us a high number of data points. Due to this wealth of data, the inclusion of other angles should not offer any additional information for the ML method explored here, but it might prove its usefulness in, for example, training a convolutional neural network since in that case each data point would have to be a tuple consisting of a number of neighbouring pixels, thus decreasing the number of available data points from this data set significantly; however that is beyond the scope of this work. Regardless, we test whether the inclusion of other projections would affect the result on a subset of 200 randomly selected galaxies using a DNN (for the description, see Sect. 3.1). At five angles each, the total data set consists of 1000 “unique” observa-

¹ <https://www.ugent.be/hpc/en/infrastructure/overview.htm>

tions. The results are shown in Appendix A. For all three target parameters, the differences in accuracy of inferred galaxy properties are negligible.

2.3. Data pre-processing

The SKIRT generated mock images contain Monte Carlo noise, which increases towards fainter surface brightness levels. To assess the reliability of the surface brightness levels, we select a sample of ten galaxies for which we re-ran the SKIRT simulations with the reliability statistics module switched on. The result of these additional simulations are diagnostic statistics on the reliability of the Monte Carlo simulation in each pixel (for detailed information on this procedure, we refer to Camps & Baes 2018, 2020). Based on these statistics, we created a mask in the H_E -band images that removes all pixels where the Monte Carlo results are deemed unreliable. The motivation behind the choice of the H_E band was twofold. Traditionally, stellar mass is reconstructed as the stellar mass-to-light ratio (that is, the ratio between stellar mass and light that reaches the observer) from the optical-near-IR colour (Zibetti et al. 2009; Zibetti et al. 2020). The variations in this ratio are smaller at longer wavelengths due to NIR being dominated by the light from the low-mass long-lived stars, as well as the lower dust extinction at those wavelengths (Bell & de Jong 2001), making the H_E band at $1.65\,\mu\text{m}$ close to the ideal choice for reconstructing stellar mass (Walcher et al. 2011). The second reason for this choice, as it will be discussed in detail in Sect. 5.3, comes from the fact that it is the most important feature when modelling both Σ_\star and Z_\star . The application of this mask translates to the removal of all pixels where $I_{H_E} \leq 0.55\,\text{MJy sr}^{-1}$, which removes a large fraction of the pixels in the outskirts of the images. The full data set of pixels surviving this clipping still consists of more than 71 million individual pixels.

We convert both the surface brightness levels and the galactic parameters onto logarithmic scale, thus avoiding issues that could occur due to the large dynamic range. Surface brightness levels are in units of MJy sr^{-1} . The Σ_\star is in units of $\text{M}_\odot \text{pc}^{-2}$ and spans seven orders of magnitude in total (from -2.5 dex to 5.3 dex); however, 99% of all points span three (from 0 dex to 3 dex). The Z_\star is dimensionless but normalised with $Z_\odot = 0.02$ in the analysis, and spans less than two orders of magnitude (from -2.8 dex to -1.1 dex), with 99% of all points spanning only from -2.2 dex to -1.2 dex. The t_\star , in units of Gyr, spans two orders of magnitude in total (from -3.3 dex to 1.1 dex), and less than one for the inner 99% of all pixels within the test set (from 0.2 dex to 1.1 dex). Afterwards, we normalise the data in such a way that all the values lie between zero and one (a standard procedure when preparing a data set for ML problems) using the linear transformation

$$X_{\text{std}} = \frac{X - X_{\min}}{X_{\max} - X_{\min}}, \quad (1)$$

where X_{\min} and X_{\max} are the minimal and maximal value of the training set range, respectively.

Finally, we opt to split our data on a galaxy instead of a pixel basis. This translates to a random fraction of $\sim 10\%$ of galaxies, equalling to a total of 115 entire galaxies (of which 86 are star-forming and 29 are quiescent) being selected into the test set, which is never seen by any of the algorithms during training. The remaining 90% (totalling 1039 galaxies, of which 783 are star-forming and 256 are quiescent) is used by the algorithm for training. However, 20% of pixels are randomly set aside for the validation of the algorithm during training. This split is random

Table 1. Hyper-parameter setup for the RF algorithm.

Hyper-parameter	Setup
Number of trees in the forest	100
Minimum number of samples required in a leaf	12
Minimum number of samples to split a node	20
Maximum depth of the tree	80
Maximum number of samples at base estimator	2 000 000
Loss function	MSE

and changes at each run of the algorithm. The total data distribution can be seen in Fig. 2.

3. Methods

3.1. Machine learning approach

We tested two supervised ML algorithms, namely an RF and a DNN, to map *Euclid* mock observations of nearby galaxies to galactic physical parameters on a pixel-by-pixel basis, and to analyse the relationship between them. We compare these two algorithms against a baseline, namely an ordinary least squares linear regression, which is modelled from the Python SciPy library as implemented in Python scikit-learn (Pedregosa et al. 2011) and concerns the linear component of this relationship.

The two ML algorithms concern the non-linear component of this relationship. For the first one, we selected RF, an ensemble method algorithm (Breiman 2001) from the Python scikit-learn library. The RF algorithm builds a forest of decision trees, where each decision tree starts at the root node and then recursively splits until the minimum node size is reached. At each node, it selects a random given subset of variables, determines the best variable at this node, and splits it into two daughter nodes, then it averages the output of an ensemble of such trees to make a prediction (Hastie et al. 2009; James et al. 2023). RF is one of the most robust methods for ML, as it was designed to avoid over-fitting (the algorithm becoming too good at predicting values on the set it was trained on, but bad at generalising on new data), and it requires minimal user input when tuning the hyper-parameters (for instance, the size of the decision tree, number of samples contained within a leaf, etc). Prior to this choice, we tested a number of different algorithms from the scikit-learn Python library on a subset of our data set, and did not notice any major effect on the accuracy of our predictions. The additional benefit of RF when compared to other ML algorithms that show similar speed and accuracy is its ability to determine feature importance, which offers a certain degree of interpretability and helps in the analysis of the results.

When building the RF, we started with a brute force approach at tuning the hyper-parameters using the scikit-learn module GridSearchCV on a subset of 20% randomly selected pixels of the training set, followed by manual tuning on the entire data set until we converge to a solution that minimises computational requirements without sacrifices to the accuracy of the prediction (using metrics described in Sect. 3.3). As a final step for this method, we perform a k -fold cross-validation across the entire training set, which consists of 1039 whole galaxies. In the cross-validation procedure, we bin the data into k (in this case $k = 10$) bins that consist of random $\sim 10\%$ (~ 104) galaxies. We set aside one bin as a validation set, train the algorithm on the remaining nine bins, and repeat the aforementioned process in a loop through all ten bins. This is a standard procedure for hyper-

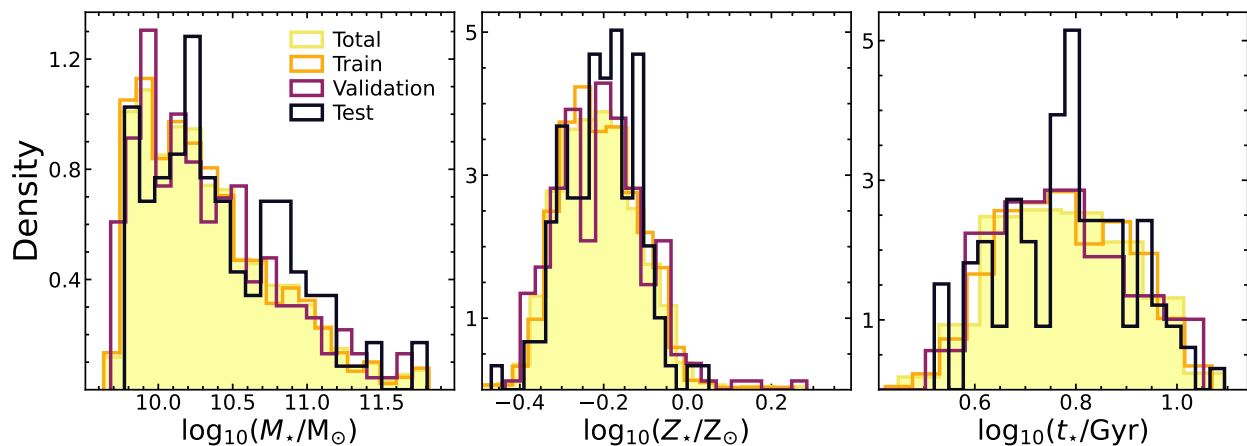


Fig. 2. Normalised distribution of $\log_{10} M_*$, $\log_{10} Z_*$, and $\log_{10} t_*$ global values of the galaxy properties for the entire data set, training set, and test set (see legend).

parameter tuning for RF, but it is also a useful additional step to avoid any possibility of over-fitting that might occur with the sample selection itself, as there is always a small risk of the algorithm behaving well on one validation set, but not on the others.

The RF setup used in this work is described in Table 1. We limit the maximum size of a sub-sample required to build one tree to two million, as this quantity is the most responsible for high memory requirements of RF. We decrease this value from the size of the full set until we reach the minimum at which the accuracy of the algorithm is not affected. We test the RF on a desktop machine to determine feasibility of modelling galactic parameters in optimal amount of time without requiring dedicated infrastructure. With the final model, we are able to train the algorithm on the entire data set in less than an hour on an 11th generation Intel i7 2.30 GHz 16-core laptop with 62.5 GiB memory and 128 GiB swap.

The third and final approach consists of building a DNN using the Google Python library TensorFlow (Abadi et al. 2015). An artificial neural network is a type of a ML technique that is inspired by the functioning of neurons in a biological brain. It consists of a series of layers of a certain width, that is the number of neurons within a layer. The architecture of a DNN used in this work is summarised in Table 2 and consists of six layers. The first layer is the input layer, its size is determined by the number of features loaded into the algorithm. This is followed by four hidden layers that consist of 64 neurons each, which are fully connected with all the neurons in the previous and following layers. Each neuron receives a signal in the form of all the features (x_i) whose number is determined by the first layer, multiplies them by a certain weight (w_i), which is different for each feature, and is arbitrarily determined at the initialisation of the network. The weighted features are then summed up together with a bias (b). A final, single output layer returns the predictions of a single of the three galaxy parameters (\hat{y}). Each of these transformations would be linear; however, a non-linear so-called activation function is then applied on this output, returning a non-linear transformation:

$$\hat{y} = f\left(\sum_{i=1}^n (w_i \times x_i) + b\right) \quad (2)$$

The most common activation function used for regression problems (e.g. Dahl et al. 2013; Zeiler et al. 2013; Krizhevsky et al. 2017), as well as in this work, is a rectified linear unit

activation function (ReLU; Nair & Hinton 2010; Glorot et al. 2011).

The network takes a sub-sample of 8192 points from our data set. This batch sub-sample is then propagated through the network, obtaining predicted values of the target parameter, which is then assessed for accuracy against a true value in this batch via a loss function. For the loss function we choose the mean squared error (MSE). MSE calculates the average of a squared distance between the predicted and the real value, it is the most common loss function used for regression problems, and it has proven to work the best in our case as well. The result of this comparison between the predicted and true values in the batch determine the direction in which the weights are updated, and this information is then back-propagated through the network. This process is repeated until the entire data set is used.

Another important hyper-parameter in building a DNN is the learning rate, which determines the size of the step the algorithm takes in the optimisation process (Murphy 2012). Instead of choosing a constant learning rate, we decide to implement a learning rate scheduler, whose main function is decreasing the network training time. The network starts with a learning rate of 10^{-4} , a lower end of a fairly common starting point. This learning rate is then multiplied by 0.8 every time the loss function does not improve for more than 300 iterations. A learning rate that is too low requires a large number of iterations before the network converges, and it also increases the risk of the network becoming stuck in a local minimum, and never reaching a global one. On the contrary, a learning rate that is too high can cause the network to overshoot, that is arbitrarily jump between points without ever “settling” at a minimum. A learning rate scheduler is a simple way to circumvent that common downfall while still benefiting from increased performance earlier in the learning. And finally, we also set an early stop, signalling to the algorithm to stop its learning process if the loss function hasn’t decreased for 2000 iterations. Even though it is rather large, this value allows the learning rate scheduler at least five updates, allowing it to decrease to around one third of its initial value. From our tests, this is a large enough number of epochs to ensure the best result, but it also still decreases the training time significantly.

For the optimiser, that is the algorithm which dynamically adjusts the learning rate for each parameter, we choose Adam (adaptive moment estimation; Kingma & Ba 2015), which has shown superior performance for large data sets (Duchi et al.

Table 2. Hyper-parameter setup for the DNN.

Hyper-parameter	Setup
Number of hidden layers	4
Neurons per layer	64
Training epoch	10 000
Activation function	ReLU
Batch size	8192
Optimiser	Adam
Initial learning rate	10^{-4}
Learning rate scheduler patience	300
Learning rate scheduler factor	0.8
Loss function	MSE
Early stop	2 000

2011). For the details on how DNNs function, we recommend Goodfellow et al. (2016).

Unlike RF, a DNN requires a high level of engagement when tuning its hyper-parameters and its training is much more computationally demanding. In this work, it requires a dedicated GPU architecture to be viable, and was run on the supercomputer facilities of the VSC². In case of all three galaxy parameters, specifically the ones that show strong non-linearity, a DNN might offer a better performance over RF, but at the expense of computational power, though as it will be shown in this work, the difference between the predictive power of an RF and a DNN is negligible. Regardless, we present the results of both non-linear methods.

3.2. Input data

The input features from the mock observations consist of four *Euclid* bands. To test our three ML methods, we organised them into the following three different categories:

- Each of the single bands, and we selected the best fit (according to metrics presented in Sect. 3.3), which is H_E in all cases.
- All four bands: I_E, Y_E, J_E, H_E .
- Four bands, I_E, Y_E, J_E, H_E , as well as six colours modelled from them: $(I_E - Y_E), (I_E - J_E), (I_E - H_E), (Y_E - J_E), (Y_E - H_E)$, and $(J_E - H_E)$.

3.3. Metrics of model reliability

Throughout this work, we use a number of metrics to determine the reliability of ML. The coefficient of determination (R^2 ; Wright 1921) is a commonly used term in regression analysis that can span from $-\infty$ to $+1$, where the latter denotes a model that perfectly matches the ground truth. R^2 determines how much the dependent variable, which in the case of ML is the prediction of the model, is determined by the independent variable, or the ground truth. R^2 is considered one of the most informative metrics in regression ML (Chicco et al. 2021). It is calculated as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{\sum_{i=1}^n (\bar{y} - y_i)^2}, \quad (3)$$

² Specifically, the *accelgor* GPU cluster with the NVIDIA Ampere A100 GPU and no parallelisation

where y_i is the actual i -th value, \hat{y}_i is the predicted i -th value, and \bar{y} is the mean of true values, which is defined as

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i. \quad (4)$$

The root-mean-square error (RMSE) of the estimation determines how much the predicted value deviates from the ground truth,

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2}, \quad (5)$$

and it can span from zero to $+\infty$, with the former denoting the case where the prediction perfectly matches the ground truth. The RMSE is monotonically anti-correlated to R^2 .

The normalised median absolute deviation (NMAD) represents a mean absolute deviation that is normalised by a factor of ~ 1.4826 , and it is considered the equivalent of standard deviation that is more resilient to outliers (Höhle & Höhle 2009),

$$\text{NMAD} = 1.4826 \times \text{median} |\hat{y}_i - y_i|. \quad (6)$$

The bias is represented as

$$\langle \Delta Y \rangle = \text{median} (\hat{y}_i - y_i), \quad (7)$$

where Y denotes any of the target galaxy properties (that is, Σ_* , Z_* , or t_*).

We calculated the fraction of catastrophic outliers (f_{out}), that is, the number of predicted values that are approximately two times (0.3 dex) above or below the ground truth, matching the uncertainties at recovering stellar mass values via traditional methods (Conroy 2013),

$$f_{\text{out}} = \frac{n(|\hat{y}_i - y_i| > 0.3 \text{ dex})}{n_{\text{tot}}}. \quad (8)$$

And finally, we calculated the Pearson correlation coefficient (ρ ; Pearson 1895):

$$\rho = \frac{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sigma_y \sigma_{\hat{y}}}, \quad (9)$$

where \bar{y} is the mean of true values (y_i); $\bar{\hat{y}}$ is the mean of predicted values (\hat{y}_i), as defined in Eq. (4); and σ_y and $\sigma_{\hat{y}}$ are the standard deviations of true and predicted values, respectively.

Additionally, as is visible in Fig. 2, there is a slight bias, especially for Z_* and t_* , in the distribution of the test set. To confirm that this bias does not affect the final result, we calculate the metrics of model reliability on both the test and the training sets. The difference between them is minor (within 10%). The values are shown in Appendix B.

3.4. Reliability of a machine learning algorithm

To assess the reliability of a ML algorithm, there are two types of uncertainties: aleatoric and epistemic. Aleatoric (statistical) uncertainty is caused by the presence of random noise in the data set, that is, the irreducible noise from observations, while epistemic (systemic) is caused by the lack of knowledge of the best model (Hüllermeier & Waegeman 2021). To assess the former, we would be required to perform the same ML procedure on idealised data, as well as the data with added noise. Since the goal of

this work is to examine the amount of physical information that can be extracted from the *Euclid* observations, and the feasibility of using ML in that process, we are currently only working with idealised data. Introducing noise into the data set would complicate the relationships between the observations and physical parameters with additional variables, so this type of uncertainty cannot be assessed. In the future, we plan to adapt this data set to the actual observations, making this type of error assessment possible. Epistemic uncertainty, on the other hand, informs us of how uncertain the model is about its prediction (Kendall & Gal 2017), and is assessed by running the same algorithm on the same data set a certain number of times. In this work, we opt for ten.

As a deterministic model (that is, an algorithm that creates the same prediction on the same set of data every time), the ordinary least squares linear regression has an epistemic error of zero. RF and DNN, on the other hand, build a model that is conditional on the previous decisions made on the data set, giving them a certain level of randomness that will change the model on every new run during training.

In case of our two ML algorithms, the situation is more complicated due to the algorithms requiring a validation set during training, and the way this validation set is created. While the test data is set to the bin consisting of $\sim 10\%$ (i.e. 115) whole galaxies, the training data is randomly split into a training and validation set used by the algorithm in a 80%/20% ratio.

Additionally, each of the two ML algorithms contribute to the uncertainty due to the stochasticity of the learning process itself. In case of RF, the uncertainty is related to the fact that, for each time the tree splits, a random set of features is selected for training, leading to a different path from the root node to the final leaf. However, in all cases this uncertainty is very small, in the worst case scenario being when the standard deviation ($\pm 1\sigma$) is only $\sim 2\%$ of the mean of R^2 . In all other cases, it is 20–100 times smaller.

Every time a DNN is initialised, it starts with a different randomly assigned set of weights with which it updates the training data, that is the algorithm starts at a different, random starting point in the optimisation process, and for every epoch, it shuffles the set of samples anew, which affects the update of the weights, and maps a different path through the optimisation. However, these values are still fairly small, with a first standard deviation accounting for 5% of the mean of R^2 , a slightly higher value when compared to RF, but still low, and over 20 times higher than the second largest error.

4. Results

In this section, we describe the results for each of our three target values: stellar mass surface density (Σ_\star), stellar-mass-averaged stellar metallicity (Z_\star), and stellar-mass-averaged stellar age (t_\star). These values were inferred using three methods: ordinary least squares LR and two ML algorithms, namely, RF and DNN. Each algorithm was trained independently for three different selections of input values: single H_e band, all four *Euclid* bands, and when using four *Euclid* bands and six colours modelled from them. Altogether, this totals to 27 separate cases, as shown in Table 3.

4.1. Stellar mass

Most of the information on stellar mass can be gained from the NIR $1.65\,\mu\text{m}$ H_e band (Zibetti et al. 2009; Walcher et al. 2011).

In our case, when modelling Σ_\star using just one *Euclid* band, the strongest relationship shown is between Σ_\star and the H_e -band surface brightness ($R^2 \approx 0.85$), regardless of the algorithm, pointing to a very strong linear relationship between H_e and Σ_\star . When we model Σ_\star using all four *Euclid* bands, the increase in accuracy is lower for LR ($R^2 \approx 0.873$), and higher for RF and DNN ($R^2 \approx 0.927$). This indicates that there is a certain non-linear component to the relationship between the other three bands and Σ_\star that is uncovered similarly well by both non-linear algorithms, with the RMSE for DNN with four *Euclid* bands being 0.130 dex. The addition of six colours modelled from the four *Euclid* bands does seem to offer a minor benefit ($R^2 \approx 0.908$) for LR, but not for RF or DNN.

We show the relationship between the true Σ_\star and the predicted Σ_\star in the case of using only the four *Euclid* bands for all three methods in Fig. 3. The diagonal line marks the identity, with two additional dashed lines on each side that mark f_{out} (Eq. 8). For higher values of Σ_\star , the scatter is fairly low, especially in case of RF and DNN. However, in all three methods, the scatter increases as Σ_\star decreases, especially at $\Sigma_\star \leq 30\,\text{M}_\odot\,\text{pc}^{-2}$. This is expected, as all our observations were clipped for values $I_{H_e} \leq 0.55\,\text{MJy sr}^{-1}$. With the strong linear relationship between Σ_\star and H_e , the effect of this clipping is fairly prominent as a lower boundary for inferred $\Sigma_\star \approx 4\,\text{M}_\odot\,\text{pc}^{-2}$ for LR. Simultaneously, we can expect a stronger effect of Monte Carlo noise at lower surface brightnesses, increasing the uncertainty of ML predictions. For RF and DNN the $f_{\text{out}} \approx 3\%$, but over double that value ($f_{\text{out}} \approx 7\%$) for LR (Table 3).

In conclusion, Σ_\star can be inferred with a high accuracy from *Euclid* observations using ML (with the Pearson ρ coefficient spanning from 0.923–0.963 in all cases). While the most information is extracted from the *Euclid* H_e band that scales with Σ_\star in a linear fashion, the other three bands offer additional non-linear information that justifies the usage of non-linear ML algorithms such as RF or DNN.

4.2. Stellar metallicity

Stellar metallicity can be extracted from the spectral absorption features (line-strength indices, e.g. Maiolino & Mannucci 2019) in the optical range of the spectrum. However many of these optical absorption lines are narrow and shallow, making them difficult to measure (Ditrani et al. 2023). For *Euclid* only the I_e band covers the optical range of the spectrum, but due to its broadness, it would make inferring Z_\star with *Euclid* imaging very difficult.

The process of predicting Z_\star in this work follows similar steps to Σ_\star . Out of the single band fits, the most important feature for predicting metallicity is the H_e band ($R^2 \approx 0.36$ for LR and RF, and $R^2 \approx 0.39$ for DNN), pointing to a mild but mostly linear relationship between Z_\star and H_e . Similarly to Σ_\star , the accuracy increases when all four *Euclid* bands are used, and the gain is much higher for RF and DNN ($R^2 \approx 0.595$) than LR ($R^2 \approx 0.417$), pointing to the relationship between Z_\star and the other three *Euclid* bands to fall in a more non-linear regime, similarly to Σ_\star . The inclusion of colours offers a small benefit to LR ($R^2 \approx 0.435$), but extremely minor to RF, and none to DNN. We recover Z_\star with a RMSE < 0.1 dex in all cases, with the best case scenario being with a DNN (RMSE = 0.076 dex), in case when four *Euclid* bands are used.

We plot the relationship between the true and predicted Z_\star in Fig. 3. The relationship is somewhat similar to Σ_\star , where highest values show the least amount of scatter, which again increases steadily until $Z_\star \approx 0.5\,Z_\odot$. Compared to the prediction of Σ_\star , the

Table 3. Accuracy scores (as defined in Sect. 3.3) for the three galaxy properties (Σ_* , Z_* , and t_*) for the three methods (ordinary least squares LR, and two ML algorithms, RF and DNN) for the three distinct cases (single H_E , four *Euclid* bands, and four *Euclid* bands as well as the six colours modelled from them).

Target	Algorithm	Input features	R^2	RMSE	f_{out}	NMAD	$\langle\Delta Y\rangle$	ρ
$\Sigma_*/(\text{M}_\odot \text{pc}^{-2})$	LR	I_{H_E}	0.852	0.186	0.0757	0.1676	-0.1676	0.923
		Bands	0.873	0.147	0.0725	0.1455	-0.0055	0.935
		Bands + colours	0.908	0.150	0.0645	0.1410	-0.0059	0.939
	RF	I_{H_E}	0.853	0.185	0.0758	0.1685	-0.0185	0.923
		Bands	0.927	0.130	0.0325	0.1102	-0.0021	0.963
		Bands + colours	0.927	0.130	0.0324	0.1102	-0.0022	0.963
	DNN	I_{H_E}	0.854	0.185	0.0736	0.1673	-0.0193	0.924
		Bands	0.927	0.130	0.0322	0.1090	-8.45×10^{-5}	0.963
		Bands + colours	0.927	0.130	0.0322	0.1089	-8.03×10^{-5}	0.963
Z_*	LR	I_{H_E}	0.358	0.096	0.0047	0.0937	-0.0142	0.598
		Bands	0.417	0.091	0.0030	0.0863	-0.0063	0.655
		Bands + colours	0.435	0.095	0.0024	0.0857	-0.0075	0.666
	RF	I_{H_E}	0.356	0.096	0.0050	0.0947	-0.0141	0.590
		Bands	0.592	0.077	0.0007	0.0725	0.0016	0.769
		Bands + colours	0.595	0.077	0.0007	0.0724	0.0020	0.770
	DNN	I_{H_E}	0.385	0.096	0.0048	0.0938	-0.0145	0.599
		Bands	0.595	0.076	0.0007	0.0720	0.0029	0.773
		Bands + colours	0.595	0.076	0.0007	0.0719	0.0028	0.773
t_*/Gyr	LR	I_{H_E}	0.010	0.163	0.0509	0.1586	-0.0435	0.104
		Bands	0.199	0.146	0.0368	0.1335	-0.0206	0.467
		Bands + colours	0.162	0.150	0.0354	0.1328	-0.0216	0.476
	RF	I_{H_E}	0.014	0.162	0.0509	0.1576	-0.0424	0.096
		Bands	0.382	0.128	0.0285	0.1008	-0.0061	0.618
		Bands + colours	0.381	0.129	0.0286	0.1007	-0.0051	0.618
	DNN	I_{H_E}	0.130	0.162	0.0498	0.1584	-0.0430	0.130
		Bands	0.382	0.128	0.0281	0.0997	-0.0024	0.622
		Bands + colours	0.382	0.128	0.0280	0.0994	-0.0025	0.623

Notes. Left to right: Target; ML algorithm; input features; coefficient of determination (R^2); root-mean-square error (RMSE); number of catastrophic outliers (f_{out}); normalised median absolute deviation (NMAD); and bias ($\langle\Delta Y\rangle$, where Y represents the target galaxy property); Pearson ρ coefficient. All values are defined in Sect. 3.3.

inferred value of Z_* shows a much larger scatter. RF and DNN perform significantly better than LR, which seems to struggle with inferring values for $Z_* < 0.4 Z_\odot$. Even for higher ranges (where $Z_* \geq 1 Z_\odot$), RF and DNN capture the distribution of Z_* much more accurately and with a lot less scatter. When only H_E is used, $f_{\text{out}} \approx 0.5\%$ (Table 3), a value similar to all three methods. This value halves with the inclusion of additional data for LR ($f_{\text{out}} \approx 0.25\%$), and drops seven-fold for RF and DNN ($f_{\text{out}} \approx 0.07\%$). It is important to note that the range for Z_* is much smaller than Σ_* , making these results seem better in comparison, when in fact Σ_* is inferred better in all cases. For details on the different ranges of our target values, we refer to Sect. 2.3.

We conclude that Z_* can be inferred from *Euclid* observations with a fairly high accuracy, but the prediction is less robust than in case of Σ_* (Pearson correlation coefficient $\rho = 0.77$ for Z_* , compared to $\rho = 0.96$ for Σ_*). Similarly to Σ_* , the information contained in the *Euclid* H_E band is related to Z_* in a linear fashion, while the other three bands contain the information in a more non-linear regime.

4.3. Stellar age

The procedure for modelling stellar ages is tightly entangled with Z_* and requires the information located in the 4000 Å Balmer break (Poggianti & Barbaro 1997) which is outside of the *Euclid* observational window for local Universe galaxies. As was the case with Z_* , *Euclid* observations do not carry enough information to directly extract stellar ages for our data set either.

When using a single *Euclid* band to model t_* , the predictions are equivalent to random chance ($R^2 \approx 0$), as shown in Table 3. The other two attempts yield similar results, with RF and DNN both offering a slightly higher accuracy than LR; however, these values are still low ($R^2 < 0.4$), giving credence to the conclusion that t_* cannot be accurately modelled from *Euclid* observational data alone. As with Z_* , RF and DNN can mimic the distribution of t_* values somewhat, but the accuracy is too low for them to be a feasible method of modelling t_* with *Euclid* observations.

While the scatter is lower for the higher range of ages, it shows an offset where regardless of the algorithm, t_* is systematically under-predicted, worsening with increasing age (Fig. 3). On the lower side of the range, the scatter quickly increases,

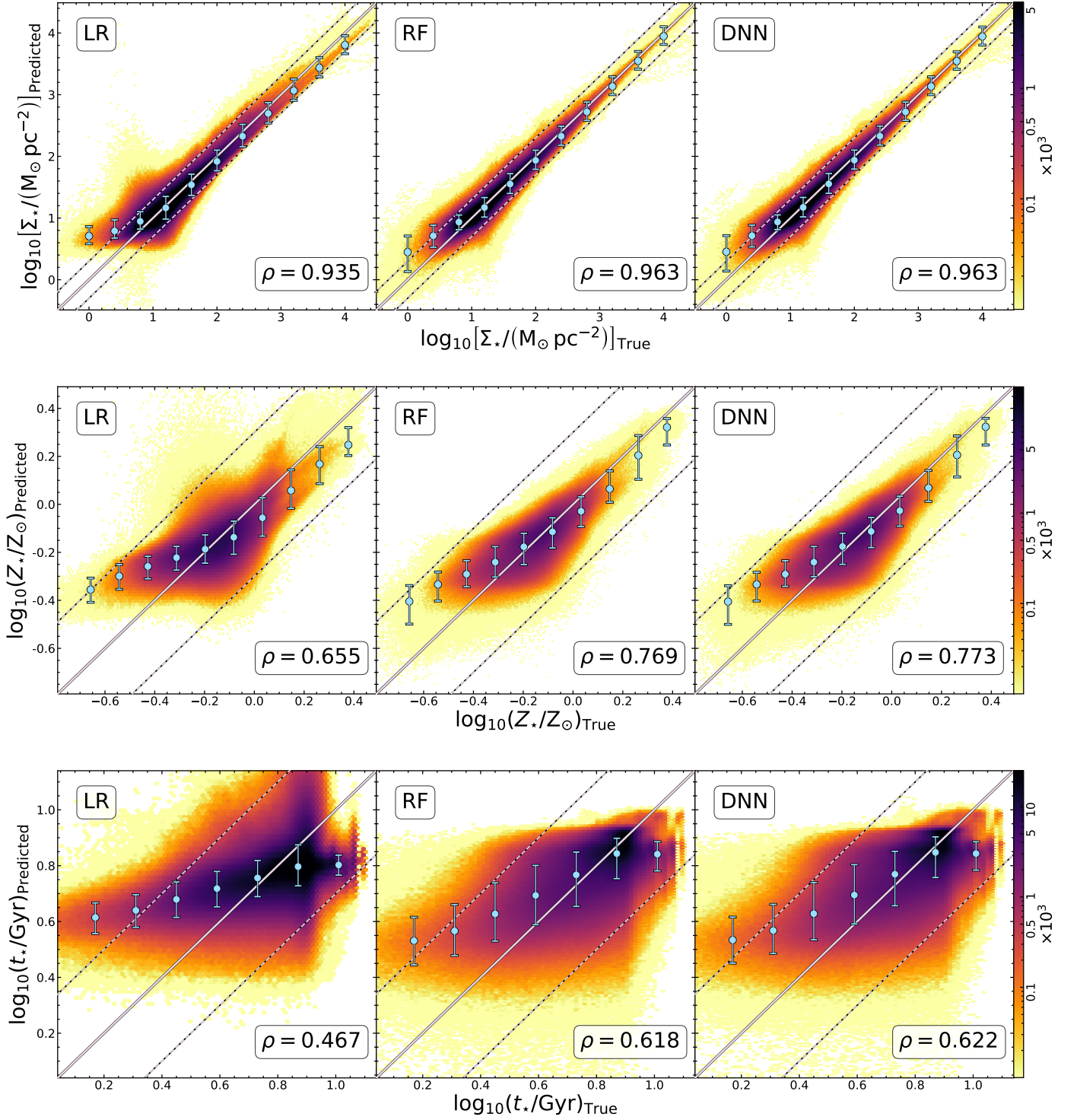


Fig. 3. Predictions for Σ_\star (top), Z_\star (middle), and t_\star (bottom) using the four *Euclid* bands as the input features for the three models: linear regression (left), RF (middle), and DNN (right) as a function of their true value. A diagonal line marks the identity, and the additional two dashed lines mark the limit for the catastrophic outliers (instances where the predicted value is two times larger or smaller than the true value). Running medians, with their 16th and 84th percentiles are marked with large blue points with error bars.

making predictions for regions with $t_\star \leq 9$ Gyr extremely difficult, and for regions with $t_\star \leq 2.5$ Gyr predictions are almost impossible. Neither algorithm seems to be able to recover these ages accurately at all. When only H_E band is used to model t_\star , $f_{\text{out}} \approx 5\%$, regardless of the algorithm. Additional features decrease this number somewhat for LR ($f_{\text{out}} \approx 3.5\%$), and more for RF and DNN ($f_{\text{out}} \approx 2.8\%$). However, the range for t_\star is very

small (spanning from $\sim 0 - 12.7$ Gyr in our entire data set, as discussed in Sect. 2.3), making the definition of f_{out} very broad in the case of this galaxy property. If we take a narrower limit (for example, 0.2 dex, i.e. ~ 1.5 times higher or lower than the true values), for the best case scenario (DNN with all input features), f_{out} increases to over 10%.

In conclusion, while some information can be inferred ($\rho \approx 0.62$ for the best case scenario, a DNN using the four *Euclid* bands), t_\star cannot be accurately predicted from the *Euclid* observation with ML.

4.4. Example galaxies

In this section we present two galaxies taken from the test set whose properties were inferred with a DNN using the four *Euclid* bands. As a reminder, when splitting our data set into training and test sets, we ensured to select the whole galaxies, meaning the test set consists of all pixels belonging to 115 whole galaxies, which were not involved in training of any of the algorithms.

We present these two galaxies, TNG 539333 as an example of a more accurately predicted galaxy, and TNG 294869 as a less accurately predicted galaxy. To select these two galaxies, we used the following procedure: first we chose the top and the bottom quantile of R^2 values for Σ_\star , and visually inspected them, opting for galaxies that are somewhat face-on, which also have complex features visible; after we located at least 10 of such galaxies for each of the two categories, we randomly selected one. The full disc mock observations in the four *Euclid* bands for these two galaxies can be seen in Fig. 4. We show a full disc map for both the true and predicted values of Σ_\star , Z_\star and t_\star , as well as their residuals, and radial profiles which were extracted with circular annuli as both galaxies are relatively face on. We focus our analysis in the inner region marked by the dashed circle and line in the residuals map and radial profile respectively. This area spans the radius equalling 12 kpc for TNG 539333 and 13 kpc for TNG 294869. Each point in the radial profile represents a median with error bars that mark the 16th and 84th percentile of values within a 1.5 kpc bin of the galactocentric radius. The metrics of accuracy (as described in Sect. 3) for each galaxy are presented in Table 4.

4.4.1. TNG 539333

In Fig. 5, we plot the full disc of TNG 539333 as an example of a good prediction by the algorithm. This galaxy is projected face-on, with a total stellar mass of $5.6 \times 10^{10} M_\odot$ and stellar half-mass radius of 4.4 kpc. It is a star-forming ($\text{SFR} \approx 5.54 M_\odot \text{ yr}^{-1}$) galaxy. It has a bright central bar where Σ_\star is predicted quite accurately (median with 16th and 84th percentile scatter being $-0.007^{+0.052}_{-0.067}$ dex for the inner 12 kpc). The Σ_\star is mildly over-estimated throughout the bright inner region of the disc, over-estimating the complex spiral structure seen in the inner part of the disc. This then reverses into a very a mild underestimation for the outer disc, which increases with radius, towards the sparser regions more affected by the Monte Carlo noise. With an $R^2 \approx 0.973$, it is in fact predicted slightly better than the general sample (where $R^2 \approx 0.927$). The radial profile is accurately recovered, with a bright central region, followed by a stronger drop towards the inner ~ 3 kpc, but it has a generally linear decrease throughout the remainder of the disc. The central region of the modelled profile is very well recovered, followed by an over-prediction that shows the strongest difference with a maximum of 0.18 dex at around 3 kpc. After ~ 6 kpc, the radial profiles mostly begin to overlap, with the difference between them increasing mildly towards the noisy edges of the full disc map.

Z_\star is under-predicted through most of the galactic disc ($-0.028^{+0.043}_{-0.039}$ dex), with the inner more metal rich spiral arms not being as prominent in the model as they are on the actual

map. The actual map of TNG 539333 shows the entire disc to be of a higher metallicity than the model, except for the area directly around the galactic bar, which is predicted to be more metal rich. At the very outer, noisy regions, as with Σ_\star , the relationship starts to mildly reverse. In essence, while the Z_\star map manages to recover a number of galactic features seen in the actual map, they are not as prominent, nor do they extend as far into the disc, and in fact assume a shape more reminiscent of the predicted Σ_\star map than the features seen in the actual Z_\star map. With $R^2 \approx 0.682$, the Z_\star is predicted with a higher accuracy than that shown in the general sample ($R^2 \approx 0.595$, as seen in Table 3). The radial profile is somewhat similar to Σ_\star , with a slightly stronger dip separating the central bin from the remainder of the disc where the profile shows a steady decline. The Z_\star is systematically under-predicted through the entire profile, with the highest difference being 0.056 dex at ~ 6 kpc, prominent as a spiral feature that is not well recovered in the full disc maps.

Of the three properties, t_\star is predicted with the lowest accuracy ($-0.041^{+0.048}_{-0.039}$ dex), in fact, with $R^2 \approx -2.653$ (Table 4), the prediction is worse than that expected of a random chance. Most features in the actual map are not being mimicked by the model. For example, the central region shows several large spiral arms expanding beyond the galactic bar filled with young stars, but the model only manages to mimic the shape of the galactic bar somewhat, and fails at the remainder of the structure, with the only older feature visible being the most prominent, extended spiral arm. This makes the model predict t_\star rather accurately around the galactic bar, but it over-predicts the ages strongly outside of this small ring, and under-predicts the ages inside the bar. The algorithm manages to mildly mimic the spiral structure outside of the central region populated with old stars, since the structure is visible both for Σ_\star and Z_\star , but t_\star is over-predicted and very flat throughout the entire disc. The shape of the radial profile is not that well recovered either. We expect an older central bin, followed by a steady, mild decrease to ~ 4 kpc, then a mild reversal. The modelled radial profile shows a younger centre which jumps to higher values at around 2 kpc, and remains fairly flat throughout the disc. The highest difference (~ 0.082 dex) is seen at the very centre.

We stress once more that Σ_\star spans a wider range compared to Z_\star and t_\star , as described in Sect. 2.3. These differences should be considered in regard to the results, meaning that Σ_\star is recovered more accurately than Z_\star and t_\star even if it presents with a higher scatter. In fact, the Pearson ρ coefficient (Table 4) shows a strong correlation with Σ_\star ($\rho \approx 0.995$), followed by Z_\star ($\rho \approx 0.878$) which is still recovered better than the general sample. However, $\rho \approx 0.013$ for t_\star points to stellar ages not being well recovered for that galaxy at all.

4.4.2. TNG 294869

TNG 294869 (Fig. 6) is an example of a galaxy in which some of the parameters are recovered less accurately. It is a moderately star-forming ($\text{SFR} \approx 3.04 M_\odot \text{ yr}^{-1}$) galaxy with total stellar mass $2.37 \times 10^{10} M_\odot$ and stellar half-mass radius 8.6 kpc. This galaxy has a bright centre with visible spiral arms, a feature where the algorithm generally struggles more in other galaxies as well. Both arms extend out of the disc; however, the bottom arm extends further. The Σ_\star is predicted rather accurately ($-0.052^{+0.074}_{-0.086}$ dex scatter for the inner 13 kpc), but on a somewhat lower scale in comparison to TNG 539333, and with $R^2 \approx 0.890$ (Table 4), it shows lower accuracy in comparison to the entire test sample. In most regions, except for a few inner spiral fea-

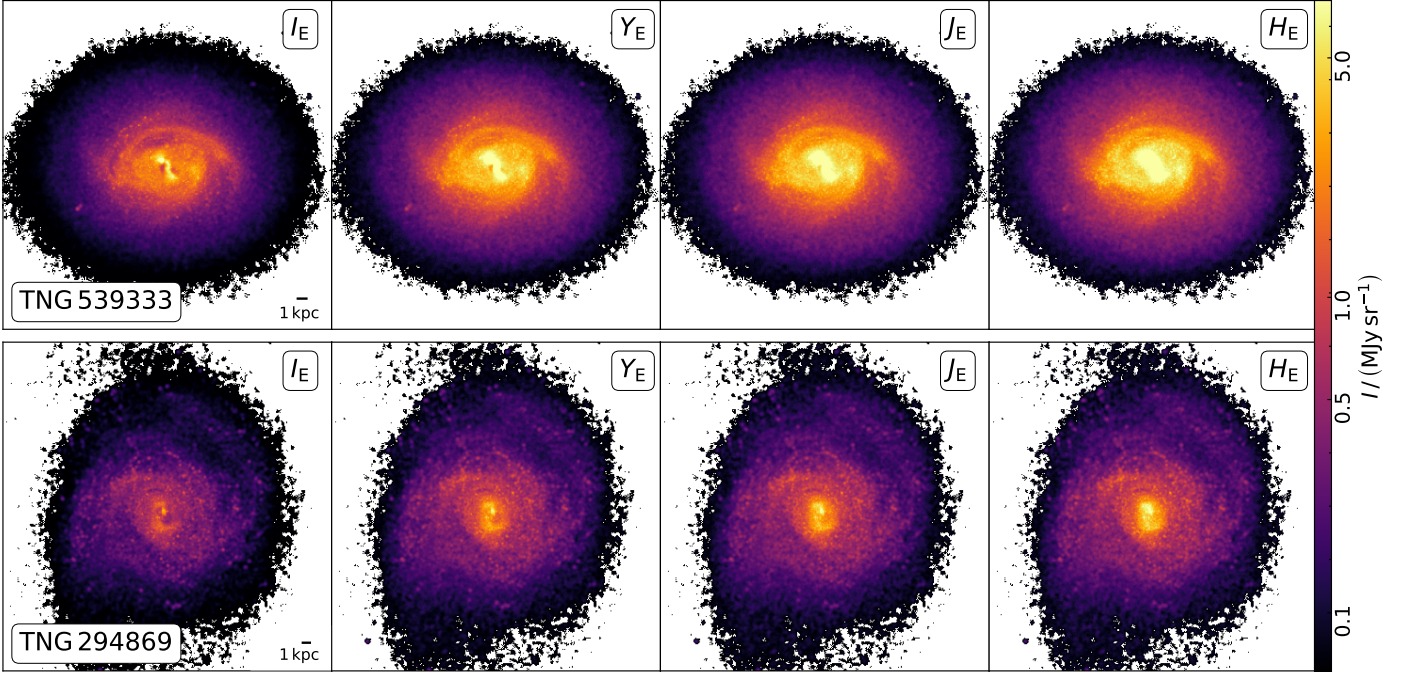


Fig. 4. Mock observations for our example galaxies TNG 539333 (top) and TNG 294869 (bottom) in four *Euclid* bands. From left to right: I_E , Y_E , J_E , and H_E . The bottom-right corner of the leftmost image for each galaxy shows the 1 kpc scale, which is equal to ten pixels with a 100 pc side.

Table 4. Accuracy scores (as defined in Sect. 3.3) for the three galaxy properties (Σ_\star , Z_\star , and t_\star) using four *Euclid* bands and a DNN for the two example galaxies TNG 539333 and TNG 294869.

Galaxy	Pixel count	Target	R^2	percentile	RMSE	NMAD	$\langle\Delta Y\rangle$	ρ
TNG 539333	54 030	$\Sigma_\star / (\text{M}_\odot \text{pc}^{-2})$	0.973	87	0.078	0.0518	-0.0016	0.995
		Z_\star	0.682	90	0.053	0.0502	-0.0259	0.878
		t_\star / Gyr	-2.653	26	0.059	0.0628	0.0349	0.013
TNG 294869	76 151	$\Sigma_\star / (\text{M}_\odot \text{pc}^{-2})$	0.890	12	0.115	0.1140	0.0702	0.968
		Z_\star	0.192	41	0.062	0.0629	-0.0337	0.693
		t_\star / Gyr	-2.520	28	0.117	0.1185	0.0717	0.082

Notes. Left to right: Galaxy; number of pixels; target; coefficient of determination (R^2); the percentile of the coefficient of determination for that galaxy; root-mean-square error (RMSE); normalised median absolute deviation (NMAD); and bias ($\langle\Delta Y\rangle$, where Y represents the target galaxy property); Pearson ρ coefficient.

tures, the algorithm over-predicts Σ_\star and this inaccuracy grows further out, a feature visible in the radial profiles as well. The Σ_\star takes the maximum in the central bin both for the true values, as well as the model, followed by a stronger decrease until ~ 4 kpc where the decline somewhat evens out. The highest difference (~ 0.29 dex) is seen in the central bin, as the algorithm seems to struggle the most with the central region of the galaxy. With $\rho \approx 0.968$, this galaxy shows a very similar correlation between predicted and true values for Σ_\star as compared to the full sample.

Z_\star model shows a somewhat larger scatter to TNG 539333 ($0.033^{+0.045}_{-0.047}$ dex). However, as with TNG 539333, Z_\star shows somewhat of a reversal in the predicted values when compared to Σ_\star . The innermost area of the galaxy shows a spherical feature populated with high metallicity stars. The shape of this feature is predicted by the algorithm; however, the values of Z_\star belong to the most under-predicted values in the entire galaxy. The stark contrast between the central bright feature, followed by a much darker, metal poor disc is not recovered, as it seems that the Z_\star map is again a scaled version of the Σ_\star map, where this contrast

between the centre and the remainder of the disc is not as prominent. In fact, a number of bright, overestimated pixels around the bright centre seen on the predicted Σ_\star map are also visible on the Z_\star map, albeit scaled down. In all, Z_\star is not well predicted, with $R^2 \approx 0.192$ (Table 4), which is significantly lower than the entire test sample. The radial profile for Z_\star assumes a shape similar to Σ_\star but with a steeper drop from the metal-rich central region; however, the area with $r \geq 4$ kpc is somewhat constant for the rest of the radial profile. The Z_\star radial profile is rather well recovered in this area as well, and the highest difference (0.13 dex) belongs to the central bin. Similar to Σ_\star , the Pearson correlation coefficient (Table 4) $\rho \approx 0.693$ is fairly close to the full test set ($\rho \approx 0.773$).

Surprisingly, the actual spiral structure outside of the very centre seen in the t_\star map is copied quite well, albeit mostly in reverse. The algorithm over-predicts the age in the outer parts of the galaxy, while under-predicting it in the spiral arms, with a $R^2 \approx -2.520$ (Table 4), pointing to the algorithm not being able to predict t_\star with any accuracy. Though interestingly, this value

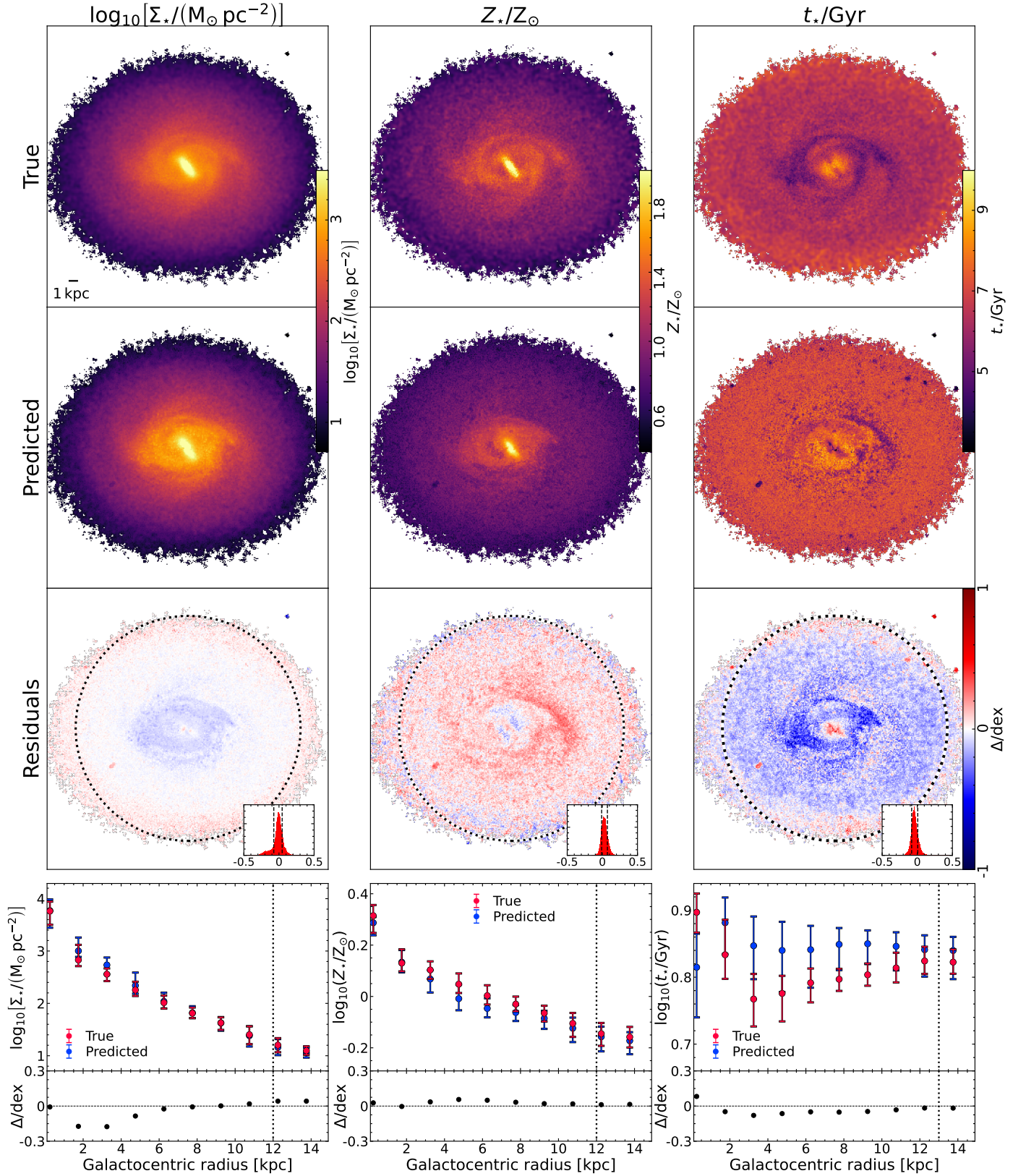


Fig. 5. Full disc map of galaxy TNG 539333 (O5 projection) for $\log_{10}[\Sigma_{*}/(\text{M}_{\odot} \text{pc}^{-2})]$ (left), $\log_{10}(Z_{*}/Z_{\odot})$ (middle), and $\log_{10}(t_{*}/\text{Gyr})$ (right) as an example of a more accurate model. From top to bottom: The top row shows the actual values for each property followed by the model created by a DNN algorithm on four *Euclid* bands, and the third row shows the residuals, where points in red (blue) represent pixels where the actual values are higher (lower) than those in the model. Histograms contained in the lower right corner of the residual maps show the distribution of error prediction ($\log_{10} y_i - \log_{10} \hat{y}_i$, where y_i is the actual i -th value, and \hat{y}_i is the predicted i -th value of the pixel), dashed lines mark the 16th and 84th percentile of the distribution for the pixels contained within the dashed circle of a 12 kpc radius. Bottom row shows radial profiles for the true (red) and predicted (blue) values where large points mark the median, and error bars mark the 16th and 84th percentile of the distribution contained within the radial bin, as well as their residuals plotted underneath.

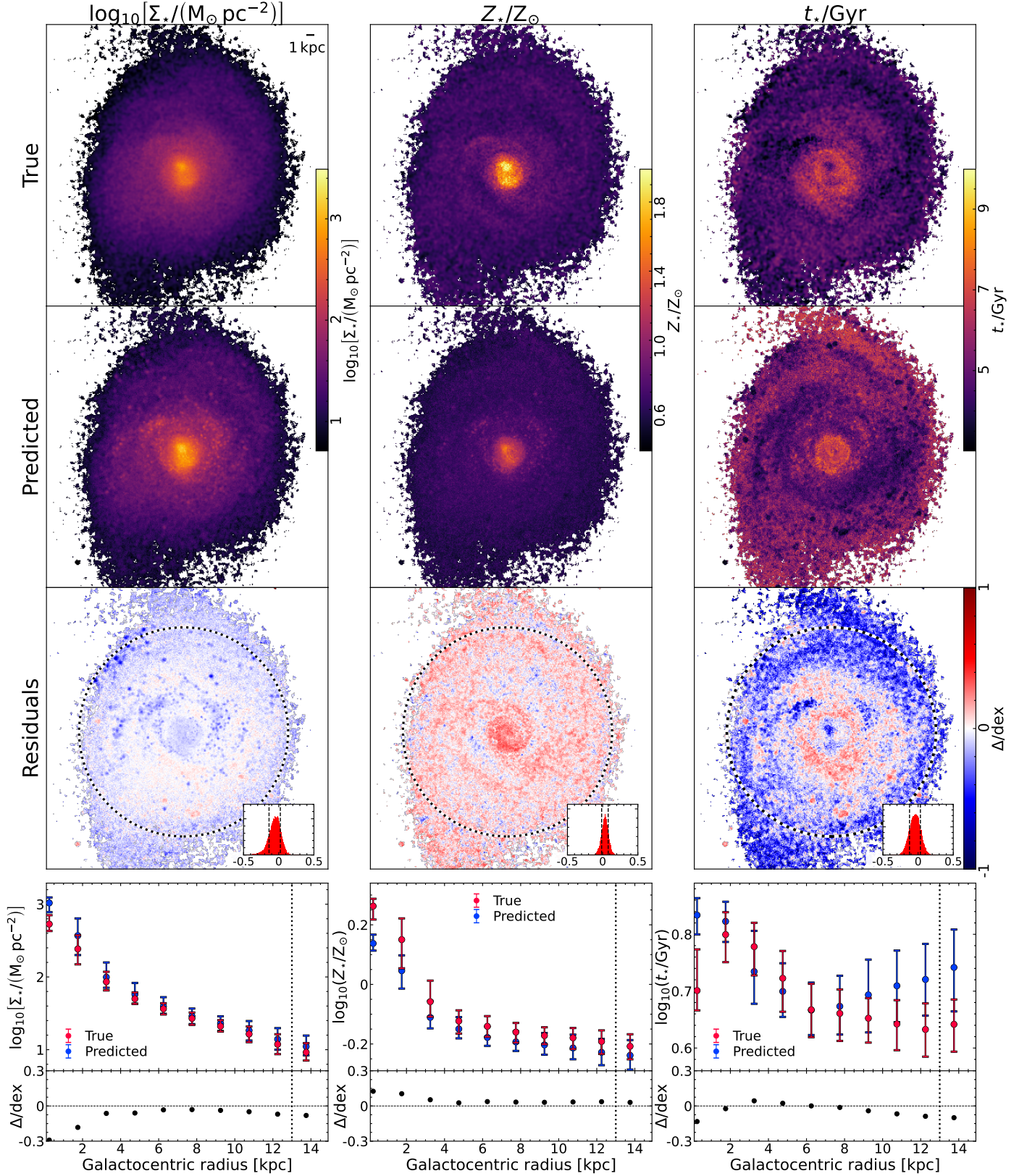


Fig. 6. Same as Fig. 5 but for the galaxy TNG 294869. The dashed circle marks the 13 kpc radius.

is somewhat better than for TNG 539333. Stellar age seems rather flat in the central region, while the actual map has visible structure, with younger stars concentrated in the very centre, surrounded by a thick disc of older stars. The algorithm shows a younger, albeit much smaller area at the very centre in a few

pixels; however, it struggles with the shape of the entire structure. The distant spiral arms in the outermost area of the galaxy show some of the youngest ages in the entire galaxy that the algorithm seems to struggle the most with, showing some of the largest under-predictions. However, it is worth noting, as seen in

Fig. 3, lower values of t_* , especially $t_* \leq 2.5$ Gyr, are very badly recovered regardless of the ML algorithm employed. In the case of this galaxy, the outermost regions fall within that range. t_* for TNG 294869 shows the highest scatter ($-0.044^{+0.076}_{-0.078}$ dex) out of all six cases discussed here, pointing to age being recovered the least accurately when taking into account that it also spans the smallest range. The radial profile is not that well recovered. The central bin is strongly over-predicted, with the highest difference (0.13 dex) in the entire disc. The model shows a young centre, with age that decreases consistently until ~ 6 kpc, where it then reverses and continues increasing, but with a very high scatter in the outer bins. The true values, on the other hand, show a young central bin, followed by a jump in age at the second bin (~ 2 kpc), and a decrease similar to the model. However, this decrease continues throughout the disc with a somewhat less steep fall at $r \geq 4$ kpc, but there is no reverse as there is in the model at $r \approx 7$ kpc. With $\rho \approx 0.082$, this galaxy, as well as TNG 539333 show a much lower correlation for t_* when compared to the full test set.

It is worth noting that Z_* is fairly under-predicted, while Σ_* and t_* are both under-predicted for both of our example galaxies. In general, the trend for the entire test set, in case of RF and DNN for when at least all four *Euclid* bands are used, shows the reversal in the bias for Z_* in comparison to Σ_* and t_* (Table 3). Not shown here, but when visually inspecting full disc maps of other test set galaxies, this trend was noticeable, though sometimes it is reversed, that is Z_* may be generally over-predicted, which is followed with Σ_* and t_* being under-predicted for the same galaxy.

4.5. Effects of the projection angle: Face-on versus edge-on

When inferring galaxy properties, the orientation of the observed galaxy should be taken into account. Each of the galaxies in the TNG50-SKIRT atlas is projected at five random angles fixed with respect to the simulation volume. When training our machine learning algorithm, we only use one image per galaxy; however, in case of our two example galaxies, we can directly compare inferred galaxy properties both in face-on and edge-on cases, as shown in Fig. 7.

The total accuracy scores for the edge-on case (Table 5) do not deviate much from the face-on case for either galaxy; however, the distribution of the pixels throughout the galactic disc does. For example, in TNG 539333, Z_* is under-predicted throughout most of the disc in the face-on projection, but in the edge-on case, it shows a strong over-prediction above and below the galactic disc. Similarly, t_* is over-predicted through most of the disc, which slowly turns into an under-prediction at the very edges in the face-on case, but in the edge-on case we can also observe a strong under-prediction above and below the disc.

TNG 294869 shows a similar behaviour as well, where the edge-on case prediction behaves quite similarly to its face-on counterpart, while showing a strong deviation from the fit in the area above and below the galactic disc. In these two example galaxies we can conclude that, even if the DNN can predict galaxy properties quite well, as is the case for Σ_* , for example, it struggles in the outer parts of the edge-on cases. In fact, it is the edge-on cases where it becomes very prominent that the Z_* and t_* maps behave in a manner very similar to a scaled version of the Σ_* map. This shows quite strongly in the residuals maps for Z_* and t_* , specifically as the true values for these two properties have a much stronger gradient throughout the galactic disc when compared to Σ_* , causing a very strong under- or over-prediction in the edge-on cases of both galaxies.

5. Discussion

5.1. Non-linearity and the importance of colour

We use three different methods to predict Σ_* , Z_* , and t_* , and to test the linearity of their relationship in regard to *Euclid* observations. We compare our results in three different approaches: using a single (best) *Euclid* band which in all cases proves to be H_E (though in the case of stellar age, a single band cannot be used to model it, with $R^2 \approx 0$ in all cases), using all four bands, and using ten features, consisting of four bands and six colours modelled from them. The choice for this approach was partly inspired by Acquaviva (2015), where the author shows that it is possible to recover gas-phase metallicity from SDSS five-band photometry with a RMSE < 0.1 dex, regardless of the ML algorithm chosen. They attempt to include both colours and squared colours into their input features, and show that, after M_* , $(g - i)$ colour is the most important feature for low- z , and $(g - z)^2$ for high- z test sets. Our results behave differently, as the inclusion of colour does not offer any additional benefit for any of the galaxy parameters (at least in the case of RF and DNN, which out-compete LR in every case). We believe that the difference lies in the fact that Acquaviva (2015) was using optical colours, which contain information on metallicity (Sanders et al. 2013), while in our case, we are using IR colours which do not seem to contain any additional information that the algorithms cannot already glean from the information contained in the four *Euclid* bands. While in this work we focus on the four *Euclid* bands, as well as colours modelled from them, we did also test whether additional information can be inferred from squared colours. The results are shown in Appendix C. The improvements to the accuracy of our method are negligible, and for that reason we do not include them in the main analysis.

Unfortunately, it can be difficult to disentangle the underlying processes that drive a ML algorithm to make the predictions that it makes. However, in this case, we assume that the reason why the six additional colours do not offer any additional benefit comes from the fact that they are easily derived from the four *Euclid* bands (essentially by dividing two input features), which should make them easily inferred by the algorithm, and thus they do not offer any new information that cannot already be extracted from the four bands during training. We test this by building a model that predicts *Euclid* colours from the four bands, and in all cases the algorithm converges very quickly with the worst accuracy ($R^2 = 0.9981$, RMSE = 0.0113 dex) belonging to $I_E - H_E$. The results can be seen in Appendix D. Either way, the addition of six colours offers negligible benefits for RF at best, and while in the case of DNN it helps the network to converge faster during training, their inclusion becomes negligible during deployment, and then it even comes with an additional computational cost, however minor, that would have to be expended on the creation of the colour maps for the new data. That is why, in the following discussion, we focus solely on the case of the four *Euclid* bands.

Regardless of the choice of ML algorithm, Σ_* of the simulated galaxy is predicted with a very high level of accuracy ($R^2 = 0.927$, RMSE = 0.130 dex, $\rho = 0.963$), followed by Z_* ($R^2 = 0.595$, RMSE = 0.076 dex, $\rho = 0.773$), then t_* ($R^2 = 0.382$, RMSE = 0.128 dex, $\rho = 0.623$) which neither of the approaches is able to predict accurately. Of course, we would like to stress once more that Σ_* range spans several orders of magnitude more than the other two properties (Sect. 2.3), making the Z_* and t_* models seem better in comparison in regard to the values of RMSE if that is not taken into account. Additionally, the Pearson correlation coefficient, ρ shows a clearer picture, pointing to a strong correlation between inferred and real

Table 5. Accuracy scores (as defined in Sect. 3.3) for the three galaxy properties (Σ_* , Z_* , and t_*) using four *Euclid* bands and a DNN for the galaxy TNG 539333 O2 projection and TNG 294869 O1 projection as an example of edge-on cases.

Galaxy	Pixel count	Target	R^2	RMSE	NMAD	$\langle\Delta Y\rangle$	ρ
TNG 539333	30 171	$\Sigma_*/(\text{M}_\odot \text{ pc}^{-2})$	0.971	0.098	0.0852	-0.0489	0.994
		Z_*	0.665	0.081	0.0673	0.0394	0.893
		t_*/Gyr	-1.522	0.097	0.0878	-0.0155	-0.345
TNG 294869	38 116	$\Sigma_*/(\text{M}_\odot \text{ pc}^{-2})$	0.963	0.093	0.0991	0.0298	0.983
		Z_*	0.175	0.089	0.1041	0.0664	0.769
		t_*/Gyr	-0.918	0.130	0.1479	0.0778	0.015

Notes. Left to right: Galaxy; number of pixels; target; coefficient of determination (R^2); root-mean-square error (RMSE); normalised median absolute deviation (NMAD); and bias ($\langle\Delta Y\rangle$, where Y represents the target galaxy property); Pearson ρ coefficient.

Σ_* , followed by Z_* , and then t_* which is very hard to infer with ML.

Σ_* shows a very strong linear relationship with the H_e -band surface brightness, and this is a preferred band regardless of the choice of the algorithm. The accuracy increases with the addition of the other three *Euclid* bands, but the increase is higher in the case of RF and DNN, pointing to the fact that their relationship to Σ_* comes with more non-linear features. The case is similar for Z_* ; however, the accuracy using only the H_e -band surface brightness is low enough that the addition of the other three bands creates a much better prediction, pointing to a stronger non-linear relationship between the features. ML performs poorly inferring t_* , regardless of the choice of algorithm.

5.2. Mass-metallicity relation

Generally, stellar metallicity is a difficult quantity to model, as it requires high-resolution observations in the optical range. The fact that any of the ML algorithms are capable of modelling it with such accuracy from *Euclid* observations requires a more focused discussion.

Stellar metallicity scales primarily with stellar mass (and quantities strongly related to it). This so-called stellar mass-metallicity relation (sMZR; for an overview, see Sect. 5.1.1 in [Maiolino & Mannucci 2019](#)) is originally based on global values due to the historical lack of spatially resolved data (e.g. [González Delgado et al. 2014](#)), which makes it difficult to determine the influence of local processes within galaxies. This relationship also appears when modelling stellar metallicity using ML. For example, [Simet et al. \(2021\)](#) use a neural network to predict the distribution of metallicity on semi-analytic galaxies tuned to match CANDELS observations ([Lu et al. 2014](#)), and they find a strong relationship between stellar mass and stellar metallicity in the simulations, which influences the accuracy of their stellar metallicity predictions, even though the algorithm itself is not aware of the complexity of underlying processes which guide this relationship. [Acquaviva \(2015\)](#), building on the work by [Sanders et al. \(2013\)](#) on the SDSS-DR7 ([Abazajian et al. 2009](#)) data, test the non-linearity of a relationship between a number of input features and gas-phase metallicity on the same data set. In the performance diagnostics of their non-linear extremely randomised trees (ERT) estimator, they show that mass is the most important feature for predicting gas-phase metallicity; however, information about luminosity, colours, and squared colours is required to better constrain the relationship.

Our data is spatially resolved at $100 \text{ pc} \times 100 \text{ pc}$ pixel scale, and yet we notice a similarly strong relationship between Σ_* and Z_* that positively affects the accuracy of predicting stellar metallicity. While the relationship between stellar mass and gas-phase metallicity in the TNG has been discussed before ([Torrey et al. 2019](#)), in this section we focus on the relationship between Σ_* and Z_* found in the simulation. We plot Z_* against Σ_* in Fig. 8 for our entire data set (left), as well as for the test data set, plotting the predicted Z_* values (top right), as well as the true Z_* values (bottom right), against true Σ_* . We calculate a running median, 16th and 84th percentile (light-blue points with error bars) for each of the Σ_* bins, showing a strong relationship within the TNG50 simulation itself. We notice a tighter relationship between the predicted Z_* , when compared to the true Z_* , and Σ_* , giving credence to our assumption that the surprisingly high accuracy of modelling Z_* is explained with this underlying Σ_* - Z_* relation.

We also plot a number of examples from the literature over the entire data set. For example, [González Delgado et al. \(2014\)](#) investigate the relationship between stellar metallicity and local mass surface density in 300 CALIFA integral survey ([Sánchez et al. 2012](#)) spheroidal and disc dominated galaxies. They find that metallicity in galaxies is driven by both global (M_*) and local (Σ_*) processes, and conclude that the local Σ_* is an important driver of Z_* in galaxies everywhere but their innermost, densest regions. [Zibetti & Gallazzi \(2022\)](#) use more than 600 000 individual regions of $\sim 1 \text{ kpc}$ linear size from a sample of 362 galaxies from the CALIFA integral survey as well, and investigate how local properties are affected by both the local stellar mass surface density, as well as the global stellar mass. The authors derive a local stellar mass surface density- r -band light weighted metallicity relation with a $< 0.1 \text{ dex}$ scatter, independent of stellar age, beyond a slightly higher ($\leq 0.1 \text{ dex}$) metallicity at younger regions. In an earlier work, [Zibetti et al. \(2020\)](#) already analysed a sample of 69 early-type galaxies from the CALIFA survey, noting that the stellar metallicity and local mass surface density relation is tighter (0.07 dex) than the relation between galactocentric radius and Z_* (0.10 dex), which points to stellar mass surface density being a better predictor of stellar metallicity than radial distance. [Neumann et al. \(2021\)](#) analyse more than 2.6 million spatial bins from 7439 nearby galaxies in the SDSS-IV MaNGA survey ([Bundy et al. 2015](#)) and derive a local stellar mass surface density-metallicity relation (rMZR), while also noting a positive trend of stellar age with both stellar mass surface density and stellar metallicity. [Baker et al. \(2022\)](#) discuss local mass-gas metallicity relation as well, and mention gas-stellar metallicity correlation derived by [Kauff-](#)

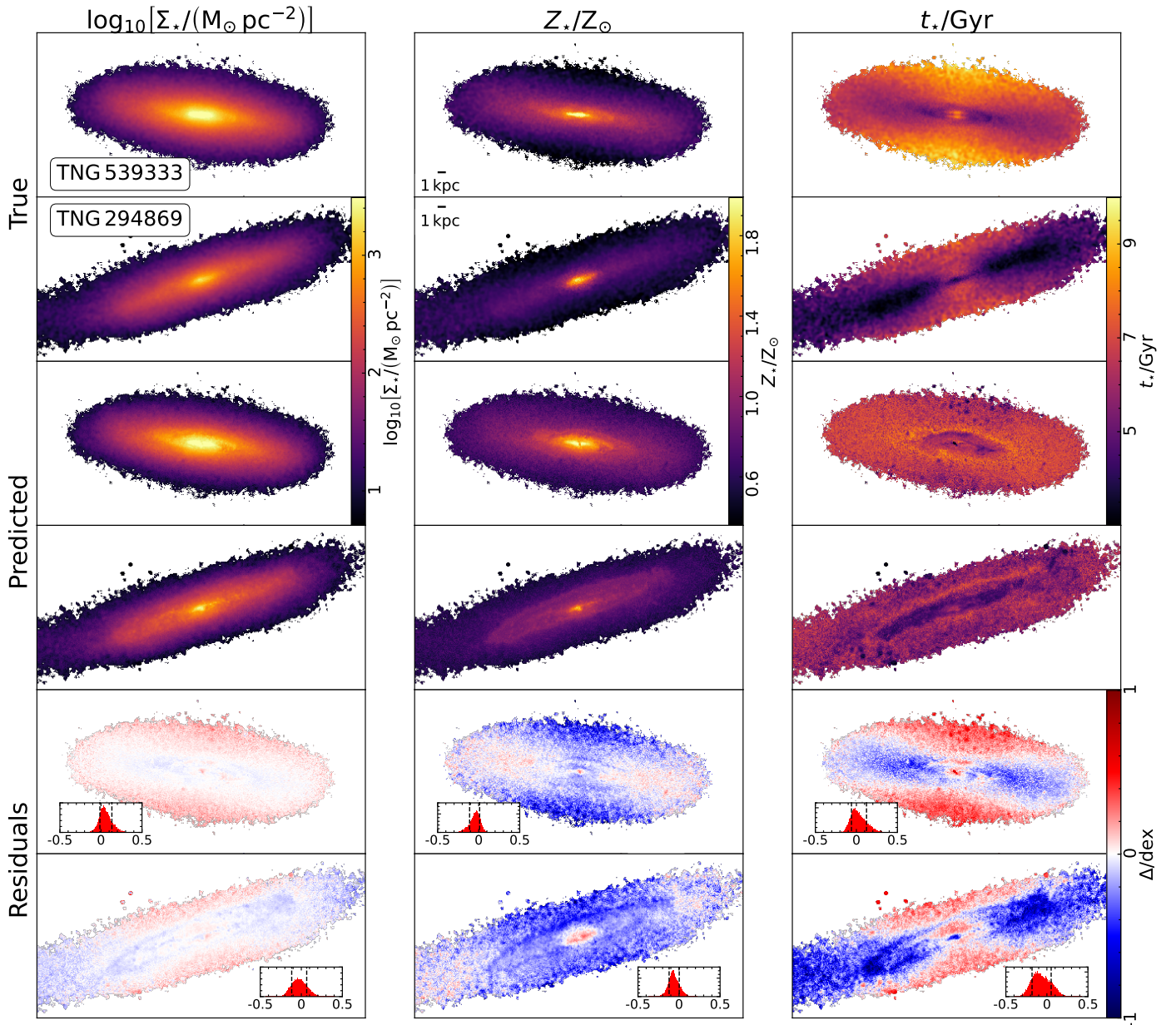


Fig. 7. Full disc map of galaxy TNG 539333 O2 projection (upper) and TNG 294869 O1 projection (lower) for $\log_{10} [\Sigma_{\star}/(\text{M}_{\odot} \text{pc}^{-2})]$ (left), Z_{\star}/Z_{\odot} (middle), and t_{\star}/Gyr (right) as an example of edge-on projection. From top to bottom: The top row shows the actual values for each property followed by the model created by a DNN algorithm on four *Euclid* bands, and the bottom row shows the residuals, where points in red (blue) represent pixels where the actual values are higher (lower) than those in the model. Histograms contained in the lower left or right corner of the residual maps show the distribution of error prediction ($\log_{10} y_i - \log_{10} \hat{y}_i$, where y_i is the actual i -th value, and \hat{y}_i is the predicted i -th value of the pixel), dashed lines mark the 16th and 84th percentile of the distribution.

mann et al. (2003). However, Sánchez (2020) warn that the local stellar metallicity- Σ_{\star} depends strongly on galaxy mass and morphology, properties which we do not consider in this procedure.

MZR has been taken as prior in some SED fitting codes on global scales (e.g. Webb et al. 2020; Bellstedt et al. 2021; Thorne et al. 2022; Euclid Collaboration: Abdurro’uf et al., in prep), with some promising results, especially in cases where photometry is too poor to measure metallicity. However, while the derived relations may show marginally improved results in extracting stellar metallicity, a caution should be taken since the relation itself can be affected by sample selection and methodology used.

While taking into account a number of issues with photometric based SED fitting (see, for example Nersesian et al. 2024, and references within), the strong relationship we find between Σ_{\star} and Z_{\star} on local scales in TNG50, as well as in several surveys outlined earlier makes for an interesting case for the addition of a rMZR prior to the SED fitting procedure to better constrain stellar metallicities on resolved scales.

5.3. Linear regression as a baseline

Using ordinary least squares LR, which assumes a simple linear relationship between the input features and the output target, we can extract the slopes for each of our input parameters, as well

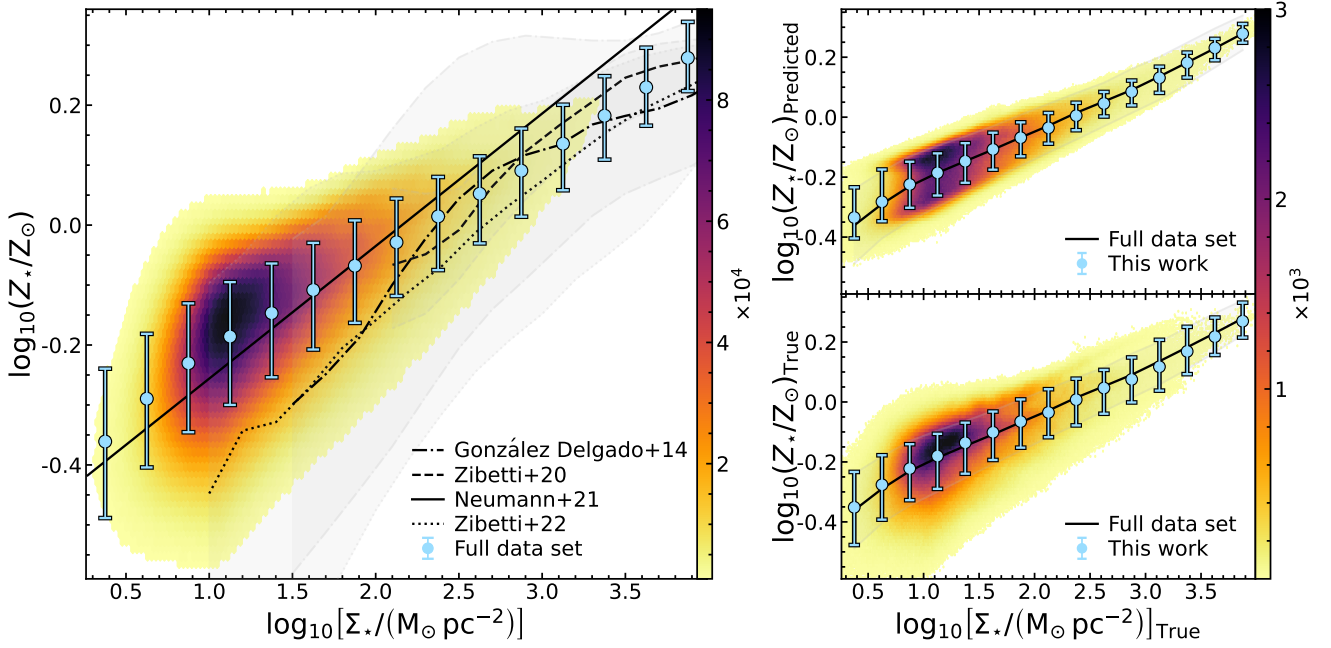


Fig. 8. Density plots of the distribution of $\log_{10}(Z_{\star}/Z_{\odot})$ as a function of $\log_{10}[\Sigma_{\star}/(\text{M}_{\odot} \text{pc}^{-2})]$. Large light-blue points with error bars represent the median, 16th and 84th percentile of the values of $\log_{10} Z_{\star}$ per each $\log_{10} \Sigma_{\star}$ bin. *Left:* The entire data set, consisting of 71 120 654 individual pixels. Only areas with ≥ 1000 pixels are included. Full, dashed, dash-dotted, and dotted lines represent fits from González Delgado et al. (2014), Zibetti et al. (2020), Neumann et al. (2021), and Zibetti & Gallazzi (2022), respectively. *Right:* The test data set (consisting of 9 035 270 pixels from 150 galaxies), predicted (top) and true (bottom) Z_{\star} as a function of true Σ_{\star} . Only areas with ≥ 10 pixels are included. Black line and the grey area marks the median, 16th and 84th percentile of the whole TNG50 data set (the light-blue points with error bars from the plot on the left), while the light-blue points with error bars represent the median, 16th and 84th percentile of $\log_{10} Z_{\star}$ for each $\log_{10} \Sigma_{\star}$ bin in the test set.

Table 6. Coefficients for the ordinary least squares LR, as used in Eq. (10), with *Euclid* bands in units of MJy sr^{-1} .

\hat{Y}	I_E	I_{Y_E}	I_{J_E}	I_{H_E}	b
$\Sigma_{\star}/(\text{M}_{\odot} \text{pc}^{-2})$	-11.113	4.941	2.156	7.924	1.446
Z_{\star}	-4.823	0.549	-0.318	3.178	0.045
t_{\star}/Gyr	-8.035	1.058	0.031	3.870	3.650

as the intercept, used to model our target feature

$$\log_{10} \hat{Y} = \sum_{i=1}^4 a_i \log_{10} X_i + b, \quad (10)$$

where \hat{Y} is the modelled galaxy property (target: $\Sigma_{\star}/(\text{M}_{\odot} \text{kpc}^{-2})$, Z_{\star} , or t_{\star}/Gyr), and X are the observational surface brightnesses of the four *Euclid* bands (input features: I_E , Y_E , J_E , and H_E) in units of MJy sr^{-1} . We show these values in Table 6. H_E most strongly correlated with all three galaxy properties, while I_E shows the strongest anti-correlation. However, as discussed in Sect. 4, H_E shows a strong linear relationship with Σ_{\star} , while other relationships fall into a more complex, non-linear regime.

5.4. Interpretability of machine learning predictions

Machine learning could offer a fast alternative to traditional methods when predicting galaxy parameters from mock observations, and with modern libraries does not come with a steep learning curve. Prior to our choice of RF, and DNN as the algorithms used to constrain galactic parameters from *Euclid* obser-

vations, we tested a number of other ML algorithms (AdaBoost, Extremely randomised trees, Ridge regression, Support vector machine, and MLPRegressor from the scikit-learn Python library) on a subset of our data set, and they all showed a similar level of accuracy with the default values of their hyperparameters. It is highly unlikely that the choice of algorithm would have a drastic effect on the final result, with the exception of the linear methods, as a number of relationships within our data set are non-linear, an effect that most likely stems from the fact that galaxies are complex systems with a complex formation and evolution history.

To test the non-linear component we choose RF, as it is a fairly beginner-friendly algorithm well guarded from user error, and is inherently good at avoiding over-fitting. In our case it also shows really high accuracy that can compete with a far more complex and computationally demanding DNN. RF takes up to several hours to train on an average desktop machine on a sample of over 1000 resolved galaxies when used out-of-the-box, and in this work we dedicated an effort to refine its hyper-parameters until the learning process took less than an hour without any losses to the accuracy.

In fact, due to its speed and interpretability, it should be a preferred algorithm for this work, since a DNN introduces a lot of complexity that is much harder to disentangle with the current tools. One of the main issues with using ML in general arises due to the opaque way in which ML algorithms operate. The goal of this work also extends to understanding the relationship between the observations and the galaxy properties that were mapped to them.

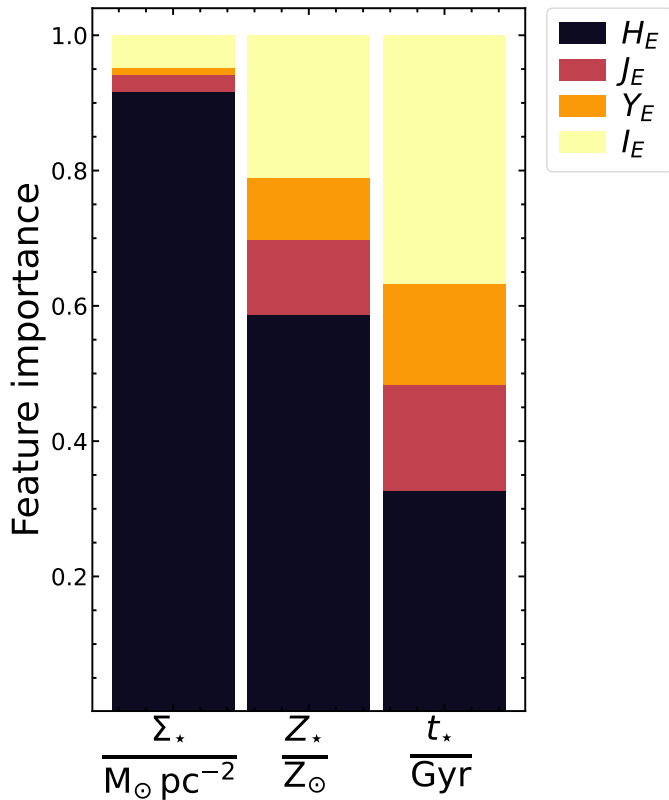


Fig. 9. Feature ranking for RF predictions of Σ_* , Z_* , and t_* . H_E is ranked as the most important feature for Σ_* and Z_* (91.5% and 58.6% respectively), while I_E (36.8%) and H_E (32.7%) are the two most important features for t_* .

5.4.1. Random forest and feature importance

As an ensemble method, RF offers feature importance (Gini index; Jost 2006), which ranks the features according to how much they contribute to the final prediction of the algorithm. The algorithm also introduces a certain level of randomness into feature selection, meaning that at each start of training, the algorithm chooses a new, random subset of input features, thus avoiding the possibility of a strong feature taking over during training, but also removes the requirement of having to shuffle features when loading them into the algorithm to avoid the risk of the first loaded feature being assigned higher importance than the ones loaded later. The values for feature importance are shown in Fig. 9. H_E ranks as by far the most important feature for Σ_* , and as we have seen earlier, mostly relates to Σ_* in a linear regime; however, the rest of the bands do offer an additional non-linear information that increases the predictive power for RF. The situation is somewhat similar, albeit weaker, for Z_* , which is an expected result due to the resolved mass-metallicity relation inherent in the cosmological simulation. None of the *Euclid* bands show a strong importance in modelling t_* , and t_* cannot be recovered accurately with ML from *Euclid* observations.

5.4.2. Choice of algorithm and missing data

In addition to feature importance offered by the RF, we also used the leave-one-feature-out (LOFO) method to assess the amount of information contributed to our result from each of the four *Euclid* surface brightnesses. In this method, we repeat the ML algorithm training process for each of the three galaxy param-

eters, but each time we leave out one of the four input features and compare the changes in the accuracy on a metric of choice. In our case, we decided on RMSE:

$$\text{LOFO}_X = \frac{\text{RMSE}_X - \text{RMSE}}{\text{RMSE}}, \quad (11)$$

where X is one of the input features left out during training (I_E , Y_E , J_E , or H_E). We show the results in Table 7. We repeat the same procedure for both our ML algorithms, RF and DNN. For RF, the accuracy of the prediction is most strongly affected with the removal of I_E , with LOFO_{I_E} being at least several times higher than the next highest value), and then H_E , but in the case of the latter, the decrease is very minor. This is followed by J_E , and then Y_E , whose removal has the smallest effect on RMSE for each of the three galaxy parameters. This may seem to contradict the feature importance values shown in Sect. 5.4.1, pointing to the necessity of using several methods when trying to interpret a ML model.

We explore this further in Fig. 10, where we plot the distribution of both the true and the inferred values for our entire data set. There it becomes obvious that the removal of I_E mostly affects the lower part of the distribution in case of all three galaxy parameters, while the effect of other three *Euclid* bands seems very minor and spread through the distribution. Only in case of Z_* does H_E have a prominent effect mostly on the lowest end.

DNN behaves quite similar to RF for most of our target values, except in the case where we try to predict t_* without including the I_E . As mentioned earlier, t_* cannot be inferred accurately even when using the entire data set, and the removal of Y_E , J_E , or H_E has only a minor effect on the already quite low accuracy. However, as most information on the lower end of the distribution seems to be contained within I_E , this effect is very strong in the case of RF, where the Pearson ρ coefficient drops from 0.6 to 0.36, while in the case of DNN ρ remains quite high at 0.59. We believe this makes a strong case for considering different algorithms when trying to infer data with ML. While both RF and DNN show similar results when using the entire data set and, as we mentioned earlier (Sect. 5.4), due to its speed and better interpretability, RF should be the better choice of an algorithm for our work, LOFO analysis shows the superior power of neural networks when managing missing data.

Unlike RF, DNN requires a dedicated computer architecture. Still, the algorithm was trained in under three days on the VSC Tier dedicated GPU machine. Once trained, the model can easily be saved and deployed on new galaxies, as well as new computer infrastructure. In fact, it only takes several minutes to finish modelling a map of a single galaxy on a sub-kpc (100 pc \times 100 pc) resolution on an 11th generation Intel i7 2.30 GHz 16-core laptop, something that is unfortunately still not possible with traditional methods, such as SED fitting. As the most complex method of the three used in this work, DNN is also hardest to interpret, but with the recent advances in ML, a number of methods have been developed to offer better interpretability of various algorithms, which should make using neural networks to solve astrophysical problems more feasible in the future (for an overview on interpreting neural networks, see e.g. Vilone & Longo 2021; Longo et al. 2024).

Regardless, as our results show, the choice of the algorithm only has a very minor effect on the accuracy of inferring galaxy parameters from *Euclid* observations on the full data set, with the differences only becoming prominent in the case of missing data and hard to infer values. All of these properties need to be taken

Table 7. Accuracy scores (as defined in Sect. 3.3 for the three galaxy properties (Σ_\star , Z_\star , and t_\star) for the LOFO analysis for the two ML algorithms (RF DNN).

Target	Feature	RF							DNN						
		R^2	RMSE	LOFO	f_{out}	NMAD	$\langle\Delta Y\rangle$	ρ	R^2	RMSE	LOFO	f_{out}	NMAD	$\langle\Delta Y\rangle$	ρ
$\Sigma_\star / (\text{M}_\odot \text{pc}^{-2})$	I_E	0.872	0.1727	0.3199	0.0712	0.1502	-0.0066	0.934	0.872	0.1724	0.3250	0.0706	0.1498	-0.0063	0.934
	Y_E	0.920	0.1365	0.0431	0.0379	0.1143	-0.0010	0.959	0.912	0.1428	0.0975	0.0450	0.1173	-0.0034	0.955
	J_E	0.918	0.1383	0.0570	0.0400	0.1143	-0.0010	0.958	0.918	0.1379	0.0597	0.0397	0.1140	-0.0009	0.958
	H_E	0.912	0.1432	0.0944	0.0450	0.1171	-0.0013	0.955	0.913	0.1427	0.0966	0.0445	0.1169	-0.0020	0.955
Z_\star	I_E	0.477	0.0866	0.1274	0.0029	0.0806	0.0007	0.692	0.479	0.0864	0.1327	0.0029	0.0802	0.0011	0.694
	Y_E	0.549	0.0804	0.0469	0.0009	0.0758	0.0000	0.743	0.524	0.0826	0.0823	0.0012	0.0784	-0.0023	0.726
	J_E	0.544	0.0808	0.0524	0.0010	0.0763	0.0000	0.740	0.546	0.0806	0.0567	0.0010	0.0761	0.0003	0.741
	H_E	0.531	0.0820	0.0678	0.0012	0.0769	0.0000	0.731	0.533	0.0818	0.0723	0.0012	0.0768	0.0011	0.732
t_\star / Gyr	I_E	0.123	0.1531	0.1861	0.0462	0.1252	-0.0123	0.359	0.338	0.1330	0.0355	0.0317	0.1046	-0.0072	0.587
	Y_E	0.349	0.1319	0.0215	0.0305	0.1045	-0.0081	0.595	0.351	0.1317	0.0255	0.0303	0.1043	-0.0079	0.597
	J_E	0.349	0.1319	0.0219	0.0307	0.1042	-0.0073	0.596	0.351	0.1317	0.0256	0.0306	0.1039	-0.0070	0.597
	H_E	0.336	0.1332	0.0321	0.0319	0.1049	-0.0073	0.585	0.337	0.1331	0.0360	0.0318	0.1047	-0.0070	0.586

Notes. Left to right: Target; *Euclid* surface brightness removed for LOFO analysis; coefficient of determination (R^2); root-mean-square error (RMSE); LOFO importance; fraction of catastrophic outliers (f_{out}); normalised median absolute deviation (NMAD); and bias ($\langle\Delta Y\rangle$, where Y represents the target galaxy property); Pearson ρ coefficient.

into account when choosing the best machine learning algorithm for the task.

6. Summary

In this work, we have tested the viability of using supervised ML to map mock *Euclid* observations of nearby galaxies with galaxy parameters extracted from the Illustris TNG50 simulation and processed by the Monte Carlo radiative transfer code SKIRT on sub-kiloparsec scales. We tested three distinct methods, an ordinary least squares LR and two ML algorithms (RF and DNN) with three different choices of input parameters, single *Euclid* band, all four *Euclid* bands, and ten input parameters consisting of both the four *Euclid* bands and six *Euclid* colours modelled from them, for three galaxy parameters, stellar mass surface density, mass-averaged stellar metallicity, and stellar age.

We showed that when using a single band, the H_E band performs the best when predicting Σ_\star and Z_\star in the case of all three algorithms. It is the most important input feature according to RF and has the second-strongest effect on the accuracy during the LOFO analysis. On the other hand, I_E is shown to be the second most important feature when building a tree in RF, but it has the strongest effect in the LOFO analysis, showcasing the importance of using several methods when trying to analyse ML results. Specifically, I_E seems to carry the most information from the lower end of the distribution for all three galaxy parameters, while H_E affects the entire range. Any single band cannot be used to model stellar age (with $R^2 \approx 0$ in all four cases).

We showed that the accuracy of modelling Σ_\star with only the H_E band is very high, with similar values of R^2 for all three algorithms, pointing to a strong linear relationship between Σ_\star and the H_E -band surface brightness. However, the accuracy increases slightly with the introduction of the three other bands, with this increase being higher in the case of RF and DNN, indicating that the information gained from the I_E , Y_E , and J_E bands belongs to a more non-linear regime. The introduction of colour does not offer any significant increase to accuracy, pointing to the fact that ML can extract all the necessary information from the four *Euclid* bands.

We also presented full disc maps of our model for two example galaxies: TNG 539333, which represents a better outcome, and TNG 294869, as an example of a worse outcome. We note

that the model performance struggles around more complex features (for example, spiral arms or dusty features) and on the outskirts of the galactic disc with low surface brightness, where the effects of Monte Carlo noise become more prominent. Additionally, we compared both face-on and edge-on projections for these galaxies and showed that even in cases of good prediction, the algorithm struggles in the outer parts of the edge-on cases, that is, the most distant regions above and below the galactic disc.

As expected, we cannot model stellar age from the *Euclid* observations; however, we do get an unexpectedly high accuracy when modelling Z_\star , and the behaviour of the predictions is similar to Σ_\star . We would not expect to be able to extract the information on Z_\star from the *Euclid* observations due to the fact that Z_\star is extracted from spectral absorption features in the optical regime, which only the I_E -band covers, but the I_E -band itself is too broad. However, we noticed a strong relationship between Σ_\star and Z_\star within the TNG50 simulation, which explains the unexpected quality of the Z_\star prediction. We compared this finding to earlier studies done on simulated (EAGLE) and observational (CALIFA, MaNGA) data. While the MZR is a well-known galactic relation on global data, in this work we hypothesise that this relation could be used on resolved (sub-kiloparsec) data as well.

7. Implications for *Euclid*

With the observations of thousands of nearby galaxies that we expect to get access to with *Euclid*, traditional methods used to extract galaxy parameters, such as SED fitting, become unfeasible due to their large computational time requirements. In this work, we offer an alternative method that has become standard in analysing large amounts of data. We trained two distinct ML algorithms on mock images of 1154 nearby galaxies extracted from the TNG50 simulation, compared them to an ordinary least squares LR, and showed that it is possible to map *Euclid* observations to stellar mass surface density on the sub-kiloparsec level with a high accuracy in a short amount of time. We also showed that though mapping *Euclid* observations to stellar metallicity and age is less robust, the information still holds value as a first-order approximation of stellar metallicities across large samples.

It is nevertheless important to stress that this work was performed on idealised data. The ML procedure offers a very lim-

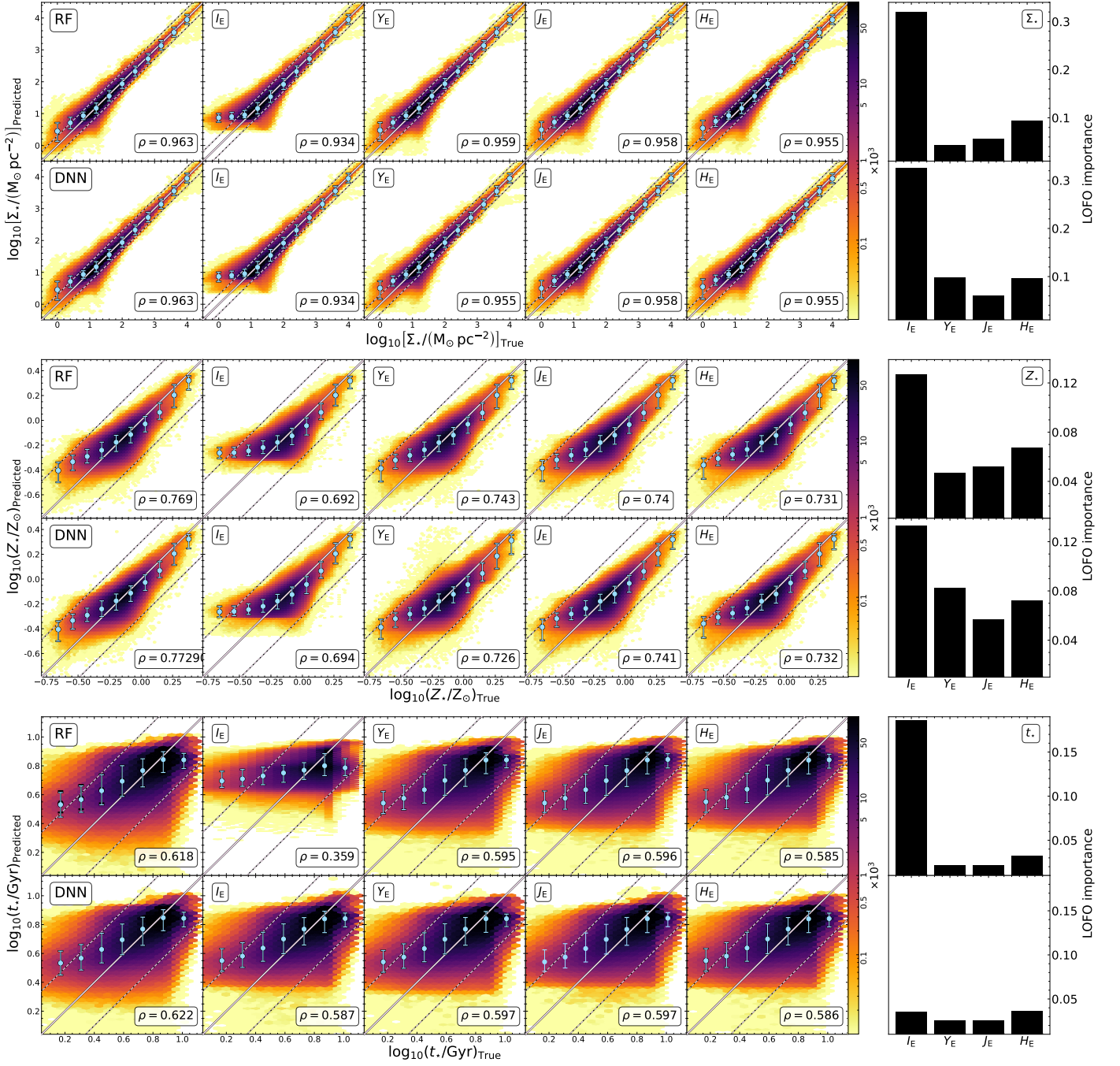


Fig. 10. Same as Fig. 3 but for LOFO analysis. Left to right: Predicted value when all four *Euclid* bands are used followed by four cases where the model was trained without one of the *Euclid* bands (marked in the upper-left corner). The LOFO importance is shown at the far right, calculated as Eq. (11). For each of the three target values, the top row shows the case when using RF and the bottom row when using DNN.

ited ability to analyse the underlying relationship between the input parameters and the output targets. The goal of this work was to focus on the pure physical properties, as this is the necessary first step in trying to apply machine learning algorithms to real life data. The inclusion of noise could potentially introduce additional complex underlying relationships that could be difficult to properly account for without the baseline set here. For this reason, we could not determine the aleatoric uncertainty of our algorithms. As a reminder, there are two types of uncertainties possible in an ML procedure: epistemic uncertainty, which is caused by the stochasticity present in the algorithm itself and informs one of how uncertain the algorithm is in its prediction,

and aleatoric uncertainty, which is caused by the presence of observational noise in the data set. For a discussion on epistemic uncertainty, which has been assessed in this work, we refer to Sect. 3.4. We plan to assess the aleatoric uncertainty in the future, for which we will have to perform the entire ML procedure first on idealised data and then on data with noise added to it. This will require adding a field of stars to mimic the presence of contaminants, Poissonian noise to mimic effects of photon shot noise, Gaussian noise to simulate observational background, and RMS noise to assign uncertainties to the images themselves. For details on this procedure, we refer to *Euclid* Collaboration: Merlin et al. (2023). The resolution of the images will differ from

the current $100\text{ pc} \times 100\text{ pc}$ once the *Euclid* PSF is taken into account, potentially affecting the accuracy of our method. However, we have performed some initial tests on different resolutions, and they have shown that the effects of resolution should be very minor at best.

Additionally, we have treated each pixel as an individual data point, ignoring the spatial information that could be gathered from the neighbouring pixels. This information could be accessed, for instance, via a CNN setup, which we plan to deploy in a future work. That way, having both the information contained in individual pixels as well as when including the spatial information from the neighbouring pixels, we should be able to gain a better understanding of how much spatial information contributes to the accuracy of inferring galaxy parameters with ML.

Another potential way to improve these results would be to take advantage of the relationships between our three targets and use chained regression (e.g. Euclid Collaboration: Humphrey et al., in prep). In our initial tests, the inclusion of Σ_* in the input features increased the accuracy in modelling Z_* , and the inclusion of Z_* increased the accuracy in t_* . We also plan to combine the *Euclid* data with other (specifically optical) data to test how much we can increase the accuracy of ML at predicting galaxy properties from observations. We are also aware of the fact that algorithms on simulated data might not transfer well into actual observations (e.g. Zanisi et al. 2021, where the authors notice a discrepancy between simulated galaxies from TNG50/TNG100 simulations, and SDSS). And finally, we will test our model on actual observations and compare it with traditional methods currently being developed (Euclid Collaboration: Abdurro'uf et al., in prep; Euclid Collaboration: Nersesian et al., in prep) in order to see how feasible it is to use supervised ML to accurately infer galaxy parameters, such as Σ_* , Z_* , and t_* , from actual *Euclid* observations.

Acknowledgements. We thank the referee for useful comments that helped to improve the quality of the manuscript. Co-funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them. IK, MB and AN acknowledge support from the Belgian Science Policy Office (BELSPO) through the PRODEX project “Belgian Euclid Science Exploitation (BESE)” (No. 4000143202). JHK acknowledges grant PID2022-136505NB-I00 funded by MCIN/AEI/10.13039/501100011033 and EU, ERDF. We wish to thank the “Summer School for Astrostatistics in Crete” for providing training on the statistical methods adopted in this work. The Euclid Consortium acknowledges the European Space Agency and a number of agencies and institutes that have supported the development of *Euclid*, in particular the Academy of Finland, the Agenzia Spaziale Italiana, the Belgian Science Policy, the Canadian Euclid Consortium, the French Centre National d’Etudes Spatiales, the Deutsches Zentrum für Luft- und Raumfahrt, the Danish Space Research Institute, the Fundação para a Ciência e a Tecnologia, the Ministerio de Ciencia e Innovación, the National Aeronautics and Space Administration, the National Astronomical Observatory of Japan, the Nederlandse Onderzoeksschool Voor Astronomie, the Norwegian Space Agency, the Romanian Space Agency, the State Secretariat for Education, Research and Innovation (SERI) at the Swiss Space Office (SSO), and the United Kingdom Space Agency. A complete and detailed list is available on the *Euclid* web site (<http://www.euclid-ec.org>).

References

Abadi, M., Agarwal, A., Barham, P., et al. 2015, TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems, software available from tensorflow.org
 Abazajian, K. N., Adelman-McCarthy, J. K., Agüeros, M. A., et al. 2009, ApJS, 182, 543
 Abdalla, F. B., Banerji, M., Lahav, O., & Rashkov, V. 2011, MNRAS, 417, 1891
 Abdurro'uf, Accetta, K., Aerts, C., et al. 2022a, ApJS, 259, 35
 Abdurro'uf, Coe, D., Jung, I., et al. 2023, ApJ, 945, 117

Abdurro'uf, Lin, Y.-T., Hirashita, H., et al. 2022b, ApJ, 935, 98
 Acquaviva, V. 2015, MNRAS, 456, 1618
 Alsing, J., Peiris, H., Mortlock, D., Leja, J., & Leistedt, B. 2023, ApJS, 264, 29
 Alsing, J., Thorp, S., Deger, S., et al. 2024, ApJS, 274, 12
 Baes, M. & Dejonghe, H. 2001, MNRAS, 326, 722
 Baes, M., Gebek, A., Trčka, A., et al. 2024, A&A, 683, A181
 Baes, M., Verstaappen, J., De Looze, I., et al. 2011, ApJS, 196, 22
 Baker, W. M., Maiolino, R., Belfiore, F., et al. 2022, MNRAS, 519, 1149
 Baron, D. 2019, arXiv:1904.07248
 Bell, E. F. & de Jong, R. S. 2001, ApJ, 550, 212
 Bellstedt, S., Robotham, A. S. G., Driver, S. P., et al. 2021, MNRAS, 503, 3309
 Breiman, L. 2001, Machine Learning, 45, 5
 Brescia, M., Cavuoti, S., D'Abrusco, R., Longo, G., & Mercurio, A. 2013, ApJ, 772, 140
 Brinchmann, J., Charlot, S., White, S. D. M., et al. 2004, MNRAS, 351, 1151
 Bruzual, G. & Charlot, S. 1993, ApJ, 405, 538
 Bruzual, G. & Charlot, S. 2003, MNRAS, 344, 1000
 Bundy, K., Bershady, M. A., Law, D. R., et al. 2015, ApJ, 798, 7
 Calzetti, D., Armus, L., Bohlin, R. C., et al. 2000, ApJ, 533, 682
 Camps, P. & Baes, M. 2018, ApJ, 861, 80
 Camps, P. & Baes, M. 2020, Astronomy and Computing, 31, 100381
 Camps, P., Misselt, K., Bianchi, S., et al. 2015, A&A, 580, A87
 Carliles, S., Budavári, T., Heinis, S., Priebe, C., & Szalay, A. S. 2010, ApJ, 712, 511
 Cavuoti, S., Brescia, M., Longo, G., & Mercurio, A. 2012, A&A, 546, A13
 Chabrier, G. 2003, PASP, 115, 763
 Chamba, N., Trujillo, I., & Knapen, J. H. 2022, A&A, 667, A87
 Charlot, S. & Bruzual, G. 1991, ApJ, 367, 126
 Charlot, S. & Fall, S. M. 2000, ApJ, 539, 718
 Chicco, D., Warrens, M., & Jurman, G. 2021, PeerJ Computer Science, 486, 5104
 Collister, A. A. & Lahav, O. 2004, PASP, 116, 345
 Conroy, C. 2013, ARA&A, 51, 393
 Cropper, M., Pottinger, S., Azzollini, R., et al. 2018, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 10698, Space Telescopes and Instrumentation 2018: Optical, Infrared, and Millimeter Wave, ed. M. Lystrup, H. A. MacEwen, G. G. Fazio, N. Batalha, N. Siegler, & E. C. Tong, 1069828
 D'Abrusco, R., Staiano, A., Longo, G., et al. 2007, ApJ, 663, 752
 Dahl, G. E., Sainath, T. N., & Hinton, G. E. 2013, in 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, 8609–8613
 Dale, D. A. & Helou, G. 2002, ApJ, 576, 159
 Declair, M., De Looze, I., Boquien, M., et al. 2019, MNRAS, 486, 743
 Delli Veneri, M., Cavuoti, S., Brescia, M., Longo, G., & Riccio, G. 2019, MNRAS, 486, 1377
 D'Isanto, A. & Polsterer, K. L. 2018, A&A, 609, A111
 Ditrani, F. R., Longhetti, M., La Barbera, F., et al. 2023, A&A, 677, A93
 Dobbels, W. & Baes, M. 2021, A&A, 655, A34
 Dobbels, W., Baes, M., Viaene, S., et al. 2020, A&A, 634, A57
 Draine, B. T. 2003, ARA&A, 41, 241
 Duchi, J., Hazan, E., & Singer, Y. 2011, Journal of Machine Learning Research, 12, 2121
 Euclid Collaboration: Aussel, B., Kruk, S., Walmsley, M., et al. 2024, A&A, 689, A274
 Euclid Collaboration: Bisigello, L., Conselice, C. J., Baes, M., et al. 2023, MNRAS, 520, 3529
 Euclid Collaboration: Bretonnière, H., Huertas-Company, M., Boucaud, A., et al. 2022, A&A, 657, A90
 Euclid Collaboration: Bretonnière, H., Kuchner, U., Huertas-Company, M., et al. 2023, A&A, 671, A102
 Euclid Collaboration: Desprez, G., Paltani, S., Coupon, J., et al. 2020, A&A, 644, A31
 Euclid Collaboration: Enia, A., Bolzonella, M., Pozzetti, L., et al. 2024, A&A, 691, A175
 Euclid Collaboration: Humphrey, A., Bisigello, L., Cunha, P. A. C., et al. 2023, A&A, 671, A99
 Euclid Collaboration: Jahnke, K., Gillard, W., Schirmer, M., et al. 2024, A&A, accepted, arXiv:2405.13493
 Euclid Collaboration: Leuzzi, L., Meneghetti, M., Angora, G., et al. 2024, A&A, 681, A68
 Euclid Collaboration: Mellier, Y., Abdurro'uf, Acevedo Barroso, J., et al. 2024, A&A, accepted, arXiv:2405.13491
 Euclid Collaboration: Merlin, E., Castellano, M., Bretonnière, H., et al. 2023, A&A, 671, A101
 Euclid Collaboration: Scaramella, R., Amiaux, J., Mellier, Y., et al. 2022, A&A, 662, A112
 Euclid Collaboration: Schirmer, M., Jahnke, K., Seidel, G., et al. 2022, A&A, 662, A92
 Faber, S. M. 1973, ApJ, 179, 731

- Gallazzi, A., Charlot, S., Brinchmann, J., White, S. D. M., & Tremonti, C. A. 2005, *MNRAS*, 362, 41
- Galliano, F., Nersesian, A., Bianchi, S., et al. 2021, *A&A*, 649, A18
- Gao, Y., Bao, M., Yuan, Q., et al. 2018, *ApJ*, 869, 15
- Genel, S., Vogelsberger, M., Springel, V., et al. 2014, *MNRAS*, 445, 175
- Gilda, S., Lower, S., & Narayanan, D. 2021, *ApJ*, 916, 43
- Glorot, X., Bordes, A., & Bengio, Y. 2011, in *Proceedings of Machine Learning Research*, Vol. 15, *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, ed. G. Gordon, D. Dunson, & M. Dudík (Fort Lauderdale, FL, USA: PMLR), 315–323
- González Delgado, R. M., Cid Fernandes, R., García-Benito, R., et al. 2014, *ApJ*, 791, L16
- Goodfellow, I., Bengio, Y., & Courville, A. 2016, *Deep Learning* (MIT Press), <http://www.deeplearningbook.org>
- Hastie, T., Tibshirani, R., & Friedman, J. 2009, *The elements of statistical learning: Data mining, inference, and prediction*, second edition (Springer Science & Business Media)
- Hildebrandt, H., Arnouts, S., Capak, P., et al. 2010, *A&A*, 523, A31
- Höhle, J. & Höhle, M. 2009, *ISPRS Journal of Photogrammetry and Remote Sensing*, 64, 398
- Hüllermeier, E. & Waegeman, W. 2021, *Machine Learning*, 110, 457
- Iglesias-Navarro, P., Huertas-Company, M., Martín-Navarro, I., Knapen, J. H., & Pernet, E. 2024, *A&A*, 689, A58
- Ivezić, Ž., Kahn, S. M., Tyson, J. A., et al. 2019, *ApJ*, 873, 111
- James, G., Witten, D., Hastie, T., Tibshirani, R., & Taylor, J. 2023, *An Introduction to Statistical Learning: with Applications in Python* (Springer)
- Jost, L. 2006, *Oikos*, 113, 363
- Kauffmann, G., Heckman, T. M., White, D. M. S., et al. 2003, *MNRAS*, 341, 33
- Kendall, A. & Gal, Y. 2017, in *Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS'17* (Red Hook, NY, USA: Curran Associates Inc.), 5580–5590
- Kingma, D. P. & Ba, J. 2015, in *3rd International Conference on Learning Representations, ICLR 2015*, San Diego, CA, USA, May 7–9, 2015, *Conference Track Proceedings*, ed. Y. Bengio & Y. LeCun
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. 2017, *Commun. ACM*, 60, 84–90
- Laigache, G., Dole, H., Puget, J. L., et al. 2004, *ApJS*, 154, 112
- Laureijs, R., Amiaux, J., Arduini, S., et al. 2011, *arXiv:1110.3193*
- Law, K.-H., Gordon, K. D., & Misselt, K. A. 2018, *ApJS*, 236, 32
- Leistedt, B., Alsing, J., Peiris, H., Mortlock, D., & Leja, J. 2023, *ApJS*, 264, 23
- Leja, J., Carnall, A. C., Johnson, B. D., Conroy, C., & Speagle, J. S. 2019, *ApJ*, 876, 3
- Leja, J., Johnson, B. D., Conroy, C., van Dokkum, P. G., & Byler, N. 2017, *ApJ*, 837, 170
- Lequeux, J., Peimbert, M., Rayo, J. F., Serrano, A., & Torres-Peimbert, S. 1979, *A&A*, 80, 155
- Lo Faro, B., Buat, V., Roehly, Y., et al. 2017, *MNRAS*, 472, 1372
- Longo, L., Brcic, M., Cabitza, F., et al. 2024, *Information Fusion*, 106, 102301
- Lovell, C. C., Acquaviva, V., Thomas, P. A., et al. 2019, *MNRAS*, 490, 5503
- Lu, Y., Wechsler, R. H., Somerville, R. S., et al. 2014, *ApJ*, 795, 123
- Luhman, M. L., Satyapal, S., Fischer, J., et al. 2003, *ApJ*, 594, 758
- Ly, C., Malkan, M. A., Rigby, J. R., & Nagao, T. 2016, *ApJ*, 828, 67
- Maiolino, R. & Mannucci, F. 2019, *A&A Rev.*, 27, 3
- Marinacci, F., Vogelsberger, M., Pakmor, R., et al. 2018, *MNRAS*, 480, 5113
- Matteucci, F. 2008, in *Massive Stars as Cosmic Engines*, ed. F. Bresolin, P. A. Crowther, & J. Puls, Vol. 250, 391–400
- Murphy, K. 2012, *Machine Learning: A Probabilistic Perspective*, *Adaptive Computation and Machine Learning series* (MIT Press)
- Naiman, J. P., Pillepich, A., Springel, V., et al. 2018, *MNRAS*, 477, 1206
- Nair, V. & Hinton, G. E. 2010, in *Proceedings of the 27th International Conference on International Conference on Machine Learning, ICML'10* (Madison, WI, USA: Omnipress), 807–814
- Nelson, D., Pillepich, A., Springel, V., et al. 2019a, *MNRAS*, 490, 3234
- Nelson, D., Springel, V., Pillepich, A., et al. 2019b, *Computational Astrophysics and Cosmology*, 6, 2
- Nersesian, A., van der Wel, A., Gallazzi, A., et al. 2024, *A&A*, 681, A94
- Nersesian, A., Xilouris, E. M., Bianchi, S., et al. 2019, *A&A*, 624, A80
- Neumann, J., Thomas, D., Maraston, C., et al. 2021, *MNRAS*, 508, 4844
- Pacifici, C., Iyer, K. G., Mobasher, B., et al. 2023, *ApJ*, 944, 141
- Pearson, K. 1895, *Proceedings of the Royal Society of London Series I*, 58, 240
- Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, *Journal of Machine Learning Research*, 12, 2825
- Pellerin, A. & Finkelstein, S. L. 2010, *Proceedings of the International Astronomical Union*, 262, 283
- Pillepich, A., Nelson, D., Hernquist, L., et al. 2017, *MNRAS*, 475, 648
- Pillepich, A., Nelson, D., Springel, V., et al. 2019, *MNRAS*, 490, 3196
- Pillepich, A., Springel, V., Nelson, D., et al. 2018, *MNRAS*, 473, 4077
- Planck Collaboration: Ade, P. A. R., Aghanim, N., Arnaud, M., et al. 2016, *A&A*, 594, A13
- Poggianti, B. M. & Barbaro, G. 1997, *A&A*, 325, 1025
- Popesso, P., Concas, A., Cresci, G., et al. 2023, *MNRAS*, 519, 1526
- Rosales-Ortega, F. F., Sánchez, S. F., Iglesias-Páramo, J., et al. 2012, *ApJ*, 756, L31
- Sadeh, I., Abdalla, F. B., & Lahav, O. 2016, *PASP*, 128, 104502
- Salim, S., Boquien, M., & Lee, J. C. 2018, *ApJ*, 859, 11
- Salim, S. & Narayanan, D. 2020, *ARA&A*, 58, 529
- Salucci, P. 2019, *A&A Rev.*, 27, 2
- Sánchez, S. F. 2020, *ARA&A*, 58, 99
- Sánchez, S. F., Kennicutt, R. C., Gil de Paz, A., et al. 2012, *A&A*, 538, A8
- Sánchez, S. F., Rosales-Ortega, F. F., Jungwiert, B., et al. 2013, *A&A*, 554, A58
- Sanders, N. E., Levesque, E. M., & Soderberg, A. M. 2013, *ApJ*, 775, 125
- Schmidt, M. 1959, *ApJ*, 129, 243
- Siebenmorgen, R. & Krügel, E. 2007, *A&A*, 461, 445
- Simet, M., Chartab, N., Lu, Y., & Mobasher, B. 2021, *ApJ*, 908, 47
- Sorba, R. & Sawicki, M. 2015, *MNRAS*, 452, 235
- Speagle, J. S., Steinhardt, C. L., Capak, P. L., & Silverman, J. D. 2014, *ApJS*, 214, 15
- Springel, V. 2010, *MNRAS*, 401, 791
- Springel, V., Pakmor, R., Pillepich, A., et al. 2017, *MNRAS*, 475, 676
- Steinacker, J., Baes, M., & Gordon, K. D. 2013, *ARA&A*, 51, 63
- Stensbo-Smidt, K., Gieseke, F., Igel, C., Zirm, A., & Steenstrup Pedersen, K. 2017, *MNRAS*, 464, 2577
- Surana, S., Wadadekar, Y., Bait, O., & Bhosale, H. 2020, *MNRAS*, 493, 4808
- Tagliaferri, R., Longo, G., Andreon, S., et al. 2003, in *Lecture Notes in Computer Science*, Vol. 2859, 226–234
- Tammes, P. M. L. 1930, *Recueil des travaux botaniques néerlandais*, 27, 1
- Thorne, J. E., Robotham, A. S. G., Bellstedt, S., et al. 2022, *MNRAS*, 517, 6035
- Torrey, P., Vogelsberger, M., Marinacci, F., et al. 2019, *MNRAS*, 484, 5587
- Tremonti, C. A., Heckman, T. M., Kauffmann, G., et al. 2004, *ApJ*, 613, 898
- Viaene, S., Fritz, J., Baes, M., et al. 2014, *A&A*, 567, A71
- Vilone, G. & Longo, L. 2021, *Machine Learning and Knowledge Extraction*, 3, 615
- Vogelsberger, M., Genel, S., Springel, V., et al. 2014, *MNRAS*, 444, 1518
- Walcher, J., Groves, B., Budavári, T., & Dale, D. 2011, *Ap&SS*, 331, 1
- Watkins, A. E., Salo, H., Laurikainen, E., et al. 2022, *A&A*, 660, A69
- Webb, K., Balogh, M. L., Leja, J., et al. 2020, *MNRAS*, 498, 5317
- Weinberger, R., Springel, V., Hernquist, L., et al. 2016, *MNRAS*, 465, 3291
- Whitaker, K. E., van Dokkum, P. G., Brammer, G., & Franx, M. 2012, *ApJ*, 754, L29
- Wright, S. 1921, *Journal of Agricultural Research*, 20, 557
- Zahid, H. J., Kudritzki, R.-P., Conroy, C., Andrews, B., & Ho, I. T. 2017, *ApJ*, 847, 18
- Zanisi, L., Huertas-Company, M., Lanusse, F., et al. 2021, *MNRAS*, 501, 4359
- Zeiler, M., Ranzato, M., Monga, R., et al. 2013, in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, 3517–3521
- Zibetti, S., Charlot, S., & Rix, H.-W. 2009, *MNRAS*, 400, 1181
- Zibetti, S. & Gallazzi, A. R. 2022, *MNRAS*, 512, 1415
- Zibetti, S., Gallazzi, A. R., Hirschmann, M., et al. 2020, *MNRAS*, 491, 3562

¹ Sterrenkundig Observatorium, Universiteit Gent, Krijgslaan 281 S9, 9000 Gent, Belgium

² STAR Institute, Quartier Agora - Allée du six Août, 19c B-4000 Liège, Belgium

³ INAF-Osservatorio Astronomico di Roma, Via Frascati 33, 00078 Monteporzio Catone, Italy

⁴ Johns Hopkins University 3400 North Charles Street Baltimore, MD 21218, USA

⁵ INAF, Istituto di Radioastronomia, Via Piero Gobetti 101, 40129 Bologna, Italy

⁶ Dipartimento di Fisica e Astronomia "G. Galilei", Università di Padova, Via Marzolo 8, 35131 Padova, Italy

⁷ INAF-Osservatorio di Astrofisica e Scienza dello Spazio di Bologna, Via Piero Gobetti 93/3, 40129 Bologna, Italy

⁸ INAF-Osservatorio Astronomico di Capodimonte, Via Moirariello 16, 80131 Napoli, Italy

⁹ INFN section of Naples, Via Cinthia 6, 80126, Napoli, Italy

¹⁰ Jodrell Bank Centre for Astrophysics, Department of Physics and Astronomy, University of Manchester, Oxford Road, Manchester M13 9PL, UK

¹¹ Dipartimento di Fisica e Astronomia, Università di Bologna, Via Gobetti 93/2, 40129 Bologna, Italy

¹² INAF-Osservatorio Astrofisico di Arcetri, Largo E. Fermi 5, 50125, Firenze, Italy

¹³ Instituto de Astrofísica de Canarias, Calle Vía Láctea s/n, 38204, San Cristóbal de La Laguna, Tenerife, Spain

- 14 Departamento de Astrofísica, Universidad de La Laguna, 38206, La Laguna, Tenerife, Spain
- 15 Universität Innsbruck, Institut für Astro- und Teilchenphysik, Technikerstr. 25/8, 6020 Innsbruck, Austria
- 16 Institute of Physics, Laboratory of Astrophysics, Ecole Polytechnique Fédérale de Lausanne (EPFL), Observatoire de Sauverny, 1290 Versoix, Switzerland
- 17 Kapteyn Astronomical Institute, University of Groningen, PO Box 800, 9700 AV Groningen, The Netherlands
- 18 SISSA, International School for Advanced Studies, Via Bonomea 265, 34136 Trieste TS, Italy
- 19 INFN, Sezione di Trieste, Via Valerio 2, 34127 Trieste TS, Italy
- 20 Université de Strasbourg, CNRS, Observatoire astronomique de Strasbourg, UMR 7550, 67000 Strasbourg, France
- 21 INAF-IASF Milano, Via Alfonso Corti 12, 20133 Milano, Italy
- 22 Instituto de Astrofísica de Canarias (IAC); Departamento de Astrofísica, Universidad de La Laguna (ULL), 38200, La Laguna, Tenerife, Spain
- 23 Institute of Space Sciences (ICE, CSIC), Campus UAB, Carrer de Can Magrans, s/n, 08193 Barcelona, Spain
- 24 School of Mathematics and Physics, University of Surrey, Guildford, Surrey, GU2 7XH, UK
- 25 INAF-Osservatorio Astronomico di Brera, Via Brera 28, 20122 Milano, Italy
- 26 IFPU, Institute for Fundamental Physics of the Universe, via Beirut 2, 34151 Trieste, Italy
- 27 INAF-Osservatorio Astronomico di Trieste, Via G. B. Tiepolo 11, 34143 Trieste, Italy
- 28 INFN-Sezione di Bologna, Viale Berti Pichat 6/2, 40127 Bologna, Italy
- 29 Max Planck Institute for Extraterrestrial Physics, Giessenbachstr. 1, 85748 Garching, Germany
- 30 Universitäts-Sternwarte München, Fakultät für Physik, Ludwig-Maximilians-Universität München, Scheinerstrasse 1, 81679 München, Germany
- 31 INAF-Osservatorio Astrofisico di Torino, Via Osservatorio 20, 10025 Pino Torinese (TO), Italy
- 32 Aix-Marseille Université, CNRS, CNES, LAM, Marseille, France
- 33 Dipartimento di Fisica, Università di Genova, Via Dodecaneso 33, 16146, Genova, Italy
- 34 INFN-Sezione di Genova, Via Dodecaneso 33, 16146, Genova, Italy
- 35 Department of Physics "E. Pancini", University Federico II, Via Cinthia 6, 80126, Napoli, Italy
- 36 Instituto de Astrofísica e Ciências do Espaço, Universidade do Porto, CAUP, Rua das Estrelas, PT4150-762 Porto, Portugal
- 37 Faculdade de Ciências da Universidade do Porto, Rua do Campo de Alegre, 4150-007 Porto, Portugal
- 38 Dipartimento di Fisica, Università degli Studi di Torino, Via P. Giuria 1, 10125 Torino, Italy
- 39 INFN-Sezione di Torino, Via P. Giuria 1, 10125 Torino, Italy
- 40 Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT), Avenida Complutense 40, 28040 Madrid, Spain
- 41 Port d'Informació Científica, Campus UAB, C. Albareda s/n, 08193 Bellaterra (Barcelona), Spain
- 42 Institute for Theoretical Particle Physics and Cosmology (TTK), RWTH Aachen University, 52056 Aachen, Germany
- 43 Institut d'Estudis Espacials de Catalunya (IEEC), Edifici RDIT, Campus UPC, 08860 Castelldefels, Barcelona, Spain
- 44 Dipartimento di Fisica e Astronomia "Augusto Righi" - Alma Mater Studiorum Università di Bologna, Viale Berti Pichat 6/2, 40127 Bologna, Italy
- 45 Institute for Astronomy, University of Edinburgh, Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ, UK
- 46 European Space Agency/ESRIN, Largo Galileo Galilei 1, 00044 Frascati, Roma, Italy
- 47 ESAC/ESA, Camino Bajo del Castillo, s/n., Urb. Villafranca del Castillo, 28692 Villanueva de la Cañada, Madrid, Spain
- 48 Université Claude Bernard Lyon 1, CNRS/IN2P3, IP2I Lyon, UMR 5822, Villeurbanne, F-69100, France
- 49 Institut de Ciències del Cosmos (ICCUB), Universitat de Barcelona (IEEC-UB), Martí i Franquès 1, 08028 Barcelona, Spain
- 50 Institució Catalana de Recerca i Estudis Avançats (ICREA), Passeig de Lluís Companys 23, 08010 Barcelona, Spain
- 51 UCB Lyon 1, CNRS/IN2P3, IUF, IP2I Lyon, 4 rue Enrico Fermi, 69622 Villeurbanne, France
- 52 Departamento de Física, Faculdade de Ciências, Universidade de Lisboa, Edifício C8, Campo Grande, PT1749-016 Lisboa, Portugal
- 53 Instituto de Astrofísica e Ciências do Espaço, Faculdade de Ciências, Universidade de Lisboa, Campo Grande, 1749-016 Lisboa, Portugal
- 54 Department of Astronomy, University of Geneva, ch. d'Ecogia 16, 1290 Versoix, Switzerland
- 55 INAF-Istituto di Astrofisica e Planetologia Spaziali, via del Fosso del Cavaliere, 100, 00100 Roma, Italy
- 56 Université Paris-Saclay, CNRS, Institut d'astrophysique spatiale, 91405, Orsay, France
- 57 INFN-Padova, Via Marzolo 8, 35131 Padova, Italy
- 58 Université Paris-Saclay, Université Paris Cité, CEA, CNRS, AIM, 91191, Gif-sur-Yvette, France
- 59 Space Science Data Center, Italian Space Agency, via del Politecnico snc, 00133 Roma, Italy
- 60 Istituto Nazionale di Fisica Nucleare, Sezione di Bologna, Via Irnerio 46, 40126 Bologna, Italy
- 61 INAF-Osservatorio Astronomico di Padova, Via dell'Osservatorio 5, 35122 Padova, Italy
- 62 Dipartimento di Fisica "Aldo Pontremoli", Università degli Studi di Milano, Via Celoria 16, 20133 Milano, Italy
- 63 Institute of Theoretical Astrophysics, University of Oslo, P.O. Box 1029 Blindern, 0315 Oslo, Norway
- 64 Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA, 91109, USA
- 65 Department of Physics, Lancaster University, Lancaster, LA1 4YB, UK
- 66 Felix Hormuth Engineering, Goethestr. 17, 69181 Leimen, Germany
- 67 Technical University of Denmark, Elektrovej 327, 2800 Kgs. Lyngby, Denmark
- 68 Cosmic Dawn Center (DAWN), Denmark
- 69 Max-Planck-Institut für Astronomie, Königstuhl 17, 69117 Heidelberg, Germany
- 70 NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA
- 71 Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, UK
- 72 Department of Physics and Helsinki Institute of Physics, Gustaf Hållströmin katu 2, 00014 University of Helsinki, Finland
- 73 Aix-Marseille Université, CNRS/IN2P3, CPPM, Marseille, France
- 74 Leiden Observatory, Leiden University, Einsteinweg 55, 2333 CC Leiden, The Netherlands
- 75 Université de Genève, Département de Physique Théorique and Centre for Astroparticle Physics, 24 quai Ernest-Ansermet, CH-1211 Genève 4, Switzerland
- 76 Department of Physics, P.O. Box 64, 00014 University of Helsinki, Finland
- 77 Helsinki Institute of Physics, Gustaf Hållströmin katu 2, University of Helsinki, Helsinki, Finland
- 78 NOVA optical infrared instrumentation group at ASTRON, Oude Hoogeveensedijk 4, 7991PD, Dwingeloo, The Netherlands
- 79 INFN-Sezione di Milano, Via Celoria 16, 20133 Milano, Italy
- 80 University of Applied Sciences and Arts of Northwestern Switzerland, School of Engineering, 5210 Windisch, Switzerland
- 81 Universität Bonn, Argelander-Institut für Astronomie, Auf dem Hügel 71, 53121 Bonn, Germany
- 82 INFN-Sezione di Roma, Piazzale Aldo Moro, 2 - c/o Dipartimento di Fisica, Edificio G. Marconi, 00185 Roma, Italy
- 83 Dipartimento di Fisica e Astronomia "Augusto Righi" - Alma Mater Studiorum Università di Bologna, via Piero Gobetti 93/2, 40129 Bologna, Italy

- ⁸⁴ Department of Physics, Centre for Extragalactic Astronomy, Durham University, South Road, Durham, DH1 3LE, UK
- ⁸⁵ Université Paris Cité, CNRS, Astroparticule et Cosmologie, 75013 Paris, France
- ⁸⁶ Institut d'Astrophysique de Paris, 98bis Boulevard Arago, 75014, Paris, France
- ⁸⁷ Institut d'Astrophysique de Paris, UMR 7095, CNRS, and Sorbonne Université, 98 bis boulevard Arago, 75014 Paris, France
- ⁸⁸ European Space Agency/ESTEC, Keplerlaan 1, 2201 AZ Noordwijk, The Netherlands
- ⁸⁹ School of Mathematics, Statistics and Physics, Newcastle University, Herschel Building, Newcastle-upon-Tyne, NE1 7RU, UK
- ⁹⁰ Institut de Física d'Altes Energies (IFAE), The Barcelona Institute of Science and Technology, Campus UAB, 08193 Bellaterra (Barcelona), Spain
- ⁹¹ DARK, Niels Bohr Institute, University of Copenhagen, Jagtvej 155, 2200 Copenhagen, Denmark
- ⁹² Centre National d'Etudes Spatiales – Centre spatial de Toulouse, 18 avenue Edouard Belin, 31401 Toulouse Cedex 9, France
- ⁹³ Institute of Space Science, Str. Atomistilor, nr. 409 Măgurele, Ilfov, 077125, Romania
- ⁹⁴ Consejo Superior de Investigaciones Científicas, Calle Serrano 117, 28006 Madrid, Spain
- ⁹⁵ Institut für Theoretische Physik, University of Heidelberg, Philosophenweg 16, 69120 Heidelberg, Germany
- ⁹⁶ Institut de Recherche en Astrophysique et Planétologie (IRAP), Université de Toulouse, CNRS, UPS, CNES, 14 Av. Edouard Belin, 31400 Toulouse, France
- ⁹⁷ Université St Joseph; Faculty of Sciences, Beirut, Lebanon
- ⁹⁸ Departamento de Física, FCFM, Universidad de Chile, Blanco Encalada 2008, Santiago, Chile
- ⁹⁹ Satlantis, University Science Park, Sede Bld 48940, Leioa-Bilbao, Spain
- ¹⁰⁰ Infrared Processing and Analysis Center, California Institute of Technology, Pasadena, CA 91125, USA
- ¹⁰¹ Instituto de Astrofísica e Ciências do Espaço, Faculdade de Ciências, Universidade de Lisboa, Tapada da Ajuda, 1349-018 Lisboa, Portugal
- ¹⁰² Universidad Politécnica de Cartagena, Departamento de Electrónica y Tecnología de Computadoras, Plaza del Hospital 1, 30202 Cartagena, Spain
- ¹⁰³ INFN-Bologna, Via Ippolito Nievo 46, 40126 Bologna, Italy
- ¹⁰⁴ Dipartimento di Fisica, Università degli studi di Genova, and INFN-Sezione di Genova, via Dodecaneso 33, 16146, Genova, Italy
- ¹⁰⁵ Astronomical Observatory of the Autonomous Region of the Aosta Valley (OAVdA), Loc. Lignan 39, I-11020, Nus (Aosta Valley), Italy
- ¹⁰⁶ ICL, Junia, Université Catholique de Lille, LITL, 59000 Lille, France
- ¹⁰⁷ ICSC - Centro Nazionale di Ricerca in High Performance Computing, Big Data e Quantum Computing, Via Magnanelli 2, Bologna, Italy
- ¹⁰⁸ Instituto de Física Teórica UAM-CSIC, Campus de Cantoblanco, 28049 Madrid, Spain
- ¹⁰⁹ CERCA/ISO, Department of Physics, Case Western Reserve University, 10900 Euclid Avenue, Cleveland, OH 44106, USA
- ¹¹⁰ Dipartimento di Fisica e Scienze della Terra, Università degli Studi di Ferrara, Via Giuseppe Saragat 1, 44122 Ferrara, Italy
- ¹¹¹ Laboratoire Univers et Théorie, Observatoire de Paris, Université PSL, Université Paris Cité, CNRS, 92190 Meudon, France
- ¹¹² Departamento de Física Fundamental. Universidad de Salamanca. Plaza de la Merced s/n. 37008 Salamanca, Spain
- ¹¹³ Istituto Nazionale di Fisica Nucleare, Sezione di Ferrara, Via Giuseppe Saragat 1, 44122 Ferrara, Italy
- ¹¹⁴ Center for Data-Driven Discovery, Kavli IPMU (WPI), UTIAS, The University of Tokyo, Kashiwa, Chiba 277-8583, Japan
- ¹¹⁵ Dipartimento di Fisica - Sezione di Astronomia, Università di Trieste, Via Tiepolo 11, 34131 Trieste, Italy
- ¹¹⁶ Minnesota Institute for Astrophysics, University of Minnesota, 116 Church St SE, Minneapolis, MN 55455, USA
- ¹¹⁷ Institute Lorentz, Leiden University, Niels Bohrweg 2, 2333 CA Leiden, The Netherlands
- ¹¹⁸ Université Côte d'Azur, Observatoire de la Côte d'Azur, CNRS, Laboratoire Lagrange, Bd de l'Observatoire, CS 34229, 06304 Nice cedex 4, France
- ¹¹⁹ Institute for Astronomy, University of Hawaii, 2680 Woodlawn Drive, Honolulu, HI 96822, USA
- ¹²⁰ Department of Physics & Astronomy, University of California Irvine, Irvine CA 92697, USA
- ¹²¹ Department of Astronomy & Physics and Institute for Computational Astrophysics, Saint Mary's University, 923 Robie Street, Halifax, Nova Scotia, B3H 3C3, Canada
- ¹²² Departamento Física Aplicada, Universidad Politécnica de Cartagena, Campus Muralla del Mar, 30202 Cartagena, Murcia, Spain
- ¹²³ Department of Physics, Oxford University, Keble Road, Oxford OX1 3RH, UK
- ¹²⁴ CEA Saclay, DFR/IRFU, Service d'Astrophysique, Bat. 709, 91191 Gif-sur-Yvette, France
- ¹²⁵ Institute of Cosmology and Gravitation, University of Portsmouth, Portsmouth PO1 3FX, UK
- ¹²⁶ Department of Computer Science, Aalto University, PO Box 15400, Espoo, FI-00076, Finland
- ¹²⁷ Instituto de Astrofísica de Canarias, c/ Via Lactea s/n, La Laguna E-38200, Spain. Departamento de Astrofísica de la Universidad de La Laguna, Avda. Francisco Sanchez, La Laguna, E-38200, Spain
- ¹²⁸ Caltech/IPAC, 1200 E. California Blvd., Pasadena, CA 91125, USA
- ¹²⁹ Ruhr University Bochum, Faculty of Physics and Astronomy, Astronomical Institute (AIRUB), German Centre for Cosmological Lensing (GCCL), 44780 Bochum, Germany
- ¹³⁰ Univ. Grenoble Alpes, CNRS, Grenoble INP, LPSC-IN2P3, 53, Avenue des Martyrs, 38000, Grenoble, France
- ¹³¹ Department of Physics and Astronomy, Vesilinnantie 5, 20014 University of Turku, Finland
- ¹³² Serco for European Space Agency (ESA), Camino bajo del Castillo, s/n, Urbanización Villafranca del Castillo, Villanueva de la Cañada, 28692 Madrid, Spain
- ¹³³ ARC Centre of Excellence for Dark Matter Particle Physics, Melbourne, Australia
- ¹³⁴ Centre for Astrophysics & Supercomputing, Swinburne University of Technology, Hawthorn, Victoria 3122, Australia
- ¹³⁵ School of Physics and Astronomy, Queen Mary University of London, Mile End Road, London E1 4NS, UK
- ¹³⁶ Department of Physics and Astronomy, University of the Western Cape, Bellville, Cape Town, 7535, South Africa
- ¹³⁷ IRFU, CEA, Université Paris-Saclay 91191 Gif-sur-Yvette Cedex, France
- ¹³⁸ Oskar Klein Centre for Cosmoparticle Physics, Department of Physics, Stockholm University, Stockholm, SE-106 91, Sweden
- ¹³⁹ Astrophysics Group, Blackett Laboratory, Imperial College London, London SW7 2AZ, UK
- ¹⁴⁰ Dipartimento di Fisica, Sapienza Università di Roma, Piazzale Aldo Moro 2, 00185 Roma, Italy
- ¹⁴¹ Aurora Technology for European Space Agency (ESA), Camino bajo del Castillo, s/n, Urbanización Villafranca del Castillo, Villanueva de la Cañada, 28692 Madrid, Spain
- ¹⁴² Centro de Astrofísica da Universidade do Porto, Rua das Estrelas, 4150-762 Porto, Portugal
- ¹⁴³ HE Space for European Space Agency (ESA), Camino bajo del Castillo, s/n, Urbanización Villafranca del Castillo, Villanueva de la Cañada, 28692 Madrid, Spain
- ¹⁴⁴ Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK
- ¹⁴⁵ Department of Astrophysics, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland
- ¹⁴⁶ INAF-Osservatorio Astronomico di Brera, Via Brera 28, 20122 Milano, Italy, and INFN-Sezione di Genova, Via Dodecaneso 33, 16146, Genova, Italy

- ¹⁴⁷ Theoretical astrophysics, Department of Physics and Astronomy,
Uppsala University, Box 515, 751 20 Uppsala, Sweden
- ¹⁴⁸ School of Physics & Astronomy, University of Southampton, High-
field Campus, Southampton SO17 1BJ, UK
- ¹⁴⁹ Department of Physics, Royal Holloway, University of London,
TW20 0EX, UK
- ¹⁵⁰ Mullard Space Science Laboratory, University College London,
Holmbury St Mary, Dorking, Surrey RH5 6NT, UK
- ¹⁵¹ Department of Physics and Astronomy, University of California,
Davis, CA 95616, USA
- ¹⁵² Department of Astrophysical Sciences, Peyton Hall, Princeton Uni-
versity, Princeton, NJ 08544, USA
- ¹⁵³ Cosmic Dawn Center (DAWN)
- ¹⁵⁴ Niels Bohr Institute, University of Copenhagen, Jagtvej 128, 2200
Copenhagen, Denmark
- ¹⁵⁵ Center for Cosmology and Particle Physics, Department of Physics,
New York University, New York, NY 10003, USA
- ¹⁵⁶ Center for Computational Astrophysics, Flatiron Institute, 162 5th
Avenue, 10010, New York, NY, USA

Appendix A: Effect of including additional projection angles

Here, we use a subset of 200 randomly selected galaxies to test whether the inclusion of other projections would affect the prediction accuracy of our ML algorithm. As a test set, we set aside 50 random galaxies, each of them at one random projection angle, totalling ~ 2.7 million pixels. The remaining 150 galaxies are used to train the algorithm. First, we use only one projection angle O1, then two (O1 and O2), in each step adding an additional angle projection until all five projection angles (O1–O5) are used. The accuracy scores are located in Table A.1.

Table A.1. Coefficient of determination (R^2) for a DNN using the four *Euclid* surface brightnesses on a subset of 200 randomly selected galaxies and including different projections.

Projection angles	Pixel count	R^2 (Σ_*)	R^2 (Z_*)	R^2 (t_*)
O1	8 912 000	0.9319	0.5452	0.3009
O1, O2	17 774 522	0.9322	0.5484	0.3015
O1, O2, O3	26 411 564	0.9320	0.5899	0.3116
O1, O2, O3, O4	34 809 544	0.9323	0.5490	0.3115
O1, O2, O3, O4, O5	43 214 225	0.9324	0.5529	0.3116

Appendix B: The effect of sample bias

There is a slight bias between our training and test sets, as shown in Fig. 2. For this reason, we apply our trained algorithms on both the test and the training sets. The accuracy scores between these two sets, shown in Tab. B.1, are within the 10% difference.

Table B.1. Accuracy scores (as defined in Sect. 3.3) for the three target galaxy properties (Σ_* , Z_* , and t_*) inferred using the four *Euclid* surface brightnesses for the training and test sets with the RF ML algorithm.

	Train	Test	Diff. (%)
Target	R^2		
$\Sigma_* / (M_\odot \text{ pc}^{-2})$	0.937	0.927	1
Z_*	0.606	0.592	2.4
t_* / Gyr	0.419	0.382	9.7
Target	RMSE		
$\Sigma_* / (M_\odot \text{ pc}^{-2})$	0.1219	0.1308	6.8
Z_*	0.0805	0.0765	5.2
t_* / Gyr	0.1279	0.1285	9.7
Target	ρ		
$\Sigma_* / (M_\odot \text{ pc}^{-2})$	0.968	0.963	0.5
Z_*	0.778	0.771	0.9
t_* / Gyr	0.647	0.622	4

Notes. Left to right: targets; coefficient of determination (R^2); root-mean-square error (RMSE); Pearson ρ coefficient.

Appendix C: The effect of squared colours

As this work was partly inspired by Acquaviva (2015), we also test the effect of using squared colour could offer a higher ac-

curacy when inferring galaxy parameters from mock *Euclid* observations. We used our DNN set-up, as described in Table 2 in Sect. 3.1. The results are shown in Table C.1. The differences, if any, are minor at best, once again giving credence to our conclusion that it is possible to infer galaxy parameters using only the four *Euclid* bands, as the algorithm is capable of extracting all of the information contained within them, without the need of modelling additional colours.

Table C.1. Accuracy scores (as defined in Sect. 3.3) for the three target galaxy properties (Σ_* , Z_* , and t_*) inferred using the four *Euclid* surface brightnesses, six colours modelled from them, and six squared colours using a DNN.

Target	R^2	RMSE	f_{out}	NMAD	$\langle \Delta Y \rangle$	ρ
$\Sigma_* / (M_\odot \text{ pc}^{-2})$	0.927	0.130	0.0325	0.1088	0.0020	0.963
Z_*	0.595	0.076	0.0007	0.0719	0.0024	0.773
t_* / Gyr	0.382	0.128	0.0281	0.0994	-0.0017	0.623

Notes. Left to right: Target; coefficient of determination (R^2); root-mean-square error (RMSE); fraction of catastrophic outliers (f_{out}); normalised median absolute deviation (NMAD); and bias ($\langle \Delta Y \rangle$, where Y represents the target galaxy property); Pearson ρ coefficient.

Appendix D: Inferring colours from *Euclid* surface brightnesses

Here, we test the accuracy of inferring the six colours modelled with the four *Euclid* surface brightnesses using a DNN with the same set-up as Table 2. In all cases, the network converges very quickly (in under 100 epochs), and with the accuracy shown in Table D.1, we conclude that using colours to infer galaxy parameters does not offer any additional benefit to our method.

Table D.1. Accuracy scores (as defined in Sect. 3.3) for the six *Euclid* colours inferred from the four *Euclid* brightnesses using a DNN.

Target colour	R^2	RMSE
$(I_E - Y_E)$	0.999	0.004
$(I_E - J_E)$	0.999	0.003
$(I_E - H_E)$	0.998	0.011
$(Y_E - J_E)$	0.999	0.009
$(Y_E - H_E)$	0.999	0.005
$(J_E - H_E)$	0.999	0.005

Notes. Left to right: Target colour; coefficient of determination (R^2); root-mean-square error (RMSE).