# LJMU Research Online

Camps, P, Kapoor, AU, Trčka, A, Font, AS, McCarthy, IG, Trayford, J and Baes, M

High-resolution synthetic UV-submm images for Milky Way-mass simulated galaxies from the ARTEMIS project

http://researchonline.ljmu.ac.uk/id/eprint/16547/

Article

For more information please contact researchonline@ljmu.ac.uk

# High-resolution synthetic UV-submm images for Milky Way-mass simulated galaxies from the ARTEMIS project

Peter Camps [ID],[1]* Anand Utsav Kapoor [ID],[1] Ana Trcka [ID],[1] Andreea S. Font [ID],[2] Ian G. McCarthy [ID],[2] James Trayford[3] and Maarten Baes [ID][1]

[1]*Sterrenkundig Observatorium, Universiteit Gent, Krijgslaan 281, B-9000 Gent, Belgium*
[2]*Astrophysics Research Institute, Liverpool John Moores University, 146 Brownlow Hill, Liverpool L53RF, UK*
[3]*Institute of Cosmology and Gravitation, University of Portsmouth, Dennis Sciama Building, Burnaby Road, Portsmouth PO1 3FX, UK*

## ABSTRACT

We present redshift-zero synthetic dust-aware observations for the 45 Milky Way-mass simulated galaxies of the ARTEMIS project, calculated with the SKIRT radiative transfer code. The post-processing procedure includes components for star-forming regions, stellar sources, and diffuse dust. We produce and publicly release realistic high-resolution images for 50 commonly used broad-band filters from ultraviolet to submillimetre wavelengths and for 18 different viewing angles. We compare the simulated ARTEMIS galaxies to observed galaxies in the DustPedia data base with similar stellar mass and star formation rate, and to synthetic observations of the simulated galaxies of the Auriga project produced in previous work using a similar post-processing technique. In all cases, global galaxy properties are derived using SED fitting. We find that, similar to Auriga, the post-processed ARTEMIS galaxies generally reproduce the observed scaling relations for global fluxes and physical properties, although dust extinction at FUV/UV wavelengths is underestimated and representative dust temperatures are lower than observed. At a resolved scale, we compare multiwavelength non-parametric morphological properties of selected disc galaxies across the data sets. We find that the ARTEMIS galaxies largely reproduce the observed morphological trends as a function of wavelength, although they appear to be more clumpy and less symmetrical than observed. We note that the ARTEMIS and Auriga galaxies occupy adjacent regions in the specific star formation versus stellar mass plane, so that the synthetic observation data sets supplement each other.

**Key words:** radiative transfer – methods: numerical – dust, extinction – galaxies: ISM.

## 1 INTRODUCTION

The plethora of data produced by current and planned earth-based observatories and space missions enable an exceedingly detailed study of cosmic structure, and in particular of the assembly and evolution of galaxies. One important method that helps us uncover and make sense of the physical mechanisms underlying galaxy formation is to emulate those processes in computer simulations. The most comprehensive simulations evolve dark and baryonic matter in a cosmologically relevant volume from initial conditions at high redshift to the present day (for a review, see Vogelsberger et al. 2020a), employing subgrid recipes for unresolved processes such as star formation, stellar feedback, and chemical evolution. Recent examples include EAGLE (Crain et al. 2015; Schaye et al. 2015), MassiveBlack-II (Khandai et al. 2015), Romulus25 (Tremmel et al. 2017), SIMBA (Davé et al. 2019), and Illustris-TNG50 (Nelson et al. 2019; Pillepich et al. 2019). These simulations succeed in reproducing many observed global galaxy properties to a fair degree, including for example stellar mass functions, galaxy sizes, mass–metallicity relations, star formation relations, passive fractions, and gas contents (see the references listed above for each simulation project).

The resolution of cosmological simulations is necessarily limited by the available computational resources. Smaller simulation volumes allow a somewhat better resolution but reproduce fewer massive structures and rare objects. Using an alternate approach, cosmological zoom simulations focus on a limited portion of a larger simulation volume, for example the contents of a single dark matter halo, to achieve baryon mass resolutions down to around $10^4$ M$_\odot$. Recent examples include NIHAO (Wang et al. 2015), APOSTLE (Sawala et al. 2016), Latte (Wetzel et al. 2016), Auriga (Grand et al. 2017), RomulusC (Tremmel et al. 2019), and ARTEMIS (Font et al. 2020; Font, McCarthy & Belokurov 2021). Each resolution element now has a mass at the upper end of the observed molecular cloud mass range, implying that further resolution improvements will likely need to be accompanied by enhanced subgrid recipes to better capture the physical processes on these smaller scales.

To increase the accuracy of future simulation efforts, and thus improve our understanding of the emulated physics, a detailed comparison of simulation results to observations is required. Comparing simulation results to observations, however, is often tricky. Simulations yield intrinsic galaxy properties such as mass, age, metallicity, or star formation rate (SFR) by aggregating the corresponding properties of the particles or cells used to represent physical constituents. Observations, on the other hand, yield multiwavelength fluxes and spectra which are integrated along the line of sight and thus provide a two-dimensional projection of the galaxy under study.

* E-mail: Peter.Camps@ugent.be

The observed radiation is often significantly altered by the effects of dust grains in the interstellar medium (ISM; e.g. Viaene et al. 2016) and depends non-linearly on the complex geometry of the galaxy (e.g. Saftly et al. 2015). As a result, deriving intrinsic properties from the observed data always involves some form of conversion that relies on approximating assumptions (e.g. Kennicutt & Evans 2012; Courteau et al. 2014).

Alternatively, one can bring the simulation output into the observational realm through forward modelling. In addition to assigning appropriate emission spectra, this requires simulating transport of the radiation through the ISM, including the scattering, absorption and re-emission by dust grains. This can be accomplished using radiative transfer (RT) codes such as GRASIL3D (Domínguez-Tenreiro et al. 2014), HYPERION/POWDERDAY (Robitaille 2011; Narayanan et al. 2021), or SKIRT (Baes et al. 2011; Camps & Baes 2015, 2020). While this approach obviously also relies on approximations, it offers important benefits. It allows incorporation of a wide range of physics into the model, including for example the detailed distribution of stars and dust in the simulated galaxy. The synthetic observables resulting from the model can be directly compared to observed data and can be processed or visualized using any of the tools commonly used to interpret observations. This includes deriving estimates for the physical properties from the synthetic observations, which can help evaluate the employed recipes by comparison to the known intrinsic properties of the simulated galaxies.

In the past decade, many authors have taken this route to generate synthetic observables for cosmological (zoom) simulations (e.g. Jonsson, Groves & Cox 2010; Lanz et al. 2014; Granato et al. 2015; Bignone, Pellizza & Tissera 2016; Camps et al. 2016; Santos-Santos et al. 2017; Trayford et al. 2017; Barber, Crain & Schaye 2018; Gjergo et al. 2018; Lahén et al. 2018; Narayanan et al. 2018; Cochrane et al. 2019; Liang et al. 2019; Ma et al. 2019; Rodriguez-Gomez et al. 2019; Vogelsberger et al. 2020b; Granato et al. 2021; Lovell et al. 2021; Parsotan et al. 2021). Some also prepare data sets for public use. For example, Camps et al. (2018) publish spatially integrated UV to submillimetre (submm) broad-band fluxes for nearly half a million of EAGLE galaxies up to redshift 6. More recently, Trčka et al. (2021) offer a similar data set for the Illustris-TNG50 galaxies, and Kapoor et al. (2021) provide high-resolution broad-band images for 30 present-day Auriga zoom galaxies.

In this work, we consider the recent ARTEMIS project (Font et al. 2020, 2021), encompassing 45 zoom simulations of Milky Way-mass dark matter mass haloes performed using the EAGLE cosmological simulation code (Crain et al. 2015; Schaye et al. 2015). We use SKIRT version 9 (Camps & Baes 2020) to produce synthetic observations at redshift zero for the main galaxy in each of the haloes, with a stellar mass ranging from 1 to $9 \times 10^{10}$ M$_\odot$. We calibrate our RT post-processing scheme by comparing to observed galaxies with similar stellar mass and SFR from the DustPedia data set (Clark et al. 2018), a large sample of nearby galaxies with matched aperture photometry in more than 40 bands from UV to millimetre wavelengths. For each ARTEMIS galaxy, we publish highly resolved images ($50 \times 50$ pc pixels) observed for 18 sightlines through 50 commonly used broad-band filters spanning ultraviolet (UV) to submm wavelengths.

Our set-up and procedures are nearly identical to those employed by Kapoor et al. (2021) for producing synthetic observations of the galaxies in the Auriga project (Grand et al. 2017). The latter project comprises 30 zoom simulations of isolated Milky Way-mass dark matter haloes, selected from a dark-matter-only simulation and evolved to redshift zero in a full cosmological context including baryon physics. The main galaxy in each halo has a stellar mass in a range of 3 to $11 \times 10^{10}$ M$_\odot$. Although the Auriga and ARTEMIS mass ranges are similar, the Auriga galaxies are more massive on

average and nearly all of them are spiral galaxies because of the employed selection criteria. The ARTEMIS galaxies show a much more diverse morphology and occupy a different region in the specific star formation rate (sSFR) versus stellar mass plane.

As a result, the data set for ARTEMIS prepared in this work complements and augments the data set prepared by Kapoor et al. (2021) for Auriga in several ways. The number of simulated galaxies for which high-resolution and multiwavelength synthetic observables are made available is more than doubled, from 30 to 75. The stellar mass range is extended downwards and more diverse morphology types are included, so that the combined data set allows studying simulated galaxy properties, scaling relations and dust heating on resolved scales for a wide range of Milky Way-mass galaxies. It thus becomes possible to compare spatially resolved properties of the ARTEMIS and the Auriga galaxies with observations and among each other, leading to a better understanding of how well each simulation reproduces reality and why.

In Section 2, we briefly describe the cosmological simulations, the observed data, and the software codes used in this work. In Section 3, we review our RT post-processing procedure, discuss the calibration of the associated parameters, and list the data products being made available as a result. In Section 4, we then offer an initial analysis of these synthetic observables, including global physical properties derived by SED fitting using CIGALE (Boquien et al. 2019) and non-parametric morphology properties calculated from the spatially resolved images at several wavelengths using STATMORPH (Rodriguez-Gomez et al. 2019). We compare these properties to those of similar observed DustPedia galaxies and simulated Auriga galaxies, and discuss the implications. In Section 5, we summarize and conclude.

## 2 BACKGROUND

### 2.1 ARTEMIS

The ARTEMIS project includes a set of 45 zoomed-in, high-resolution hydrodynamical simulations of galaxies residing in haloes of Milky Way mass, 42 of them presented by Font et al. (2020) and 3 more by Font et al. (2021). The baryon mass resolution is about $3 \times 10^4$ M$_\odot$. The simulations are performed with the EAGLE galaxy formation code (Crain et al. 2015; Schaye et al. 2015) using the same solvers and subgrid physics except for a re-calibrated stellar feedback recipe. The simulation set-up is fully described by Font et al. (2020) and references therein; we provide just a brief summary here.

The MUSIC code (Hahn & Abel 2011) is used to generate initial conditions at redshift 127 for a flat $\Lambda$CDM WMAP9 cosmology (Hinshaw et al. 2013) in a periodic box 25 Mpc on a side. This volume is then evolved to redshift zero using collisionless dynamics. From the completed simulation a volume-limited sample is selected of all 63 haloes within a mass range of $8 \times 10^{11} < M_{200}/$M$_\odot < 2 \times 10^{12}$, where $M_{200}$ is the mass enclosing a mean density of 200 times the critical density at redshift zero. The selection is based solely on halo mass with no conditions on the merger history or environment.

For 45 of the selected haloes, hydrodynamic zoom simulations are performed using full baryonic physics at high resolution within a region enclosing twice the halo radius, and using dark-matter-only dynamics at lower resolution in the remainder of the volume. The EAGLE code employed to run the zoom simulations is a modified version of the $N$-body smoothed particle hydrodynamics (SPH) code GADGET-3 (Springel 2005). It provides subgrid models of important processes that cannot be resolved directly in the simulations, including metal-dependent radiative cooling, star formation, stellar evolution, and chemodynamics, black hole formation and growth,

stellar feedback, and active galactic nucleus (AGN) feedback. Star formation assumes the Chabrier (2003) initial mass function (IMF).

The efficiency of the stellar feedback in the main EAGLE runs presented in Schaye et al. (2015) was fine-tuned to approximately reproduce the local galaxy stellar mass function and the size–stellar mass relation of disc galaxies. However, with increased numerical resolution, the efficiency of stellar feedback needs to be re-adjusted to recover the good match to the calibration observables. Because the resolution of the ARTEMIS simulations is about 7 times better than the finest mass resolution in the main EAGLE runs, stellar feedback efficiency was recalibrated by increasing the value of the density where the efficiency of stellar feedback transitions to its maximal value. In the EAGLE stellar feedback model, the fraction of available stellar energy used for feedback is modelled with a sigmoid function of density (and metallicity). This function asymptotes to fixed values at low and high densities, such that a higher fraction of the available energy is used at high densities in order to offset spurious (numerical) radiative cooling losses. As we increase the resolution of the simulations, the density scale at which numerical losses become important increases, motivating an increase in the transition density scale used for stellar feedback in ARTEMIS. The transition density scale was adjusted by hand so that the simulations reproduce the amplitude of the stellar mass–halo mass relation at a halo mass scale of about $10^{12}$ $M_\odot$ (see fig. 2 of Font et al. 2020). In addition, the observed sizes and star formation rates of these systems were also reproduced, without any explicit calibration to match those quantities (see the same figure).

In this work, we consider the central galaxy (i.e. the most massive object) in the redshift-zero snapshot for each of the 45 ARTEMIS haloes, excluding any satellites, or other secondary objects. Where applicable, we indicate particular galaxies using the same identifiers as introduced by Font et al. (2020, G1–G42) and Font et al. (2021, G43–G45).

## 2.2 Auriga

The Auriga project (Grand et al. 2017) includes a set of cosmological magneto-hydrodynamical zoom simulations of the formation of galaxies in isolated Milky Way mass dark haloes. The baryon mass resolution for the 30 simulations at the standard (level 4) resolution considered here is about $5 \times 10^4$ $M_\odot$. The simulation set-up is fully described by Grand et al. (2017) and references therein; we provide just a brief summary here.

The starting point for the zoom simulations is a dark-matter-only counterpart to the 100 Mpc-box Eagle simulation (L100N1504) introduced in Schaye et al. (2015) and adopting ΛCDM cosmological parameters taken from Planck Collaboration et al. (2014). This parent simulation is evolved from redshift 127 to the present day. The linear phases for the parent simulation, and for all of the zoom simulations, are taken from the public Gaussian white noise field realization, PANPHASIA (Jenkins 2013).

Host haloes are selected from the parent simulation through a mass cut criterion of $1 \times 10^{12} < M_{200}/M_\odot < 2 \times 10^{12}$ and the requirement that each candidate halo be relatively isolated at redshift zero. The degree of isolation is estimated, roughly speaking, by the distance to other haloes in the simulation relative to the virial radius of each candidate halo. From the total of 697 haloes in the chosen mass range, 174 are in the most isolated quartile, and 30 of those are randomly selected for re-simulation.

The zoom simulations are performed with the *N*-body, magnetohydrodynamics (MHD) moving mesh code AREPO (Springel 2010), equipped with a comprehensive physics model containing subgrid recipes for processes that cannot be resolved. These recipes are similar to those employed in ARTEMIS, also assuming the Chabrier (2003) IMF, but differ in many details. We summarize the more relevant ones. (1) The cold gas is not modelled in either simulation. To prevent spurious fragmentation, ARTEMIS imposes a temperature floor corresponding to the equation of state $P \propto \rho^{4/3}$. Auriga implements the two phase model introduced by Hernquist & Springel (2003). (2) ARTEMIS uses a metallicity-dependent star formation threshold, while Auriga employs a fixed threshold. (3) ARTEMIS implements stochastic thermal feedback from core-collapse supernovae; the feedback efficiency is mediated using metallicity and density-dependent factors. Auriga implements core-collapse supernovae feedback by launching wind particles that travel away from the originating site with a given velocity. The energy is deposited once certain criteria are met (Marinacci et al. 2015). (4) ARTEMIS provides a single mode of AGN feedback with a fixed efficiency. The energy is injected thermally at the location of the black hole at a rate that is proportional to the gas accretion rate. This is similar to Auriga's quasar mode. Auriga includes separate quasar and radio modes. For the quasar mode, the thermal energy is injected isotropically into neighbouring gas cells. For the radio mode, bubbles of gas are gently heated at randomly placed at locations following an inverse square distance profile around the black hole.

In addition, the Auriga simulations include prescriptions for magnetic field evolution. In the halo investigated by van de Voort et al. (2021), the central galaxy is more disc-dominated and the central black hole is more massive when magnetic fields are included. Also, the physical properties of the circumgalactic medium (CGM) change significantly. On the other hand, the global galaxy properties including stellar mass and SFR remain essentially unaffected.

In this work, we indirectly use the Auriga simulation results through the synthetic observations prepared by Kapoor et al. (2021) for the main galaxy in each of the 30 haloes.

## 2.3 DustPedia

The DustPedia project (Davies et al. 2017) combines observations from the *Herschel* and *Planck* missions and several other sources to study dust and dust-related processes in local galaxies. One outcome of the project is a public data set providing matched aperture photometry in more than 40 bands from UV to millimetre wavelengths for a sample of 875 nearby galaxies at distances up to ≈40 Mpc (Clark et al. 2018).

Casasola et al. (2020) study ISM scaling relations for the galaxies in the DustPedia data set. They report that the selection and uniform treatment of the DustPedia data leads to a complete and homogeneous galaxy sample covering a broad dynamic range of various physical properties, including stellar mass, SFR, and morphological stage. This makes the DustPedia data set ideally suited to put constraints on cosmological simulations predicting ISM properties and scaling relations. For example, Trčka et al. (2020) compare the global properties and scaling relations of galaxies produced by the EAGLE simulations to those observed for the DustPedia galaxies.

Kapoor et al. (2021) use the DustPedia data set to calibrate their RT post-processing recipe for the simulated Auriga galaxies, and subsequently compare selected scaling relations and morphological properties between simulations and observations. In this work, we follow in their path for the RT post-processing of the ARTEMIS galaxies.

## 2.4 SKIRT

The SKIRT code[1] (Baes et al. 2011; Camps & Baes 2015, 2020) is a fully three-dimensional Monte Carlo dust RT code equipped with a library of flexible input models (Baes & Camps 2015), routines to import the output from various kinds of hydrodynamical simulations (Camps & Baes 2015), a module handling stochastic heating and emission of dust grains (Camps et al. 2015), and a hybrid parallelization strategy (Verstocken et al. 2017; Camps & Baes 2020). A range of advanced spatial grids for discretizing the medium is implemented in SKIRT, including methods to efficiently traverse photons through these grids (Camps, Baes & Saftly 2013; Saftly et al. 2013; Saftly, Baes & Camps 2014).

SKIRT has been extensively used to generate synthetic UV to submm broad-band images, spectral energy distributions and polarization maps for idealized galaxies (e.g. Baes et al. 2003; Gadotti, Baes & Falony 2010; De Geyter et al. 2014; Lee et al. 2016; Peest et al. 2017), for high-resolution 3D galaxy models (e.g. De Looze et al. 2014; Verstocken et al. 2020; Nersesian et al. 2020a,b; Viaene et al. 2020), and for galaxies extracted from cosmological simulations (e.g. Saftly et al. 2015; Camps et al. 2016, 2018; Trayford et al. 2017; Barber et al. 2018; Behrens et al. 2018; Liang et al. 2018; Lahén et al. 2018; Rodriguez-Gomez et al. 2019; Ma et al. 2019; Vogelsberger et al. 2020b; Granato et al. 2021; Parsotan et al. 2021; Kapoor et al. 2021).

In this work, we use SKIRT version 9[2] to produce synthetic observations for the ARTEMIS galaxies, after extracting the relevant information from the simulation snapshots through straightforward PYTHON procedures. The full procedure and configuration details are discussed in Section 3.

## 2.5 CIGALE

The CIGALE SED-fitting code (Noll et al. 2009; Boquien et al. 2019) incorporates stellar, nebular, dust emission, and dust attenuation. It contains an implementation of a delayed and truncated star formation history (SFH; Ciesla et al. 2016), Bruzual & Charlot (2003) simple stellar population (SSP) libraries, the modified (Calzetti et al. 2000) attenuation law, and several dust models.

In this work, we use CIGALE version 0.12.1 to estimate global physical properties such as stellar mass, dust mass and SFR from the available broad-band fluxes for the various data sets under study, i.e. the ARTEMIS, Auriga and DustPedia galaxies. This allows us to compare the properties of simulated and observed galaxies on equal footing. We therefore use the same parameter settings in all cases.

Specifically, following Kapoor et al. (2021), we employ the settings used by Bianchi et al. (2018), Nersesian et al. (2019), and Trčka et al. (2020), including the THEMIS dust model (Jones et al. 2017) which we also use in our RT post-processing procedure (see Section 3), except that we specify the Chabrier (2003) IMF for SSPs, consistent with the IMF used in the ARTEMIS and Auriga simulations. For our analysis, we always use the properties corresponding to the most probable 'Bayes' model determined by CIGALE.

## 2.6 StatMorph

STATMORPH (Rodriguez-Gomez et al. 2019) is a PYTHON package for calculating many commonly used non-parametric morphological statistics of galaxy images, including the Gini-M20 (Lotz, Primack & Madau 2004) and concentration-asymmetry-smoothness (CAS; Conselice 2003) statistics, and for fitting 2D Sérsic profiles. The code can handle images with a single source each, which is the mode used in this work, as well as large mosaic images with hundreds or thousands of sources.

In this work we use the exact same procedure as Kapoor et al. (2021) to obtain multiwavelength sets of the elliptical half-light radii and of the CAS indices for a subset of ARTEMIS disc galaxies. This allows us to compare these morphological properties with those already calculated by Kapoor et al. (2021) for similarly selected Auriga galaxies and by Baes et al. (2020) for a set of well-resolved DustPedia spiral galaxies.

The four statistics studied in this work are described in detail by Rodriguez-Gomez et al. (2019) and references therein. We limit the discussion here to a very brief summary:

(i) *Half light radius* ($R_{half}/R_{80}^{opt}$): The half-light radius $R_{half}$ is calculated as the elliptical radius of the isophote that contains half of the light in the galaxy image. We normalize it with $R_{80}^{opt}$, the radius of a circular aperture containing 80 per cent of the galaxy's light in the Sloan Digital Sky Survey (SDSS) *g*-band image.

(ii) *Concentration* (C): The concentration index is defined as $5 \times \log_{10}(R_{80}/R_{20})$, where $R_{20}$ and $R_{80}$ are the radii of circular apertures containing 20 per cent and 80 per cent of the galaxy's light, respectively. The index is a measure of how concentrated the central region or bulge is with respect to the total flux of the galaxy.

(iii) *Asymmetry* (A): The asymmetry index is obtained by subtracting the galaxy image rotated by 180° from the original image. Asymmetry indicates merger events and interactions, or, in regular star-forming galaxies, structures such as spiral arms.

(iv) *Smoothness* (S): The smoothness index is computed by subtracting a lower resolution version of the galaxy image from the original galaxy image. It measures the presence of high spatial frequency features; the index value increases with clumpiness.

## 3 METHODS

Our procedure for preparing synthetic observables of the simulated ARTEMIS galaxies closely follows the procedure employed by Kapoor et al. (2021) for the simulated Auriga galaxies, which is in turn based on the procedure employed by Camps et al. (2016) and Trayford et al. (2017) for the simulated EAGLE galaxies. This similarity in approach makes the results comparable and allows the combined Auriga and ARTEMIS results to be considered as a single, consistent data set.

### 3.1 Extraction and choice of aperture

For each of the 45 ARTEMIS redshift-zero simulation snapshots, we locate the dominant halo, i.e. the halo with the largest stellar mass in its dominant sub-halo, and extract all star and gas particles from that dominant sub-halo. This represents the galaxy of interest in the zoom-in simulation. We do *not* include the other sub-halo's of the dominant halo, representing secondary structures bound to the main galaxy, because these would only interfere with the analysis of the main galaxy. We then transform the particle coordinates so that the origin is in the stellar centre of mass and the *z*-axis coincides with the

---

[1]The open-source SKIRT code is registered at the ASCL with the code entry ascl:1109.003. Documentation and other information can be found at www.skirt.ugent.be.

[2]Specifically, git commit `c70b6ef06ca5` in the master branch of the SKIRT code hosted at www.github.com/SKIRT/SKIRT9

**Table 1.** The optimal global dust-to-metal ratio $f_{\mathrm{dust}}$ found by Camps et al. (2016) for the EAGLE simulations, by Kapoor et al. (2021) for the Auriga simulations, and in Section 3.3 of this work for the ARTEMIS simulations, for the two diffuse dust allocation recipes described in Section 3.2.3.

| Allocation Recipe | $f_{\mathrm{dust}}$ EAGLE | $f_{\mathrm{dust}}$ Auriga | $f_{\mathrm{dust}}$ ARTEMIS |
|---|---|---|---|
| [d]C16 | 0.300 | 0.225 | 0.300 |
| [d]T12 | – | 0.140 | 0.275 |

stellar rotation axis. As a sanity check, we verify that the intrinsic[3] stellar masses within a spherical aperture of 30 kpc match the stellar masses listed in table 1 of Font et al. (2020).

We subsequently preserve only those particles with a position inside a spherical aperture with the largest of the following radii: the 30 kpc radius commonly used for EAGLE (Schaye et al. 2015), the radius at which the face-on stellar surface density within $\pm 10$ kpc of the mid-plane in the vertical direction falls to $2 \times 10^5 \mathrm{M}_\odot \, \mathrm{kpc}^{-2}$ (following Kapoor et al. 2021), and five times the half-stellar-mass radius or $5\mathrm{R}_{\mathrm{M50}}$ (for optimal comparison with DustPedia observables; see below). We call the largest of these radii the *extraction* aperture. The spatial domain of the SKIRT simulation and the field of view of the generated images are adjusted for each galaxy to enclose its extraction aperture.

This approach allows calculating spatially integrated fluxes from the generated images for any of the apertures listed above, or in fact for any aperture up to the extraction aperture. However, the surface brightness for each pixel, and thus the spatially integrated fluxes, will always reflect the line-of-sight radiation across the complete extraction volume. In other words, the smaller apertures are circular (or cylindrical) rather than spherical.

Tests reported by Trčka et al. (2021) indicate that the $5\mathrm{R}_{\mathrm{M50}}$ radius offers the best match to the apertures in the DustPedia galaxy sample. Therefore, all spatially integrated quantities shown and discussed in this work are calculated for that aperture. This includes the luminosities used for recipe calibration, although the choice of aperture does not seem to have a significant effect on the comparisons. Nevertheless, producing and publishing images with the full extraction aperture (see Section 3.4.1) enables other studies, such as the morphology calculations in Section 4.2, to compare with data sets that use a different aperture definition.

### 3.2 Post-processing recipes

During the calibration phase, we explore and fine-tune several variations of our RT post-processing recipe, before finally settling on a single fiducial recipe. We describe the recipe and its variations in this section and report on the calibration results in Section 3.3.

#### 3.2.1 Common procedures

In all cases, following Camps et al. (2016), Trayford et al. (2017), Trčka et al. (2020), Kapoor et al. (2021), regular stellar particles are assigned an SED from the Bruzual & Charlot (2003) template library based on metallicity and age. Also, star-forming region (SF region) particles (as defined in Section 3.2.2) are assigned an SED from the MAPPINGS III (Groves et al. 2008) template library, which

[3]We use the adjective *intrinsic* to indicate a quantity obtained by simply aggregating particle properties.

models the dust enveloping the core H II region in addition to the emission from the young stellar objects. Next to the metallicity, this library requires parameters (ambient pressure, compactness and dust covering fraction) that cannot be directly obtained from the snapshot particle properties and thus require an appropriate heuristic as described in Section 3.2.2.

Following recent work, including Nersesian et al. (2019, 2020a,b) and Verstocken et al. (2020), for constructing 3D models of nearby face-on galaxies and Kapoor et al. (2021) for post-processing the Auriga galaxies, the diffuse dust in our RT simulations is in all cases represented by the THEMIS dust model (Jones et al. 2017) as opposed to the Zubko dust model (Zubko, Dwek & Arendt 2004) used in earlier work (e.g. Camps et al. 2016; Trayford et al. 2017). Where Zubko et al. (2004) explicitly model polycyclic aromatic hydrocarbon (PAH) molecules next to non-composite graphite and silicate grains, the more recent THEMIS model is based on a mixture of amorphous hydrocarbons and amorphous silicates. Our tests indicate that, compared to the Zubko model, for an otherwise fixed recipe, the THEMIS model reduces the discrepancies between simulation and observation in some wavelength regimes but introduces extra tension in other regimes. We explore these differences in more detail in Appendix A.

An important part of our post-processing procedure is the transformation from the sets of stellar and gas particles extracted from an ARTEMIS snapshot to three distinct sets of particles presented to the SKIRT code: SF regions, regular stellar particles, and particles representing dust. We implement two distinct recipes for handling SF regions, dubbed [s]C16 and [s]K21 (see Section 3.2.2), and two distinct recipes for allocating the diffuse dust density distribution, dubbed [d]C16 and [d]T12 (see Section 3.2.3). We combine these into three different complete recipes for further calibration: [s]C16/[d]C16, [s]C16/[d]T12, and [s]K21/[d]T12 (see Section 3.2.4).

#### 3.2.2 Recipes for SF regions ([s]C16, [s]K21)

The [s]C16 recipe follows the procedure prescribed by Camps et al. (2016), Trayford et al. (2017) for the EAGLE simulations (see fig. 2 of Camps et al. 2016). Young star particles (up to 100 Myr) and gas particles with a nonzero SFR are placed in a temporary bin of candidate SF region particles. These particles are randomly resampled to sub-particles with smaller masses following a power-law molecular cloud mass function with $M \in [700, 10^6]\,\mathrm{M}_\odot$. The sub-particles also receive a random formation time. Sub-particles that have not yet formed join the gas particle bin (which will later be used to allocate dust); infant sub-particles (up to 10 Myr) go into the SF region particle bin, and the remaining sub-particles join the regular stellar particle bin. The SF region sub-particles are also randomly relocated within a small region, increasing the realism of the images.

As the final step, the [s]C16 recipe determines the values of the extra parameters needed for the MAPPINGS III templates assigned to the SF region sub-particles. The ambient pressure and compactness are estimated from snapshot particle properties; the dust covering fraction in the photodissociative region (PDR) is set to the fixed value of $f_{\mathrm{PDR}} = 0.1$.

The [s]K21 recipe is the same as the one prescribed by Kapoor et al. (2021) for the Auriga simulations. These authors note that the resolution of modern zoom-in simulations brings the mass of the baryon particles well within the range of the molecular cloud mass function so that it is no longer necessary to resample them to smaller masses, even for SF regions. The [s]K21 recipe thus simply moves

all infant stellar particles to the SF region bin, without involving the star-forming gas particles. This also avoids the need for 'converting' between gas and star particles.

The ˢK21 recipe determines the MAPPINGS III template parameter values through an alternate approach. The compactness parameter, which essentially controls the temperature of the dust in the SF region, is randomly assigned from a Gaussian distribution motivated by observations (Utomo et al. 2019) and previous studies (Kannan et al. 2020). The ambient pressure, which has only a limited effect on the continuum spectrum, is then derived from the compactness and snapshot particle properties. The dust covering fraction is calculated from the age of the SF region (i.e. the infant stellar particle) assuming a fixed molecular cloud dissipation time of $\tau_{\text{clear}} = 1$ Myr.

### 3.2.3 Recipes for allocating dust (ᵈC16, ᵈT12)

Both recipes for allocating diffuse dust first determine the subset of gas particles deemed to carry dust, and then assign dust to these particles using a fixed dust-to-metal ratio, i.e. $M_{\text{dust}} = f_{\text{dust}} Z M_{\text{gas}}$. The gas mass $M_{\text{gas}}$ and the metallicity $Z$ are taken from snapshot particle properties, and the dust fraction $f_{\text{dust}}$ is set to a fixed, global value. The two recipes differ in the heuristic for selecting 'dusty' gas particles and the value of $f_{\text{dust}}$. The latter should be re-calibrated for each recipe through comparison with observations.

The ᵈC16 recipe again follows Camps et al. (2016), Trayford et al. (2017) and is used by Kapoor et al. (2021) under the name recSF8000. This recipe allocates diffuse dust to all gas particles with a non-zero SFR or a gas temperature under 8000 K (or both).

The ᵈT12 recipe follows Torrey et al. (2015) and is used by Kapoor et al. (2021) under the name recT12. This recipe allocates dust to gas particles that are considered to be rotationally supported according to their position in temperature-density phase space; see Fig. 2 and equation (3) of Kapoor et al. (2021). Compared to ᵈC16, ᵈT12 assigns dust to a larger subset of gas particles, extending the dust density distribution to larger radii and causing it to be less compact and somewhat less clumpy. To compensate for the larger amount of 'dusty' gas, the optimal dust-to-metal ratio will be lower than for the ᵈC16 recipe.

Table 1 lists the optimal values for $f_{\text{dust}}$ obtained by Camps et al. (2016) for the EAGLE simulations and by Kapoor et al. (2021) for the Auriga simulations. The last column lists the values obtained in this work for the ARTEMIS simulations as described in Section 3.3. We will also discuss the similarities and differences in these results at the end of that section.

### 3.2.4 Combined recipes

The procedures for handling SF regions and for handling dust are independent of each other and thus the recipes discussed in Sections 3.2.2 and 3.2.3 can be combined at will. We will explore the following combinations:

(i) ˢC16/ᵈC16: resampled SF regions and basic dust allocation.
(ii) ˢC16/ᵈT12: resampled SF regions and extended dust allocation.
(iii) ˢK21/ᵈT12: plain SF regions and extended dust allocation.

We do not explore the fourth possible combination, ˢK21/ᵈC16, because Kapoor et al. (2021) found this combination to be slightly less optimal than ˢK21/ᵈT12 when comparing morphology parameters to observations.
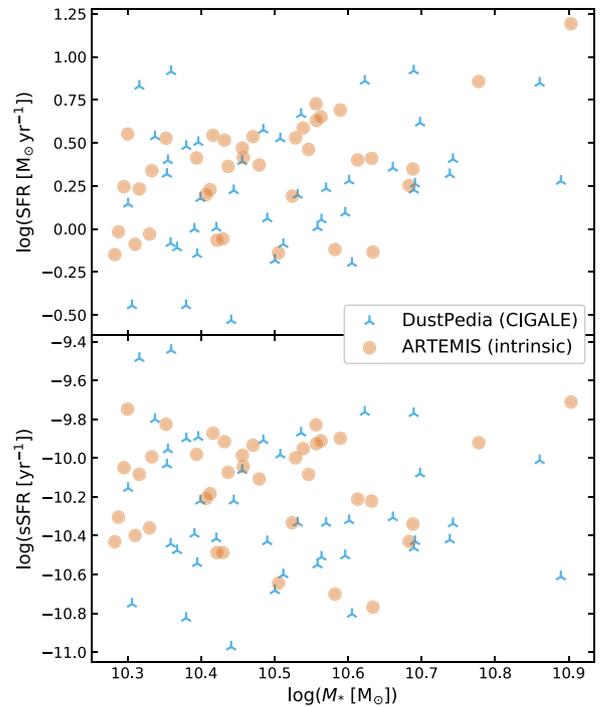


**Figure 1.** SFR (top) and sSFR (bottom) versus stellar mass for the observed DustPedia galaxies (physical properties obtained through SED fitting) and the simulated ARTEMIS galaxies (intrinsic properties from snapshot particles) used for calibrating our RT post-processing recipe. The selection criteria for the calibration sample are discussed in Section 3.3.1.

## 3.3 Calibration

### 3.3.1 Sample selection

When comparing simulation results with observed data, it is important to approximately match the overall properties of the set of galaxies on both sides. Following Kapoor et al. (2021), we use stellar mass and SFR criteria for this purpose. For the DustPedia galaxies, we use the physical properties obtained from the observed fluxes through CIGALE SED fitting as described in Section 2.5. For the ARTEMIS galaxies, we use the intrinsic quantities obtained by accumulating the stellar mass and SFR of the stellar and gas particles, respectively. This is inconsistent in the sense that we are mixing quantities inferred from observed fluxes with intrinsic quantities. Using fluxes for the ARTEMIS galaxies would be circular, however, as we do not yet have a calibrated recipe to produce such fluxes. Moreover, we are using these quantities just to construct approximately matched samples, not to perform the calibration.

It makes no sense to calibrate recipes for handling SF regions and the effects of dust using passive galaxies. We thus eliminate galaxies with sSFR $< 10^{-11}$ yr$^{-1}$ from both the ARTEMIS and DustPedia data sets. Note that we do produce data products for the omitted ARTEMIS galaxies; we just exclude them from the calibration subset. We further limit the DustPedia data set to galaxies with observed stellar mass and SFR in the range of the corresponding intrinsic properties for the ARTEMIS galaxies: $10.30 < \log_{10}(M_*/M_\odot) < 10.92$ and $-1.35 < \log_{10}(\text{SFR}/M_\odot \text{ yr}^{-1}) < 1.25$.

Fig. 1 shows the SFR and sSFR versus stellar mass for the remaining 38 ARTEMIS and 42 DustPedia galaxies. The two populations seem to be similarly distributed and sample the selected parameter space fairly well, demonstrating that the samples are sufficiently comparable for calibrating our post-processing recipe.
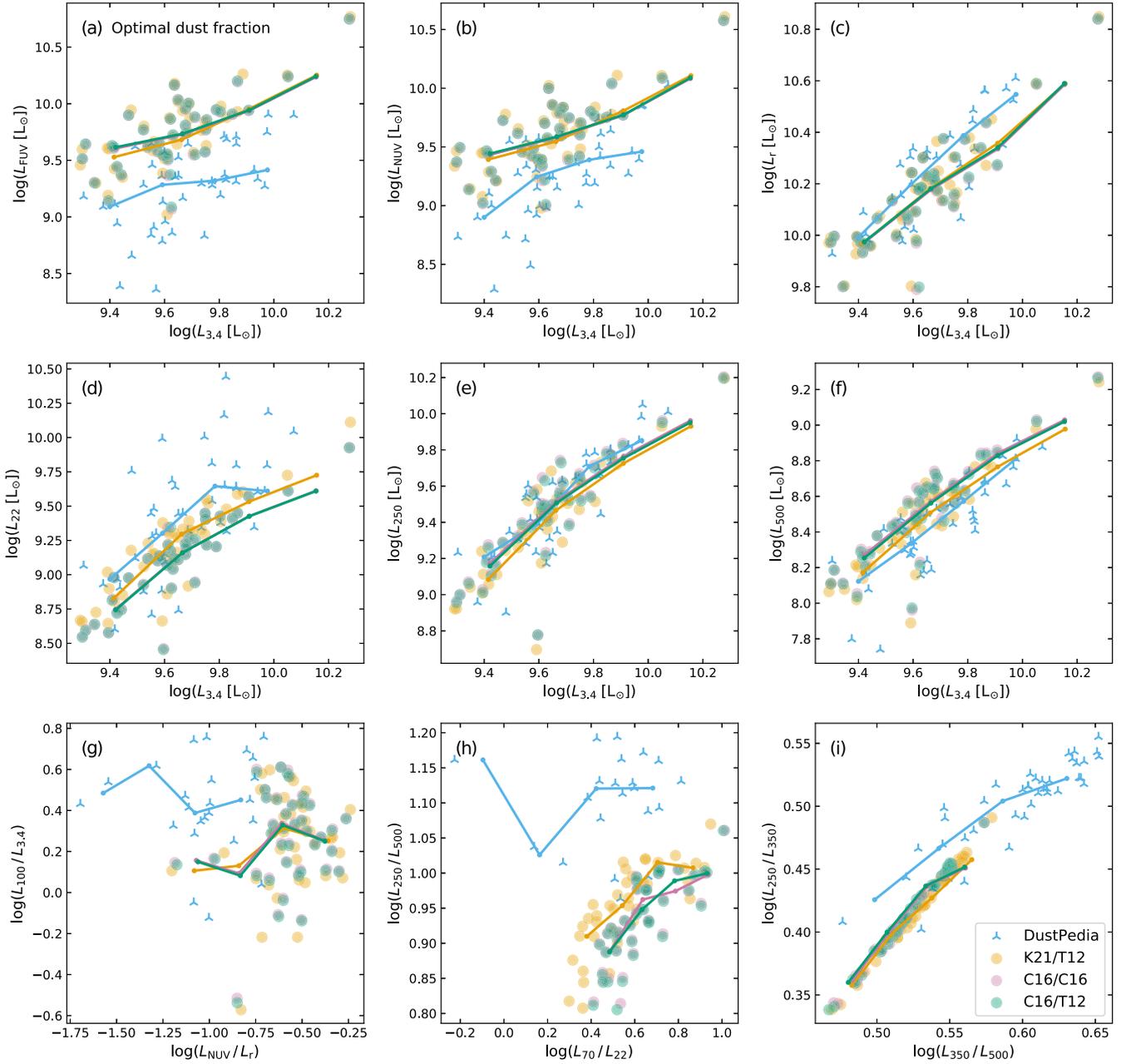
**Figure 2.** Scaling relations for the synthetic ARTEMIS luminosities calculated for a random viewing angle using our three recipes (Section 3.2.4) $^s$K21/$^d$T12 (orange), $^s$C16/$^d$C16 (purple), and $^s$C16/$^d$T12 (green), each with their optimal dust fraction $f_{dust}$ (see Table 1), versus those for the observed DustPedia luminosities (blue). The solid lines connect the median $y$-axis values in a limited number of $x$-axis bins. The data sets are limited as defined in Section 3.3.1 and, for DustPedia, to the galaxies for which the broad-band fluxes under consideration in a given panel have been observed.

We note again that Fig. 1 mixes physical and intrinsic properties. In Section 4.1, we will investigate how the physical properties derived for the ARTEMIS galaxies through SED fitting relate to their corresponding intrinsic properties.

### 3.3.2 Synthetic observations

Calibrating our RT post-processing recipe requires performing a significant number of SKIRT simulations for the ARTEMIS galaxies. Therefore, we limit the number of viewing angles and broad-bands in this phase. We include an edge-on, a face-on and a *random* view, where the latter corresponds to a sight line looking down

from the $z$-axis in the original cosmological coordinate frame (i.e. before the galaxy was rotated). For each of these sightlines, we have SKIRT produce flux densities convolved with the response curve for each of 20 common broad-bands ranging from FUV to submm wavelengths and limited to the $5R_{M50}$ aperture. We then determine the corresponding absolute luminosities $L = \nu L_\nu = \lambda L_\lambda$ taking into account the configured model-instrument distance.

We subsequently verify numerical convergence of these values with regards to discretization choices made for the simulation, including the resolution of the spatial and spectral grids and the number of photon packets launched, and taking into account the randomness inherent to the Monte Carlo radiative transfer procedure.

Our tests confirm that variations in the calculated luminosities caused by numerical issues are always below 8 per cent (0.033 dex) and substantially smaller in most cases.

### 3.3.3 Scaling relations

We employ a select set of luminosity scaling relations for comparing the synthetic ARTEMIS results to the observed DustPedia data, as shown in Fig. 2. The topmost six panels of this figure relate the luminosity for various bands to the 3.4 μm band luminosity, which can be seen as a proxy for stellar mass. Panels (a), (b), and (c) show the effects of dust attenuation at FUV, NUV, and optical wavelengths, while panels (d), (e), and (f) trace dust emission at infrared and submm wavelengths. The NUV (b) and the 22 μm (d) luminosities can also be interpreted as a proxy for SFR.

The three panels in the bottom row of Fig. 2 show colour–colour relations in various wavelength regimes. Panel (g) shows a proxy for specific dust mass versus a proxy for specific dust attenuation. Panels (h) and (i) show infrared and submm colour–colour relations indicative of representative dust temperature. In panel (h), a larger contribution of warm dust leads to a lower 70 μm luminosity relative to 22 μm and to a lower 500 μm luminosity relative to 250 μm. Consequently, data points towards the upper left indicate warmer representative dust temperatures. In panel (i), a similar reasoning leads to the conclusion that data points towards the upper right indicate warmer representative dust temperatures.

In Appendix B, we explore the variations in the ARTEMIS scaling relations for different sightlines and we determine the optimal value of the dust-to-metal fraction $f_{dust}$ for each of the recipes $^{s}$K21/$^{d}$T12, $^{s}$C16/$^{d}$C16, and $^{s}$C16/$^{d}$T12 defined in Section 3.2.4. These values are listed in Table 1. In Fig. 2 and in the discussion here, we focus on the ARTEMIS luminosities calculated for a random viewing angle using our three recipes with optimal $f_{dust}$.

There is reasonable agreement between the synthetic results and the observed data, with some significant exceptions. In the shorter wavelength regime, the FUV (a) and UV (b) luminosities are overestimated by ≈0.5 dex, while the optical luminosities are underestimated by ≈0.15 dex. These opposing differences cause a correspondingly substantial deviation in the $L_{NUV}/L_r$ colour (g). This attenuation discrepancy is in line with the findings of previous work using a similar post-processing recipe (e.g. Baes et al. 2019; Trčka et al. 2020; Kapoor et al. 2021). It cannot be resolved by a straightforward scaling of the stellar emission or of the dust mass. It appears that our procedure insufficiently captures the subgrid extinction processes in the compact and clumpy SF regions. In the infrared wavelength regime, the 22 μm luminosity (d) is underestimated by ≈0.25 dex, depending on the recipe. This is possibly related to the same limitations in our handling of SF regions.

The 100 μm luminosity (g), more or less at the top of the dust continuum emission peak, is also underestimated by ≈0.25 dex. Luminosities on the long side of the continuum peak (e, f), which may be considered basic proxies for dust mass, seem to be predicted fairly accurately, however with opposing discrepancies along the downward slope. We note ≈0.05 dex underestimation for 250 μm and up to ≈0.15 dex overestimation for 500 μm, depending on the recipe. This indicates that the emission peak is shifted to longer wavelengths, corresponding to a larger body of colder dust. This effect is also apparent in the submm–submm colour–colour relation (i) and in the submm-FIR colour–colour relation (h). In both panels, the ARTEMIS data points are significantly shifted toward colder representative dust temperatures compared to the DustPedia observations. Again, as has been noted by Camps et al. (2016) and Kapoor et al. (2021), our pro-

cedure seems to insufficiently capture the subgrid dust heating processes within and in the immediate neighbourhood of the SF regions.

The differences in the scaling relations for our three recipes are generally small, and most prominent for the longer wavelengths. The $^{s}$K21/$^{d}$T12 recipe heats the dust somewhat more efficiently than the other recipes (e,f,h,i) and also performs better in the 22 μm band (d). Both changes can be attributed to the improved handling of SF regions in the $^{s}$K21 scheme. Using the metrics discussed in Appendix B to evaluate the three recipes, the $^{s}$K21/$^{d}$T12 recipe also robustly emerges as the best recipe. We therefore use this recipe for calculating the final data products of this work.

## 3.4 Synthetic data products

### 3.4.1 Description

We use the $^{s}$K21/$^{d}$T12 recipe (see Section 3.2.4) with optimal dust fraction $f_{dust} = 0.275$ (see Appendix B) to produce broad-band images with a spatial resolution of $50 \times 50$ pc per pixel for the 45 ARTEMIS galaxies, for 50 bands ranging from UV to submm wavelengths, and for 18 sightlines with varying inclination and azimuth. The lists of broad-bands and sightlines match those of the synthetic Auriga observables produced by Kapoor et al. (2021).

Specifically, we include the 50 broad-bands listed in Table 4 of Camps et al. (2018), including transmission curves for the most commonly used instruments and observatories across the UV-submm wavelength range. Following the specifications in Section 4 of Kapoor et al. (2021), we use 11 inclinations uniformly sampled in cos $i$, with $i$ the angle between the angular momentum vector of the galaxy and the line of sight, leading to a finer grid close to edge-on positions. For the three inclinations closest to edge-on, we place observers at three different azimuths. For the remaining eight inclinations, we use just a single azimuthal position. We also consider an additional 'random' viewpoint corresponding to a sight line looking down from the $z$-axis in the original cosmological coordinate frame (i.e. before the galaxy was rotated).

All observers are placed at a distance of 20 Mpc from the simulated galaxy. To determine the field of view (in galaxy size units as opposed to angular units) of the images for a given galaxy, we use the extraction aperture defined in Section 3.1, which encloses all stellar and gas particles extracted from the corresponding ARTEMIS simulation snapshot. More precisely, the field of view in each image direction is given by twice the extraction aperture radius, rounded up to $64 \times 50$ pc = 3.2 kpc. This choice ensures that both the $5R_{M50}$ aperture used to calibrate our results against DustPedia observations in Section 3.3 and the surface density-based aperture used by Kapoor et al. (2021) are covered by each image. Furthermore, the rounding ensures that the number of pixels in each direction is always a multiple of 64, which may facilitate binning of the images in later processing steps.

These image data are available for public download at https://www.astro.ljmu.ac.uk/Artemis. As an illustration, Fig. 3 offers face-on and edge-on optical views for selected ARTEMIS galaxies, composited from the public data set.

### 3.4.2 Convergence

To avoid artefacts caused by the simulation dust grid and limit the noise level in the image pixels, we substantially increase the grid resolution and the number of photon packets compared to the calibration simulation parameters discussed in Section 3.3.2. Depending on the dust distribution in the galaxy, the number of
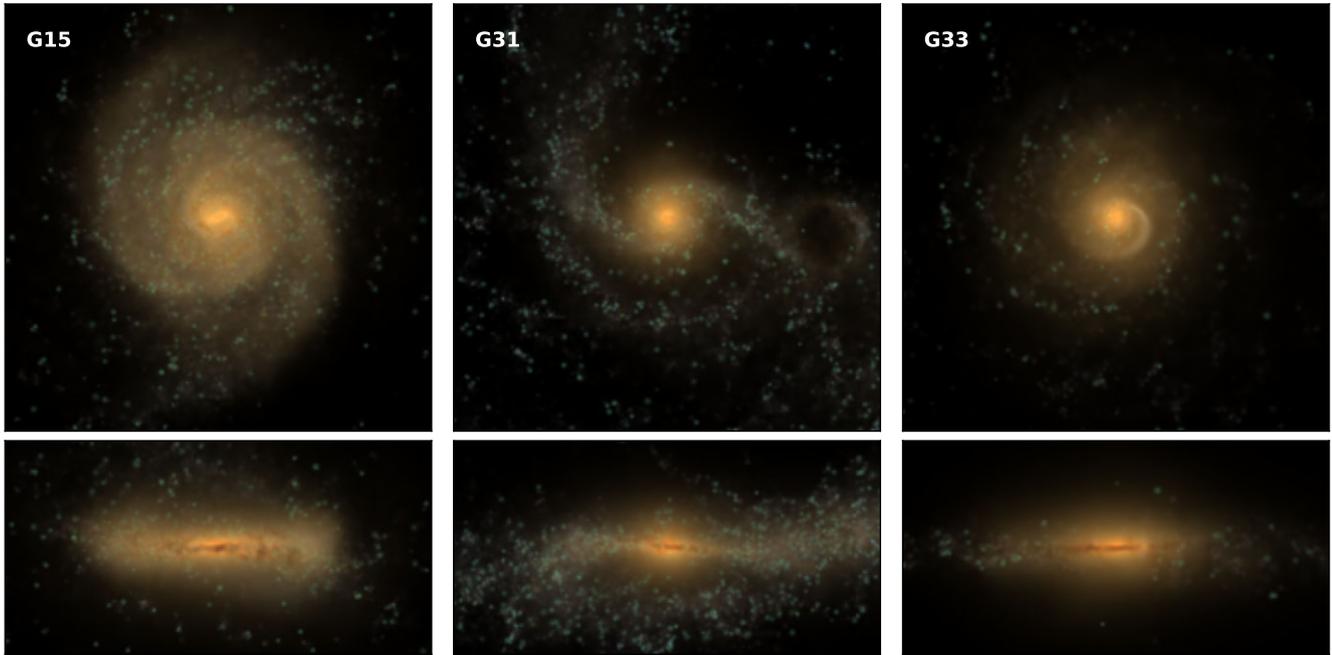
**Figure 3.** Face-on (top row) and edge-on (bottom row) colour-composite views of selected ARTEMIS galaxies using the synthetic SDSS *i*, *r*, and *g* observations produced in this work for the red, green, and blue channels, respectively. The field of view is 60 kpc.

spatial grid cells ranges from 2 to 13 million, with the cells in the densest diffuse dust regions reaching down to 5 pc on a side – far below the 50 pc image pixel size. The number of photon packets launched for both primary and secondary emission ranges from 5 to $25 \times 10^9$ and is determined for each galaxy using a heuristic as a function of the number of input particles and the number of pixels in the output images. These discretization settings cause a high level of convergence for spatially integrated quantities calculated from the images. Our tests confirm that variations on integrated luminosities caused by numerical issues are well below one per cent for all wavelengths and sightlines as long as the employed aperture is not exceedingly small.

More importantly, we need to evaluate convergence on a pixel by pixel basis. Following Kapoor et al. (2021), we calculate the relative error *R* based on Monte Carlo statistics recorded for each pixel (Camps & Baes 2020) in a representative set of broad-band images at selected viewing angles for all ARTEMIS galaxies. According to Camps & Baes (2020), the pixel value can be considered to be reliable for $R < 0.1$, it is questionable in the range $0.1 < R < 0.2$, and it is unreliable for $R > 0.2$. Fig. 4 shows these statistics for the face-on (top half) and edge-on (bottom half) views of a single representative ARTEMIS galaxy in four selected bands from UV to submm wavelengths.

The synthetic surface brightness maps in the top row of each half of the figure use a logarithmic colour scale; the transition between red and blue marks 1/100 of the maximum surface brightness. Each of the underlying data frames has $1728^2$ $50 \times 50$ pc pixels, corresponding to the 86 kpc field of view (rounded up to a multiple of 64 pixels). The second row in each half of the figure ($b = 1$) shows the corresponding *R* value for each of these pixels indicating reliable (green), questionable (orange) and unreliable (red) image areas. The third row in each half of the figure ($b = 2$) shows the same statistic after $2 \times 2$ binning into $100 \times 100$ pc pixels. As expected, this binning results in a substantial increase in reliability at the cost of lower resolution. At the binned $b = 2$ level, the reliable

area (green) essentially covers all pixels with a value down to 1/100 of the maximum surface brightness (yellow and red). At the original $b = 1$ level, one needs to include the questionable area (orange) to achieve a similar coverage.

Inspection of the *R* values in representative broad-bands for the face-on, edge-on and random viewing angles confirms that convergence for the other ARTEMIS galaxies is similar to the results shown in Fig. 4. We note that the FUV/NUV bands tend to show somewhat poorer statistics because of the relatively lower fluxes and higher extinction involved in that wavelength range.

## 4 ANALYSIS

### 4.1 Global physical properties

Given the synthetic images described in Section 3.4, we now derive global physical properties for the ARTEMIS galaxies and compare those to observations. As a first step, we calculate global fluxes by integrating the surface brightness maps within the $5R_{M50}$ aperture of the galaxy and convert these to luminosities taking into account the assumed model-instrument distance. This yields results similar to those employed during calibration (see Section 3.3.2), but now we can calculate luminosities for the full complement of 18 sightlines and 50 broad-bands. Subsequently, we use the CIGALE SED fitting code (Noll et al. 2009; Boquien et al. 2019) to estimate stellar mass, SFR, and dust mass from a relevant subset of 25 of these broad-band luminosities spanning the UV-submm wavelength range. As described in Section 2.5, we use the same parameter settings as those used to obtain physical properties of the DustPedia galaxies so that we can compare simulated and observed galaxies on equal footing.

#### 4.1.1 Inferred versus intrinsic properties

As an initial sanity check, Fig. 5 compares these inferred physical properties to the corresponding intrinsic properties for the ARTEMIS
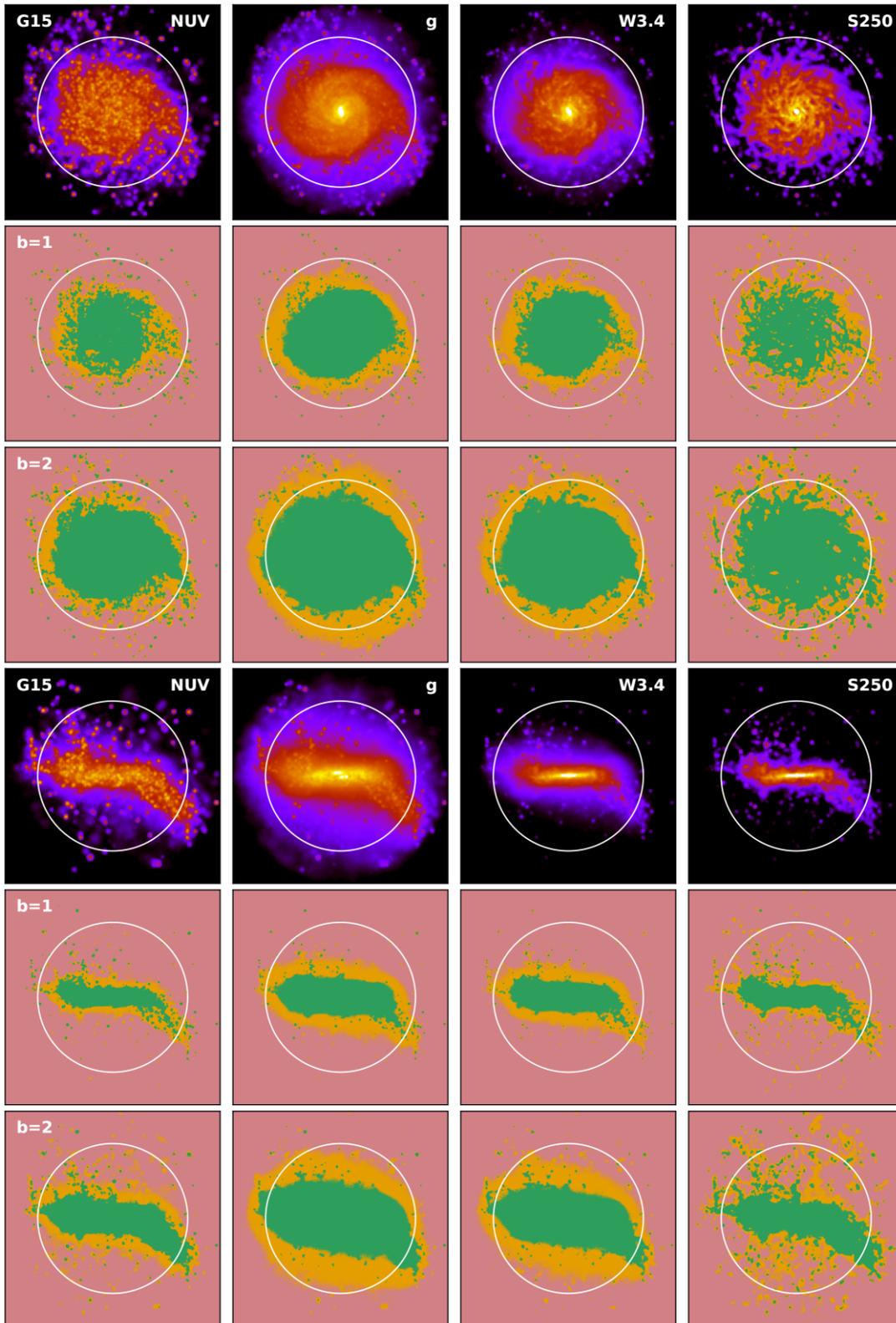
**Figure 4.** Face-on (top half) and edge-on (bottom half) views of the ARTEMIS galaxy G15 in four bands; from left to right GALEX NUV, SDSS *g*, *WISE* 3.4 μm, and *Herschel* SPIRE 250 μm. G15 has an intrinsic stellar mass of $3.57 \times 10^{10}$ M$_\odot$ inside the 30 kpc radius indicated by the white circle. The $5R_{M50}$ aperture radius is 36 kpc and the extraction aperture is 43 kpc corresponding to the 86 kpc field of view of the images. The top row in each half shows the synthetic observations produced by SKIRT using a logarithmic colour scale; the transition between red and blue marks 1/100 of the maximum surface brightness. The other two rows in each half show the corresponding convergence statistics indicating reliable (green), questionable (orange) and unreliable (red) image areas assuming 50 pc pixels ($b = 1$) and binned 100 pc pixels ($b = 2$), as discussed in Section 3.4.2.
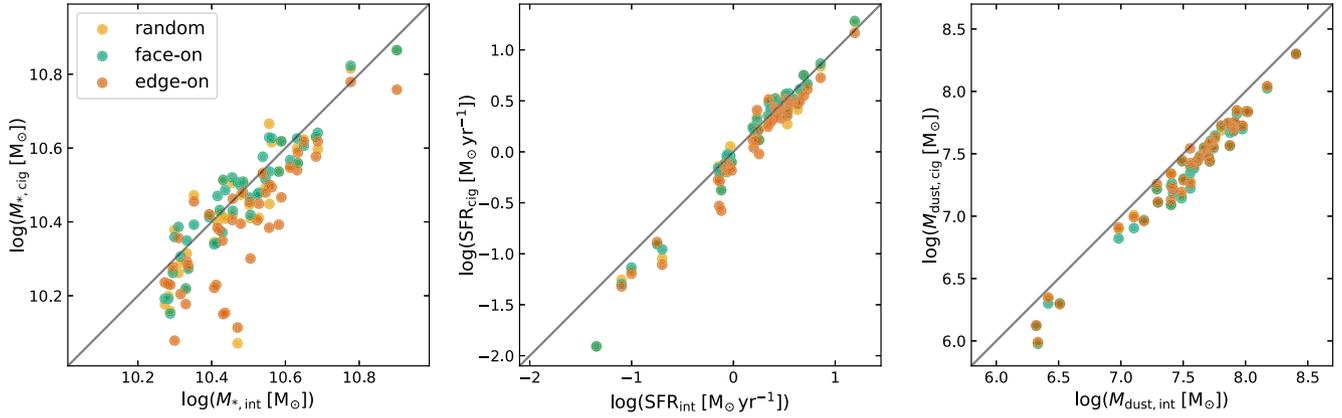
**Figure 5.** Global physical properties of the ARTEMIS galaxies derived from synthetic observations (vertical axis) for three sightlines (see legend) versus the corresponding intrinsic properties (horizontal axis). From left to right: stellar mass, SFR, and dust mass. The solid line indicates the one-to-one relation. The dustless galaxies G21 and G37 are not shown.

galaxies. The stellar mass (left-hand panel) is estimated accurately within ±0.1 dex from the face-on view but is underestimated by up to 0.4 dex from the edge-on view. We will further investigate the inclination dependence of the stellar mass estimate later in this subsection. The SFR (middle panel) is also estimated well (±0.25 dex except for one outlier) with a much smaller inclination dependence. For both stellar mass and SFR, there is an increasing underestimation for lower stellar mass/SFR values, even for the face-on results. Kapoor et al. (2021) do not see such a trend for the Auriga galaxies (private communication), but this is not really in conflict because the Auriga intrinsic stellar masses are $\gtrsim 10^{10.5}$ M$_\odot$. Trčka et al. (2020) do not find a significant trend with stellar mass in their analysis of the EAGLE galaxies, although their Fig. 4 does show some outliers in the same stellar mass/SFR range. The origin of this discrepancy is unclear. One possible cause is related to the sampling of SF regions. The $^s$C16 recipe employed for EAGLE splits SF region particles into smaller sub-particles, while the $^s$K21 recipe employed for both the Auriga and our ARTEMIS results does not. Galaxies with lower stellar mass/SFR necessarily have fewer SF particles, causing a poorer statistical sampling that may lead to systematic effects.

The dust mass (right-hand panel of Fig. 5) is systematically underestimated by 0.20 ± 0.15 dex. As expected, there is no significant inclination dependence because thermal dust emission is essentially isotropic. Kapoor et al. (2021) find a similar systematic underestimation of the dust mass for the Auriga galaxies. Hunt et al. (2019) compare methods for fitting SEDs to the most recent photometry (Dale et al. 2017) of the nearby star-forming galaxies in the KINGFISH survey (Kennicutt et al. 2011). The methods under study include the SED fitting codes CIGALE and MAGPHYS (da Cunha, Charlot & Elbaz 2008) and a method employing a library of SEDs produced from spheroidal GRASIL models (Silva et al. 1998) through an RT process. The authors note that GRASIL tends to report dust masses larger than those of CIGALE or MAGPHYS by ≈0.3 dex. They subsequently conclude that, because GRASIL is the only method that performs RT in realistic geometries, this may indicate that the other methods are underestimating dust mass.

Dudzevičiūtė et al. (2020) perform an analysis similar to ours for the 9431 galaxies at redshift $z > 0.25$ and with SFR > 10 M$_\odot$ yr$^{-1}$ in the reference EAGLE simulation (Schaye et al. 2015). The authors use MAGPHYS (da Cunha et al. 2008) to derive physical properties from synthetic SEDs produced via SKIRT (Camps et al. 2018) and then compare these inferred properties to the intrinsic properties in

their supplemental Fig. A2. Although the galaxies in their analysis are at non-zero redshift and have, on average, a much higher SFR than the galaxies in our study, it is interesting to note a number of differences and similarities. The inferred stellar mass in their analysis is systematically underestimated by ≈0.3 dex and the SFR by ≈0.1 dex. These systematic deviations might be caused in part by inclination effects (see our Fig. 5, left and middle panel; the authors presumably used a random inclination), in addition to different model assumptions in both the radiation transfer (e.g. the dust model) and the SED fitting (e.g. the IMF). On the other hand, the scatter of the inferred properties around the best-fitting line is very similar to the scatter in our results, and there is a noticeable trend towards larger stellar mass underestimation for lower-mass galaxies similar to our findings (Fig. 5, left-hand panel). The dust mass in their analysis is systematically underestimated by ≈0.2 dex, which is very similar to our results (Fig. 5, right-hand panel).

We now come back to the inclination dependence of the stellar mass estimate mentioned earlier in this subsection. Fig. 6 shows the inferred stellar mass of the ARTEMIS galaxies as a function of inclination. The curves are colour-coded for the intrinsic stellar mass of the corresponding galaxy as indicated by the colour bar. Although for some galaxies the estimated mass seems to dip and rise almost arbitrarily with inclination, most galaxies show a systematic underestimation at high inclinations. Previous studies (e.g. Małek et al. 2018; Trayford et al. 2020; Trčka et al. 2020) have found that adopting an attenuation curve with a slope that is more shallow than typically assumed observationally can lead to underestimation of the attenuation at optical wavelengths and therefore underestimation of the stellar mass. Kapoor et al. (2021) find a similar trend of underestimation at high inclinations for the Auriga galaxies. They also show that the attenuation curve fitted by the CIGALE code to the Auriga galaxies is significantly more shallow than the attenuation curves observed for the DustPedia galaxies, supporting the above reasoning.

The question remains why this effect is more prominent at high inclinations. Many SED fitting codes, including CIGALE, assume energy balance between the stellar light absorbed by dust and the thermal radiation emitted by dust. However, while the thermal emission at long wavelengths is virtually isotropic, the observed stellar light depends significantly on the dust attenuation experienced along a given sight line, breaking energy balance for that particular line of sight. We can thus intuitively expect the accuracy of SED fitting to depend on the optical attenuation.
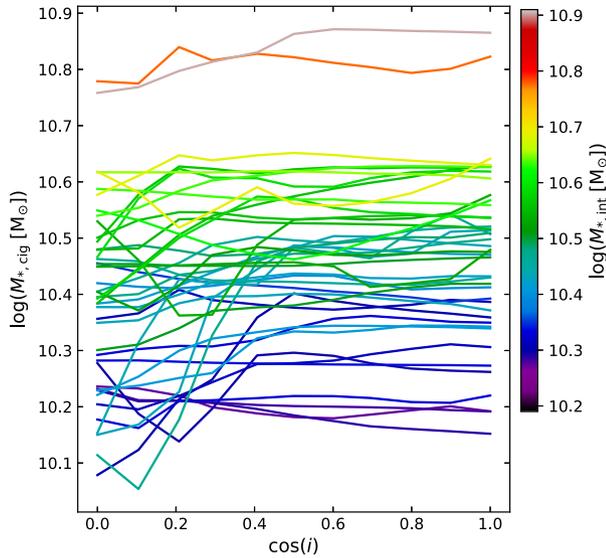
**Figure 6.** Inferred stellar mass of each ARTEMIS galaxy as a function of the inclination of the synthetic observation from which it has been derived. The curves are colour coded for the intrinsic stellar mass of the corresponding galaxy, as indicated by the colour bar. The dustless galaxies G21 and G37 are not shown.
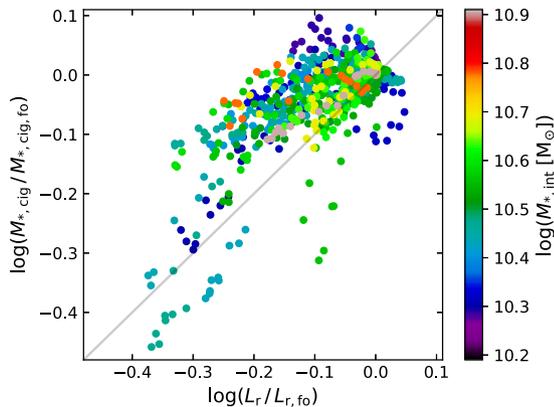


**Figure 7.** Variation in the inferred stellar mass for various inclinations of the ARTEMIS galaxies (relative to the face-on value), as a function of the corresponding variation in SDSS $r$ luminosity (also relative to the face-on value). The dots are colour-coded for the intrinsic stellar mass of the corresponding galaxy using the same colour scheme as in Fig. 6. The solid line indicates the one-to-one relation. The dustless galaxies G21 and G37 are not shown.

To investigate this further, Fig. 7 shows the variation in the inferred stellar mass for the ARTEMIS galaxies as a function of the corresponding variation in SDSS $r$ luminosity, in both cases relative to the face-on value. Each galaxy is represented by 18 dots, one for each of the simulated sightlines, and these dots are coloured for the galaxy's intrinsic stellar mass as in Fig. 6. The luminosity variation has a zero or negative value in virtually all cases. In other words, as expected, the face-on view used as a reference usually has the highest luminosity, and we can interpret the values on the horizontal axis as a proxy for attenuation at optical wavelengths. Similarly, because the face-on inferred stellar mass correlates well with the intrinsic stellar mass (Fig. 5, left-hand panel) we can interpret the values on the vertical axis as a proxy for stellar mass underestimation. With this in mind, Fig. 7 shows a clear overall correlation between the
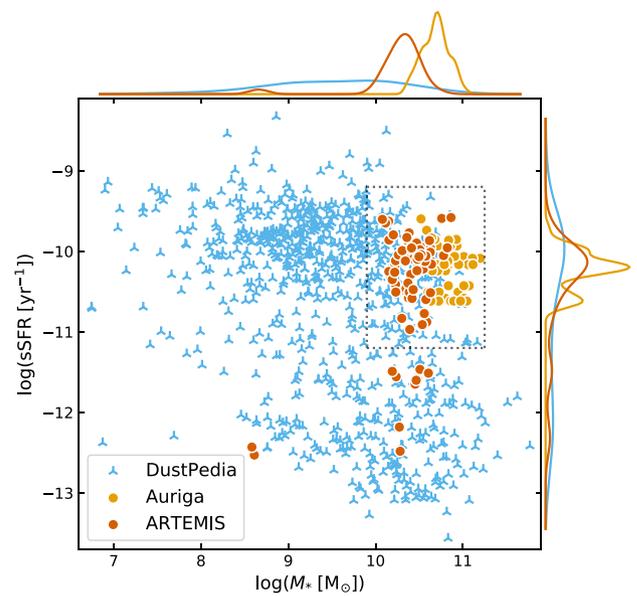


**Figure 8.** Specific SFR versus stellar mass for the ARTEMIS (red) and Auriga (orange) galaxies over-plotted on the DustPedia observations (blue). The galaxy properties are inferred through SED fitting from synthetic observations for edge-on and face-on sightlines (ARTEMIS and Auriga, using the $^s$K21/$^d$T12 recipe) or from actual observations (DustPedia). The non-star-forming ARTEMIS galaxies G21 and G37 are not shown. The dotted rectangle indicates the galaxy selection used for the bottom row of Fig. 9.

stellar mass underestimation by the SED fitting algorithm and the attenuation for a given sight line. Many individual galaxies show the same trend, with multiple dots forming an approximate line, often with a similar slope as the overall trend. On the other hand, there is a significant amount of scatter on the relation ($\approx 0.3$ dex). This is not surprising, given that the observed attenuation will depend substantially on the precise star-dust geometry, especially in near-edge-on configurations.

### 4.1.2 Dust scaling relations

We now compare simulated and observed data sets using galaxy properties inferred through SED fitting of synthetic fluxes (ARTEMIS and Auriga) or observed fluxes (DustPedia). Fig. 8 presents our three data sets in the sSFR–$M_*$ plane. The figure shows all galaxies with sSFR and $M_*$ above the lower axis limits. This excludes just a few low-mass DustPedia galaxies, two non-star-forming ARTEMIS galaxies, and no Auriga galaxies.

The Auriga zoom simulations are selected using a halo mass cut-off of $1 \times 10^{12} < M_{200}/M_\odot < 2 \times 10^{12}$ in addition to a requirement of relative environmental isolation (see Section 2.2). The resulting galaxies consequently occupy a fairly limited region in the upper right corner of the sSFR–$M_*$ plane (Fig. 8). The ARTEMIS zoom simulations use a more relaxed halo mass cutoff of $8 \times 10^{11} < M_{200}/M_\odot < 2 \times 10^{12}$ without any environmental criteria (see Section 2.1). The resulting ARTEMIS data set includes galaxies in a lower stellar mass range. Considering only active galaxies with sSFR $> 10^{-11}$ yr$^{-1}$, the low end of the stellar mass range is decreased from $\approx 3 \times 10^{10}$ M$_\odot$ to $\approx 1 \times 10^{10}$ M$_\odot$. It is evident from Fig. 8 that the ARTEMIS data set supplements the Auriga data set in this manner. Although there are also a few passive ARTEMIS galaxies with sSFR $< 10^{-11}$ yr$^{-1}$, this region of the sSFR–$M_*$ plane remains largely under-sampled.
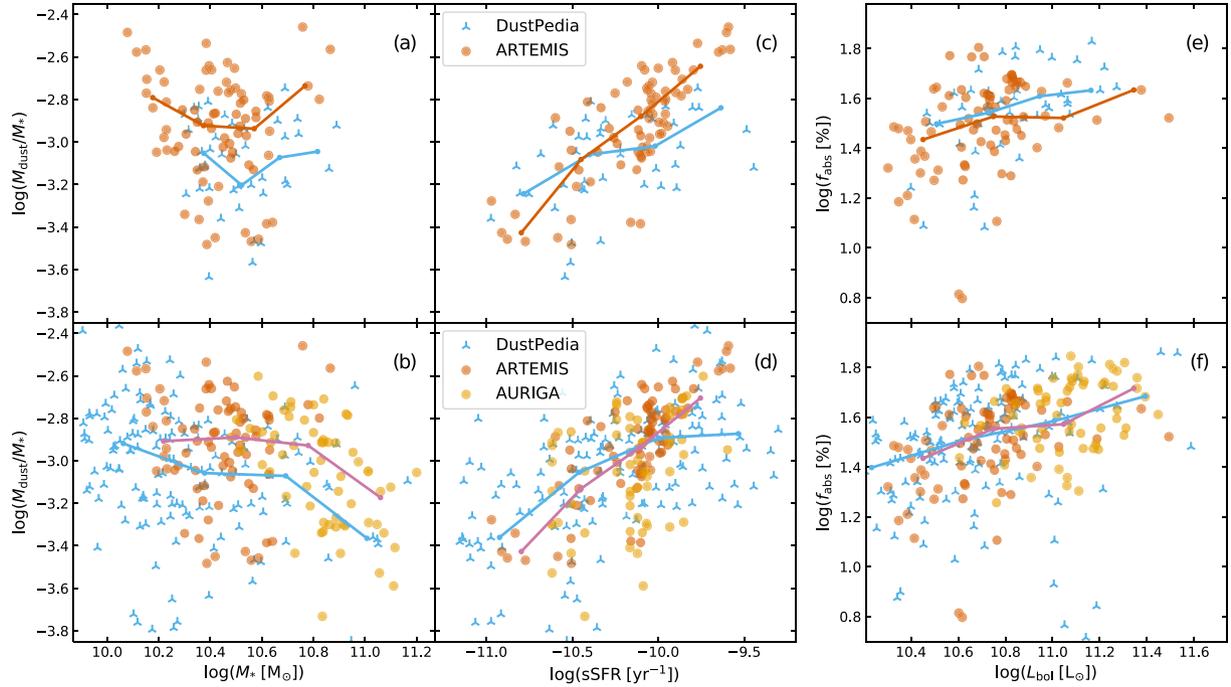
**Figure 9.** Dust scaling relations for the ARTEMIS (red), AURIGA (orange), and DustPedia (blue) galaxies. The top row compares the matched ARTEMIS and DustPedia samples defined in Section 3.3.1. The bottom row compares the combined set of ARTEMIS and Auriga galaxies to the DustPedia galaxy subset with sSFR–$M_*$ values inside the dotted rectangle shown in Fig. 8. The running median curves trace the combine data sets. The galaxy properties are inferred through SED fitting from synthetic observations for edge-on and face-on sightlines (ARTEMIS and Auriga, using the $^s$K21/$^d$T12 recipe) or from actual observations (DustPedia). The first two columns show the specific dust mass versus stellar mass and versus sSFR, respectively. The third column shows the fraction of energy absorbed by dust as a function of bolometric luminosity.

Fig. 9 shows dust scaling relations for our data sets. The top row compares the matched ARTEMIS and DustPedia samples defined in Section 3.3.1. The bottom row compares the combined set of ARTEMIS and Auriga galaxies to a DustPedia galaxy subset defined by the sSFR and $M_*$ limits indicated by the dotted rectangle in Fig. 8. These limits have been chosen to enclose the DustPedia galaxies in the sample matching ARTEMIS (defined in Section 3.3.1) and those in the sample matching Auriga (defined by Kapoor et al. 2021, DPD45). These panels again illustrate how the ARTEMIS and Auriga data sets supplement each other, although we do need to keep in mind that they originate from simulations with different assumptions and subgrid physics. Generally speaking, the synthetic galaxy scaling relations correspond to the observations fairly well. We now discuss each column in turn.

The first column of Fig. 9 shows specific dust mass versus stellar mass. We recall from Section 3.3.1 that the DustPedia sample in panel (a) has been selected using the intrinsic ARTEMIS stellar mass range. As discussed in Section 4.1.1 and illustrated in Fig. 5, the CIGALE SED fitting procedure underestimates the intrinsic stellar mass by $0.20 \pm 0.15$ dex. This causes a corresponding shift to lower masses of the ARTEMIS galaxies compared to the DustPedia sample in panel (a) of Fig. 9. Furthermore, for a given stellar mass, the ARTEMIS specific dust mass is higher than that observed for DustPedia by $\approx 0.3$ dex. The expected decrease of specific dust mass with increasing stellar mass is not seen in this panel, possibly because of the narrow stellar mass range and the restricted sSFR range (limiting the effect of an increasing passive fraction with stellar mass). The downward trend in the relation is recovered when combining the ARTEMIS and Auriga galaxies as shown in panel (b), although the specific dust mass for the synthetic galaxies generally remains

too high. Kapoor et al. (2021) argue that this discrepancy might be caused by a high gas content of the simulated galaxies compared to observations rather than by issues in the post-processing procedure.

The second column of Fig. 9 shows specific dust mass versus sSFR. The observed DustPedia relation is reproduced well by both the ARTEMIS (panels c and d) and the Auriga galaxies (panel d), although there is a slight discrepancy in the slope. Compared to observations, the synthetic galaxies show a slightly more rapid increase in specific dust mass with increasing sSFR. In any case, the panels confirm a clear correlation between (specific) dust mass and (specific) SFR.

The third column of Fig. 9 shows the fraction of energy absorbed by dust, defined as $f_{abs} = L_{dust}/L_{bol}$, versus bolometric luminosity. The observed DustPedia relation and scatter are reproduced excellently by both the ARTEMIS (panel e and f) and the Auriga galaxies (panel f). The Auriga galaxies are, on average, more luminous than the ARTEMIS galaxies, so that the two simulated data sets occupy largely distinct regions in the $f_{abs}$–$L_{bol}$ plane (panel f). Still, the slope of the $f_{abs}$–$L_{bol}$ relation for the combined simulated data set very closely follows the observed slope across the full luminosity range.

## 4.2 Morphology on resolved scales

In this section, we investigate selected non-parametric morphological parameters of the ARTEMIS disc galaxies as a function of wavelength, comparing the results to DustPedia observations and to the simulated Auriga galaxies. We use the STATMORPH package (Rodriguez-Gomez et al. 2019) to derive the CAS indices (concentration, asymmetry, smoothness) and the normalized elliptical half light radius ($R_{half}/R_{80}^{opt}$) from the ARTEMIS images in 14 broad-bands
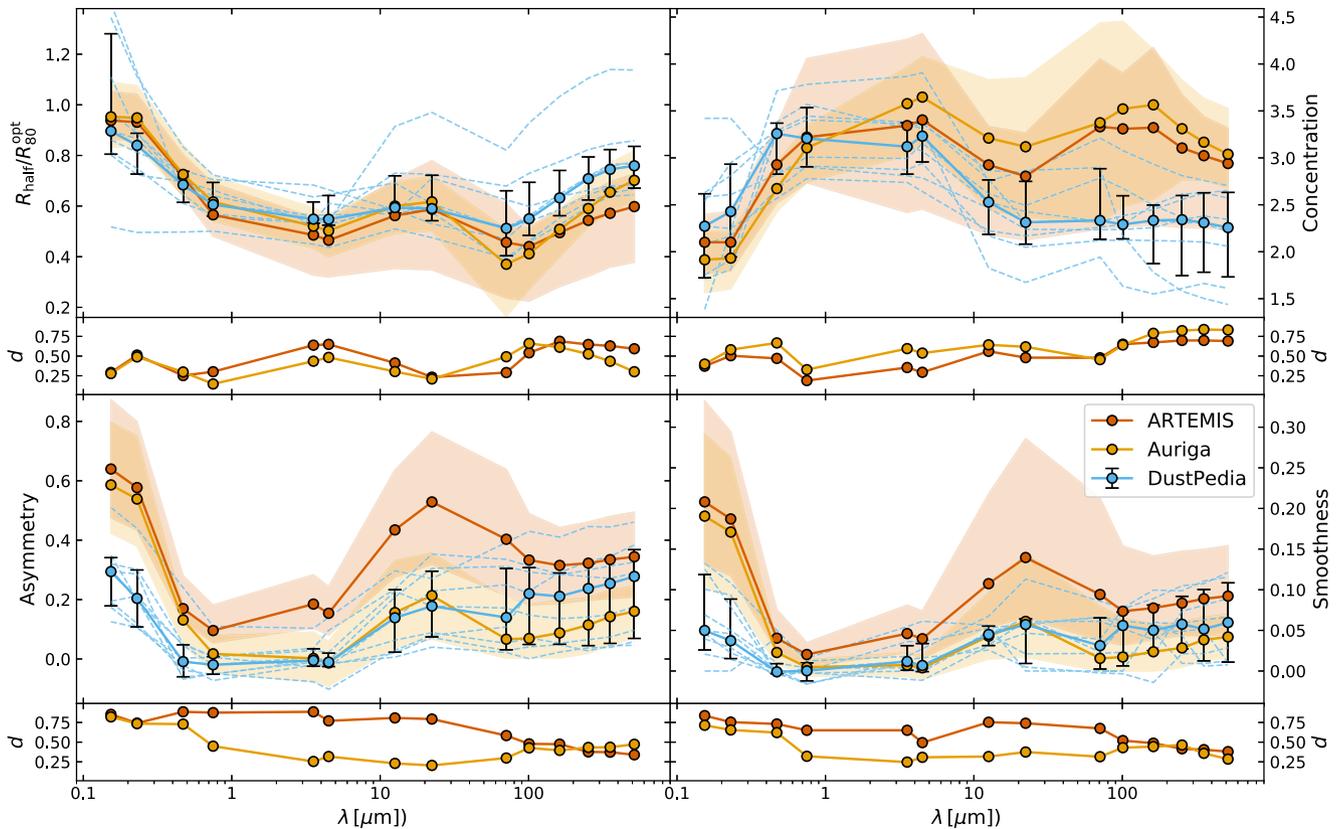
**Figure 10.** Morphological indices (half light radius, concentration, asymmetry, and smoothness; see Section 2.6 for definitions) as a function of wavelength for three sets of disc galaxies. The nine DustPedia galaxies (blue) correspond to those studied by Baes et al. (2020). The Auriga data set (orange) corresponds to the 14 galaxies studied by Kapoor et al. (2021), and the ARTEMIS data represent the 11 ARTEMIS galaxies with a disc-to-total stellar mass ratio of D/T > 0.45, using the ${}^s$K21/${}^d$T12 recipe for five inclinations ranging from face-on to $i = 73°$ in both cases. The circular markers represent the median values for each set; the error bars (DustPedia) or shaded areas (synthetic data sets) indicate the $\pm 1\sigma$ interval. The dashed lines show individual DustPedia galaxies. The small sub-panels under each panel show the K–S test distance $d$ as a function of wavelength, quantifying the dissimilarity between synthetic and observed data sets, i.e. Auriga versus DustPedia (orange) and i.e. ARTEMIS versus DustPedia (red).

ranging from UV to submm wavelengths. Section 2.6 offers some background on the STATMORPH code and a concise definition of the morphological indices used here.

We use the same wavelength bands as those employed by Baes et al. (2020) to study the morphology of 9 well-resolved spiral galaxies in the DustPedia data base and by Kapoor et al. (2021) to study a set of 14 disc galaxies from the Auriga simulations. Following Kapoor et al. (2021), we select ARTEMIS galaxies with a disc-to-total stellar mass ratio of D/T > 0.45, using the intrinsic stellar mass values listed by Font et al. (2020, Table 1). This yields a set of 11 galaxies. Just as was done for the Auriga galaxies, we employ synthetic images for five inclinations ranging from face-on to $i = 73°$ for each galaxy, leading to $5 \times 11$ data points for each wavelength. For both synthetic data sets, the images have been produced using the ${}^s$K21/${}^d$T12 recipe (see Section 3.2.4). The ARTEMIS image field of view encloses the galaxy's full extraction aperture (see Section 3.1 which always includes the stellar surface density-based aperture used by Kapoor et al. (2021) for Auriga.

Fig. 10 shows the four morphological indices under study as a function of wavelength for these three sets of disc galaxies. The circular markers represent the median values for each set; the error bars (DustPedia) or shaded areas (ARTEMIS, Auriga) indicate the $\pm 1\sigma$ interval. The dashed lines show individual DustPedia galaxies. The small sub-panels under each panel show a metric $d$ quantifying

the 'distance' between synthetic and observed data sets as a function of wavelength. This metric is calculated using the 1D two-sample Kolmogorov-Smirnov test (K–S test, Kolmogorov 1933; Smirnov 1948).

The observed normalized half light radius $R_{half}/R_{80}^{opt}$ (the upper left-hand panel of Fig. 10) shows a characteristic trend as a function of wavelength, with large values in the FUV and a gradual decrease over the optical regime to the NIR, followed by a small increase in the MIR and another dip before a final increase towards FIR and submm wavelengths. Both the Auriga and the ARTEMIS simulations reproduce the observed DustPedia trend well, although the median ARTEMIS radii are consistently smaller than the corresponding Auriga values and, for wavelengths longer than optical, also smaller than the DustPedia values. The difference is most notable in the FIR wavelength range corresponding to dust emission, implying that ARTEMIS dust discs are smaller relative to their stellar discs than those in the Auriga simulations and the DustPedia galaxies.

The observed concentration index (the upper right-hand panel of Fig. 10) shows a characteristic trend with high concentration for NIR and MIR wavelengths, tracing older stellar populations, and substantially lower concentration for shorter and longer wavelengths, tracing younger stellar populations and dust. The ARTEMIS and Auriga simulations reproduce the observed DustPedia trend fairly well for wavelengths shorter than $\approx 10$ μm. For longer wavelengths, both

simulations show substantially higher concentration than observed. Despite this discrepancy, the ARTEMIS concentration values are consistently closer to the DustPedia values than the Auriga results, particularly in wavelength regimes that trace dust either through extinction (UV and optical) or emission (FIR and submm). Because the same dust allocation scheme has been used for post-processing ARTEMIS and Auriga galaxies, and given the constant dust-to-metal ratio in this recipe, this seems to indicate that the ARTEMIS simulations include a slightly better prediction of the metal distribution in the galaxy, although still falling short of observations (at longer wavelengths, the two simulations are much closer to each other than to the data). Because there are many differences between the subgrid recipes of the two simulations, it is hard to pin down the precise cause. Most likely, differences in the stellar and AGN feedback mechanisms play a significant role, as these processes affect the metal distribution in various ways (e.g. for the effect of AGN feedback on the resolved distribution of metals in the EAGLE simulations, see Trayford & Schaye 2019).

In the UV and optical wavelength range, the asymmetry and smoothness indices (bottom row of Fig. 10) for both ARTEMIS and Auriga are very similar to each other and both are substantially higher than observed for DustPedia. These high values are probably caused by the SF regions, which are prominent at those wavelengths (see Fig. 4 for two examples). As noted by Kapoor et al. (2021), these index values may improve in case the SF regions would be resampled during post-processing as in recipe $^{\rm s}$C16 (see Section 3.2.2). At longer wavelengths, starting from the NIR, the ARTEMIS galaxies continue to show asymmetry and smoothness values well above the DustPedia reference values, while the Auriga galaxies are very close to observations. Notably, for wavelengths longer than $\approx 100\,\mu$m, the ARTEMIS galaxies continue to overpredict while the Auriga galaxies underpredict by roughly the same amount. This discrepancy might be related to differences in the merger histories for the two simulated galaxy sets. ARTEMIS includes histories with more recent mergers, whereas Auriga has a criterion for isolation.

All in all, the morphology of our ARTEMIS galaxy sample follows the same overall trends as a similar DustPedia galaxy sample, with some notable discrepancies. Using the same post-processing recipe, the ARTEMIS simulations provide better predictions for the concentration index than the Auriga simulations but falls short for the other indices under study, at least in some wavelength regimes.

# 5 SUMMARY AND CONCLUSIONS

In this work, we produce and publish multiwavelength, spatially resolved synthetic observations for the 45 simulated galaxies of the ARTEMIS project (Font et al. 2020, 2021). These galaxies were selected to have a Milky Way-like halo mass and were re-evolved to redshift zero at high resolution including full (subgrid) baryonic physics. We extract stellar and gas properties from the present-day galaxy snapshots with the purpose of calculating synthetic observables with our RT code SKIRT (Camps & Baes 2015, 2020). We assign emission spectra and dust characteristics following a combination of previously developed prescriptions (Camps et al. 2016; Trayford et al. 2017; Kapoor et al. 2021). Stellar populations are modelled through the Bruzual & Charlot (2003) template library. We include a subgrid treatment of SF regions using the MAPPINGS III template library (Groves et al. 2008) to help capture the dust RT processes in their dense and clumpy cores. We allocate diffuse interstellar dust to the galaxy's cold gas using a fixed dust-to-metal ratio, $f_{\rm dust}$, which is treated as a free parameter. The dust properties are taken from the THEMIS dust model (Jones et al. 2017). We

explore variations of the recipe with or without re-sampling of the SF regions and with a more concentrated or more extended dust allocation scheme.

We calibrate the value of $f_{\rm dust}$ for each recipe variation by comparison with observed galaxies in the DustPedia data base (Davies et al. 2017). We construct mutually matched samples including the 38 star-forming ARTEMIS galaxies on the one hand and 42 DustPedia galaxies in the same stellar mass and SFR range on the other hand. We then compare these samples through luminosity scaling relations in wavelength bands from UV to submm (Fig. 2). The resulting optimal $f_{\rm dust}$ values for each of our dust allocation recipes are listed in Table 1. We furthermore conclude that the $^{\rm s}$K21/$^{\rm d}$T12 recipe optimally reproduces the observed luminosity scaling relations, confirming the findings by Kapoor et al. (2021) for the simulated Auriga galaxies, albeit with a different value for $f_{\rm dust}$ (see Table 1).

Even with the optimal recipe, dust extinction is significantly underestimated at FUV/UV wavelengths and representative dust temperatures are lower than those observed. We attribute these discrepancies to limitations in the treatment of SF regions and their immediate environment, similar to the findings in previous studies (e.g. Camps et al. 2016; Trčka et al. 2020; Kapoor et al. 2021). These symptoms therefore seem to be a characteristic of all state-of-the-art cosmological simulation UV-submm post-processing efforts using similar SF region recipes. Resolving these issues will likely require improvements both in the modelling of the cold interstellar medium in hydrodynamical simulations of galaxy evolution and in the subgrid treatment of SF regions in the RT post-processing procedure. Concerning the latter, a crucial step is to develop an enhanced SF region model that is designed specifically for incorporation in RT simulations. Ideally, key characteristics of the model such as the SSP or dust grain properties should be configurable, and the parameters of the resulting SED template library should be easily derivable from the particle properties in the hydrodynamical simulation. More fundamentally, the model should, on average, allow more UV radiation to escape into the diffuse ISM without adversely affecting the shape of the SED at other wavelengths.

Using the optimal $^{\rm s}$K21/$^{\rm d}$T12 recipe, we produce images of all ARTEMIS galaxies at 50 pc resolution for 50 commonly used broadband filters from UV to submm wavelengths and for 18 different viewing angles. We spatially integrate these images to obtain global fluxes and use the SED fitting code CIGALE (Noll et al. 2009; Boquien et al. 2019) to derive physical galaxy properties. The inferred properties recover the known intrinsic values fairly well, except that stellar mass is often significantly underestimated for near-edge-on configurations (Figs 5 and 6). We argue that this discrepancy is related to the stronger optical dust attenuation at high inclinations (Fig. 7), which disturbs the energy balance for those sightlines, in turn confusing the SED fitting algorithm.

We use selected dust scaling relations (Fig. 9) to compare the inferred ARTEMIS galaxy properties to similarly derived properties of the DustPedia galaxies and of the simulated Auriga galaxies (Kapoor et al. 2021). We find that the ARTEMIS galaxies tend to contain more dust than comparable DustPedia galaxies, but otherwise follow the observed dust scaling relations very well. The Auriga galaxies follow the same relations but often occupy an adjacent region of the parameter space.

We subsequently use the high-resolution images at multiple wavelengths to perform a basic morphological study of the 11 ARTEMIS galaxies with a disc-to-total stellar mass ratio of D/T > 0.45. We use the STATMORPH package (Rodriguez-Gomez et al. 2019) to calculate four non-parametric morphological properties as a function of wavelength. We compare these results (Fig. 10) to

similarly derived values for 9 well-resolved spiral galaxies in the DustPedia data base and 14 simulated Auriga galaxies. We find that the ARTEMIS galaxies largely reproduce the observed trends as a function of wavelength, except that they appear to be more clumpy and less symmetrical than observed. We also highlight some differences between the ARTEMIS and Auriga data sets.

Kapoor et al. (2021) cite various types of studies of the dust-related properties of simulated Milky Way-like galaxies at redshift zero that are enabled by the availability of dust-aware high resolution images of these galaxies at multiple wavelengths. Their examples include spatially resolved SED fitting, local energy balance studies, spatially resolved dust scaling relations, dust mass maps, and the contribution to dust heating by different stellar components in various regions of a galaxy.

We similarly invite any interested party to use our ARTEMIS results in such studies (see Section 3.4). In fact, the Auriga and ARTEMIS galaxies occupy adjacent regions in the sSFR versus stellar mass plane (Fig. 8) and hence also in the dust scaling relations (bottom row of Fig. 9). This means that the data products resulting from this work are supplemental to those produced by Kapoor et al. (2021) for Auriga. Both data sets are publicly available, and because the same post-processing recipe has been used in both cases, they can be combined to achieve a wider coverage of galaxy types and/or to increase statistical significance.

## ACKNOWLEDGEMENTS

## DATA AVAILABILITY

Section 3.4 describes the public data products provided as a result of this work, available for download at https://www.astro.ljmu.ac.uk/Artemis.

## REFERENCES

Baes M. et al., 2003, MNRAS, 343, 1081
Baes M. et al., 2020, A&A, 641, A119
Baes M., Camps P., 2015, Astron. Comput., 12, 33
Baes M., Verstappen J., De Looze I., Fritz J., Saftly W., Vidal Pérez E., Stalevski M., Valcke S., 2011, ApJS, 196, 22
Baes M., Trčka A., Camps P., Nersesian A., Trayford J., Theuns T., Dobbels W., 2019, MNRAS, 484, 4069
Barber C., Crain R. A., Schaye J., 2018, MNRAS, 479, 5448
Behrens C., Pallottini A., Ferrara A., Gallerani S., Vallini L., 2018, MNRAS, 477, 552
Bianchi S. et al., 2018, A&A, 620, A112
Bignone L. A., Pellizza L. J., Tissera P. B., 2016, Boletin de la Asociacion Argentina de Astronomia La Plata Argentina, 58, 30
Boquien M., Burgarella D., Roehlly Y., Buat V., Ciesla L., Corre D., Inoue A. K., Salas H., 2019, A&A, 622, A103
Bruzual G., Charlot S., 2003, MNRAS, 344, 1000
Calzetti D., Armus L., Bohlin R. C., Kinney A. L., Koornneef J., Storchi-Bergmann T., 2000, ApJ, 533, 682
Camps P. et al., 2015, A&A, 580, A87
Camps P. et al., 2018, ApJS, 234, 20
Camps P., Baes M., 2015, Astron. Comput., 9, 20
Camps P., Baes M., 2020, Astron. Comput., 31, 100381
Camps P., Baes M., Saftly W., 2013, A&A, 560, A35
Camps P., Trayford J. W., Baes M., Theuns T., Schaller M., Schaye J., 2016, MNRAS, 462, 1057
Casasola V. et al., 2020, A&A, 633, A100
Chabrier G., 2003, PASP, 115, 763
Ciesla L. et al., 2016, A&A, 585, A43
Clark C. J. R. et al., 2018, A&A, 609, A37
Cochrane R. K. et al., 2019, MNRAS, 488, 1779
Conselice C. J., 2003, ApJS, 147, 1
Courteau S. et al., 2014, Rev. Mod. Phys., 86, 47
Crain R. A. et al., 2015, MNRAS, 450, 1937
da Cunha E., Charlot S., Elbaz D., 2008, MNRAS, 388, 1595
Dale D. A. et al., 2017, ApJ, 837, 90
Davé R., Anglés-Alcázar D., Narayanan D., Li Q., Rafieferantsoa M. H., Appleby S., 2019, MNRAS, 486, 2827
Davies J. I. et al., 2017, PASP, 129, 044102
De Geyter G., Baes M., Camps P., Fritz J., De Looze I., Hughes T. M., Viaene S., Gentile G., 2014, MNRAS, 441, 869
De Looze I. et al., 2014, A&A, 571, A69
Domínguez-Tenreiro R., Obreja A., Granato G. L., Schurer A., Alpresa P., Silva L., Brook C. B., Serna A., 2014, MNRAS, 439, 3868
Draine B. T., Li A., 2001, ApJ, 551, 807
Dudzevičiūtė U. et al., 2020, MNRAS, 494, 3828
Fasano G., Franceschini A., 1987, MNRAS, 225, 155
Font A. S. et al., 2020, MNRAS, 498, 1765
Font A. S., McCarthy I. G., Belokurov V., 2021, MNRAS, 505, 783
Gadotti D. A., Baes M., Falony S., 2010, MNRAS, 403, 2053
Gjergo E., Granato G. L., Murante G., Ragone-Figueroa C., Tornatore L., Borgani S., 2018, MNRAS, 479, 2588
Granato G. L. et al., 2021, MNRAS, 503, 511
Granato G. L., Ragone-Figueroa C., Domínguez-Tenreiro R., Obreja A., Borgani S., De Lucia G., Murante G., 2015, MNRAS, 450, 1320
Grand R. J. J. et al., 2017, MNRAS, 467, 179
Groves B., Dopita M. A., Sutherland R. S., Kewley L. J., Fischera J., Leitherer C., Brandl B., van Breugel W., 2008, ApJS, 176, 438
Hahn O., Abel T., 2011, MNRAS, 415, 2101
Hernquist L., Springel V., 2003, MNRAS, 341, 1253
Hinshaw G. et al., 2013, ApJS, 208, 19
Hunt L. K. et al., 2019, A&A, 621, A51
Jenkins A., 2013, MNRAS, 434, 2094
Jones A. P., Köhler M., Ysard N., Bocchio M., Verstraete L., 2017, A&A, 602, A46
Jonsson P., Groves B. A., Cox T. J., 2010, MNRAS, 403, 17
Kannan R., Marinacci F., Vogelsberger M., Sales L. V., Torrey P., Springel V., Hernquist L., 2020, MNRAS, 499, 5732
Kapoor A. U. et al., 2021, MNRAS, 506, 5703
Kennicutt R. C. et al., 2011, PASP, 123, 1347
Kennicutt R. C., Evans N. J., 2012, ARA&A, 50, 531
Khandai N., Di Matteo T., Croft R., Wilkins S., Feng Y., Tucker E., DeGraf C., Liu M.-S., 2015, MNRAS, 450, 1349
Kolmogorov A. N., 1933, Giornale dell'Istituto Italiano degli Attuari, 4, 83
Lahén N., Johansson P. H., Rantala A., Naab T., Frigo M., 2018, MNRAS, 475, 3934
Lanz L., Hayward C. C., Zezas A., Smith H. A., Ashby M. L. N., Brassington N., Fazio G. G., Hernquist L., 2014, ApJ, 785, 39
Lee D., Baes M., Seon K.-I., Camps P., Verstocken S., Han W., 2016, MNRAS, 463, 2912
Li A., Draine B. T., 2001, ApJ, 554, 778
Liang L. et al., 2019, MNRAS, 489, 1397

Liang L., Feldmann R., Faucher-Giguère C.-A., Kereš D., Hopkins P. F., Hayward C. C., Quataert E., Scoville N. Z., 2018, MNRAS, 478, L83

Lotz J. M., Primack J., Madau P., 2004, AJ, 128, 163

Lovell C. C., Geach J. E., Davé R., Narayanan D., Li Q., 2021, MNRAS, 502, 772

Ma X. et al., 2019, MNRAS, 487, 1844

Małek K. et al., 2018, A&A, 620, A50

Marinacci F., Vogelsberger M., Mocz P., Pakmor R., 2015, MNRAS, 453, 3999

Mathis J. S., Mezger P. G., Panagia N., 1983, A&A, 500, 259

Narayanan D. et al., 2021, ApJS, 252, 12

Narayanan D., Davé R., Johnson B. D., Thompson R., Conroy C., Geach J., 2018, MNRAS, 474, 1718

Nelson D. et al., 2019, MNRAS, 490, 3234

Nersesian A. et al., 2019, A&A, 624, A80

Nersesian A. et al., 2020a, A&A, 637, A25

Nersesian A. et al., 2020b, A&A, 643, A90

Noll S., Burgarella D., Giovannoli E., Buat V., Marcillac D., Muñoz-Mateos J. C., 2009, A&A, 507, 1793

Parsotan T., Cochrane R. K., Hayward C. C., Anglés-Alcázar D., Feldmann R., Faucher-Giguère C. A., Wellons S., Hopkins P. F., 2021, MNRAS, 501, 1591

Peacock J. A., 1983, MNRAS, 202, 615

Peest C., Camps P., Stalevski M., Baes M., Siebenmorgen R., 2017, A&A, 601, A92

Pillepich A. et al., 2019, MNRAS, 490, 3196

Planck Collaboration et al., 2014, A&A, 571, A16

Press W. H., Teukolsky S. A., Vetterling W. T., Flannery B. P., 2002, Numerical Recipes in C++: The Art of Scientific Computing, 3rd edn. Cambridge University Press

Robitaille T. P., 2011, A&A, 536, A79

Rodriguez-Gomez V. et al., 2019, MNRAS, 483, 4140

Saftly W., Camps P., Baes M., Gordon K. D., Vand ewoude S., Rahimi A., Stalevski M., 2013, A&A, 554, A10

Saftly W., Baes M., Camps P., 2014, A&A, 561, A77

Saftly W., Baes M., De Geyter G., Camps P., Renaud F., Guedes J., De Looze I., 2015, A&A, 576, A31

Santos-Santos I. M. E., Domínguez-Tenreiro R., Granato G. L., Brook C. B., Obreja A., 2017, A&A, 603, A4

Sawala T. et al., 2016, MNRAS, 457, 1931

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

Silva L., Granato G. L., Bressan A., Danese L., 1998, ApJ, 509, 103

Smirnov N., 1948, Ann. Math. Statist., 19, 279

Springel V., 2005, MNRAS, 364, 1105

Springel V., 2010, MNRAS, 401, 791

Torrey P. et al., 2015, MNRAS, 447, 2753

Trayford J. W. et al., 2017, MNRAS, 470, 771

Trayford J. W., Schaye J., 2019, MNRAS, 485, 5715

Trayford J. W., Lagos C. d. P., Robotham A. S. G., Obreschkow D., 2020, MNRAS, 491, 3937

Trčka A. et al., 2020, MNRAS, 494, 2823

Trčka A., Baes M., Camps P., Kapoor A. U., Nelson D., Pillepich A., Barrientos D., 2021, MNRAS, submitted

Tremmel M. et al., 2019, MNRAS, 483, 3336

Tremmel M., Karcher M., Governato F., Volonteri M., Quinn T. R., Pontzen A., Anderson L., Bellovary J., 2017, MNRAS, 470, 1121

Utomo D., Chiang I.-D., Leroy A. K., Sandstrom K. M., Chastenet J., 2019, ApJ, 874, 141

van de Voort F., Bieri R., Pakmor R., Gómez F. A., Grand R. J. J., Marinacci F., 2021, MNRAS, 501, 4888

Verstocken S. et al., 2020, A&A, 637, A24

Verstocken S., Van De Putte D., Camps P., Baes M., 2017, Astron. Comput., 20, 16

Viaene S. et al., 2016, A&A, 586, A13

Viaene S. et al., 2020, A&A, 638, A150

Vogelsberger M. et al., 2020b, MNRAS, 492, 5167

Vogelsberger M., Marinacci F., Torrey P., Puchwein E., 2020a, Nat. Rev. Phys., 2, 42

Wang L., Dutton A. A., Stinson G. S., Macciò A. V., Penzo C., Kang X., Keller B. W., Wadsley J., 2015, MNRAS, 454, 83

Wetzel A. R., Hopkins P. F., Kim J.-h., Faucher-Giguère C.-A., Kereš D., Quataert E., 2016, ApJ, 827, L23

Zubko V., Dwek E., Arendt R. G., 2004, ApJS, 152, 211

## APPENDIX A: COMPARING DUST MODELS

As stated in Section 3.2.1, we use a more recent dust model for representing the diffuse dust in our RT post-processing procedure than the dust model employed by Camps et al. (2016), Trayford et al. (2017), and Camps et al. (2018) for producing synthetic observables for the EAGLE galaxies. In this appendix, we study the effects of this new dust model on the scaling relations shown in Fig. 2, which we use to calibrate and compare variations of the recipes in Section 3.3.

The earlier EAGLE work used the dust model presented by Zubko et al. (2004), called Zubko in this appendix. This model includes a mixture of non-composite graphite and silicate grains and neutral and ionized polycyclic aromatic hydrocarbon (PAH) molecules, designed so that the global dust properties reproduce the extinction, emission, and abundance constraints of the Milky Way. The optical and calorimetric properties follow the prescriptions of Draine & Li (2001) and Li & Draine (2001). In this work, we use the THEMIS dust model described by Jones et al. (2017) and references therein. This model was developed in the context of the DustPedia project to explain the dust extinction and emission in the diffuse interstellar medium, and to self-consistently include the effects of dust evolution in the transition to denser regions. It includes a mixture of amorphous hydrocarbons and amorphous silicates. For the latter, it is assumed that half of the mass is amorphous enstatite and the remaining half is amorphous forsterite.

Fig. A1 shows the ratio of the extinction and absorption mass coefficients for the THEMIS dust model over those for the Zubko dust model in the UV-optical wavelength range. Fig. A2 compares the emissivity of a THEMIS and Zubko dust grain population of the same mass in response to an interstellar radiation field of varying strength. In each figure, the bottom panel shows the transmission curves for the broad-bands used in our scaling relations. From the first figure, we conclude that a THEMIS dust grain population exhibits 10–25 per cent more extinction than a Zubko grain population in the UV and optical bands under consideration. In the second figure, it is immediately obvious that the aromatic features, on average,
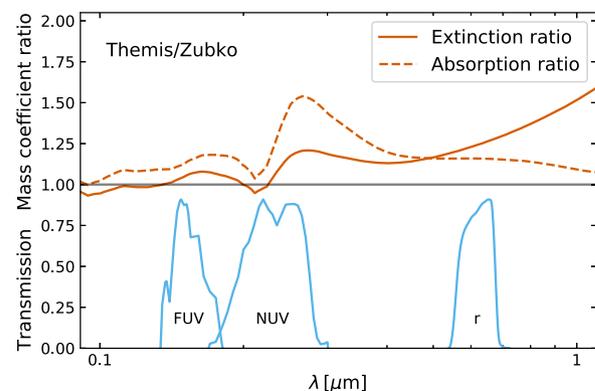


**Figure A1.** Ratio of the total extinction (solid red) and absorption (dashed red) mass coefficients for the THEMIS (Jones et al. 2017) dust model over those for the Zubko et al. (2004) dust model in the UV-optical wavelength range. The bottom panel shows the transmission curves (blue) for the broad-bands used in the scaling relations of Fig. A3.
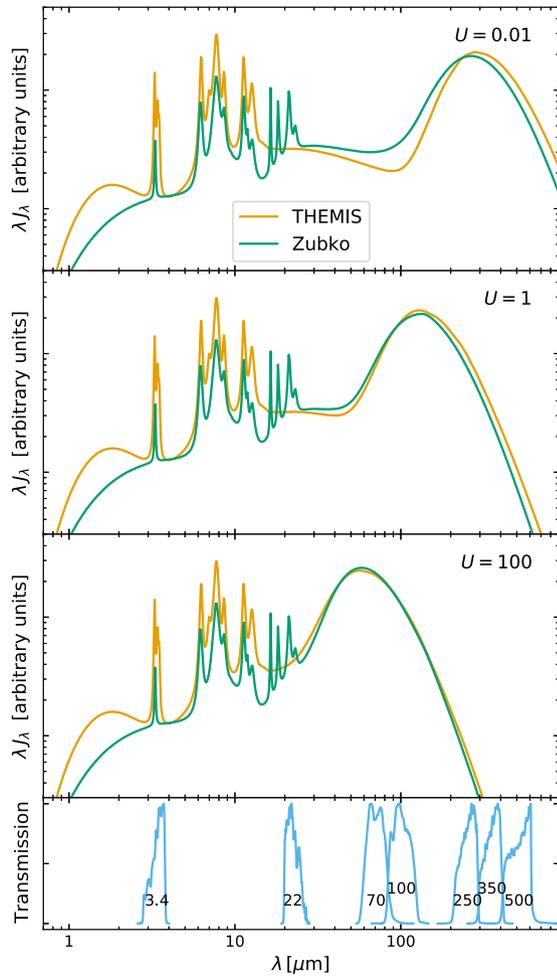
**Figure A2.** Emissivity of a dust grain population with properties of the THEMIS (Jones et al. 2017, orange) and Zubko et al. (2004, green) dust models in response to a typical interstellar radiation field (Mathis, Mezger & Panagia 1983) with various strengths $U = 0.01, 1, 100$. Details on the emission calculation, including non-equilibrium heating of smaller dust grains, are provided by Camps et al. (2015) and references therein. The bottom panel shows the transmission curves (blue) for the broad-bands used in the scaling relations of Fig. A3.

are more luminous for the THEMIS dust than for the Zubko dust. This significantly boosts the 3.4 μm band luminosity, while the 22 μm band luminosity is somewhat tempered. Furthermore, the dust continuum emission peak for weaker input fields shifts slightly to longer wavelengths for the THEMIS model, corresponding to lower representative dust temperatures. Also, the slopes on both sides of the continuum emission peak differ between the dust models.

With this information, we can interpret Fig. A3, which shows the scaling relations first presented in Fig. 2 for each of the two dust models. The top and middle row panels all have the 3.4 μm band luminosity on the horizontal axis. The THEMIS data points in these panels are therefore shifted to the right in accordance with the extra aromatic feature emission modelled in this band (by up to 0.15 dex for the most dust-luminous galaxy). The top row panels have a UV or optical band on the vertical axis, causing the THEMIS data points to shift downward reflecting the increased dust attenuation in that model. The combined result in these panels is a diagonal shift more or less orthogonal to the scaling relation. As discussed in Section 3.3.3, our fiducial recipe underestimates attenuation at UV wavelengths. We see here that the THEMIS dust mix helps decreasing this discrepancy. Unfortunately, because its extinction coefficient increases more distinctly for optical wavelengths than for UV wavelengths (see Fig. A1), the THEMIS model instead overestimates the attenuation in the optical regime.

The THEMIS 22 μm data points in panel (d) are also shifted downward, resulting in a diagonal shift similar to that in the UV and optical regimes, but now caused by the decreased aromatic feature emission in this band. The 250 and 500 μm bands are on the downward slope of the dust continuum peak and thus both see enhanced emission for the THEMIS dust model (see Fig. A2). As a result, the THEMIS data points in panels (e) and (f) shift essentially along the observed scaling relation. The shifts in the colour–colour relations on the bottom row can be similarly interpreted. Notably, panel (i) clearly shows a lower representative dust temperature for the THEMIS dust mix (the dust temperature rises diagonally to the upper right in this panel; see e.g. fig. 11 of Camps et al. 2016).

In summary, compared to the Zubko dust model, and for an otherwise fixed recipe, the THEMIS dust model reduces the discrepancies between our simulations and the DustPedia observations in some wavelength regimes but introduces extra tension in other regimes.
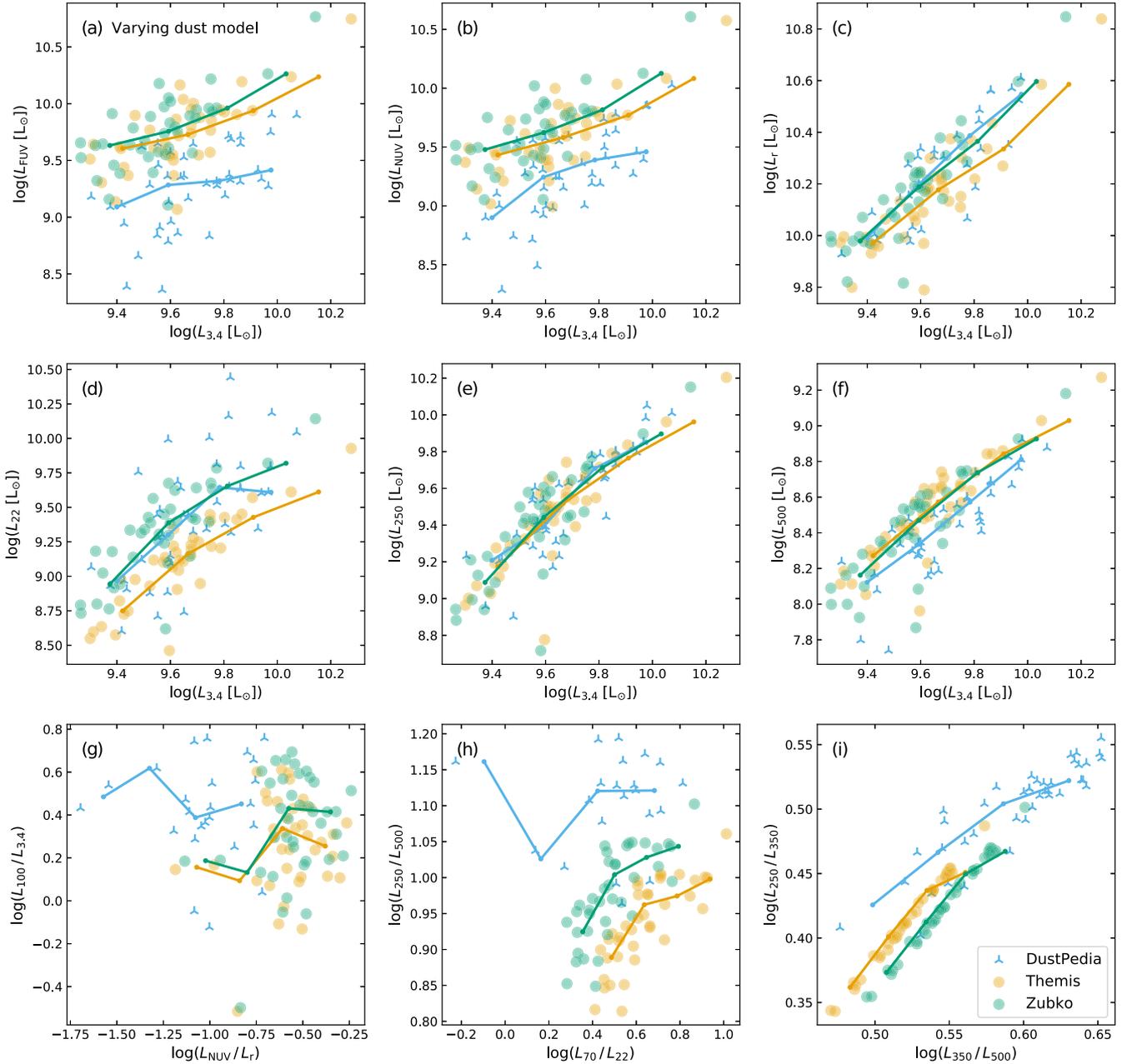
**Figure A3.** The same scaling relations as in Fig. 2, now showing ARTEMIS luminosities calculated for a random viewing angle using the $^sC16/^dC16$ recipe at optimal dust fraction $f_{dust}$ (see Table 1) with THEMIS (Jones et al. 2017, orange) and Zubko et al. (2004, green) dust models in comparison with observed DustPedia luminosities (blue).

## APPENDIX B: CALIBRATING DUST FRACTIONS

As described in Section 3.2.3, the recipes for handling diffuse dust in our RT simulations have a free parameter, the dust-to-metal fraction $f_{dust}$, which must be determined through comparison with observations. We have chosen the luminosity scaling relations shown in Fig. 2 to accomplish this comparison, because they trace the key physical galaxy properties including stellar mass, SFR, sSFR, dust mass, and dust temperature. In principle, selecting an 'optimal' $f_{dust}$ value is a straightforward optimization process. In practice, however, it is substantially complicated by several factors.

The numerical error on the ARTEMIS luminosities caused by discretization effects in the RT simulation is below 8 per cent or 0.033 dex (see Section 3.3.2). The calibration error on the DustPedia fluxes for the broad-bands used in our scaling relations is of the same order (see Table 1 of Clark et al. 2018), although these numbers probably do not capture all observational uncertainties. For our purposes, we can assume that these variations constitute a minor factor.

Another factor is the effect of the sight line on the observed luminosities. While the DustPedia galaxies are obviously observed at some fixed sight line, we can control the viewing angle for our simulated ARTEMIS galaxies. The effect on our scaling relations is shown in Fig. B1. As expected, there is a significant discrepancy between the face-on and edge-on luminosities at shorter wavelengths,
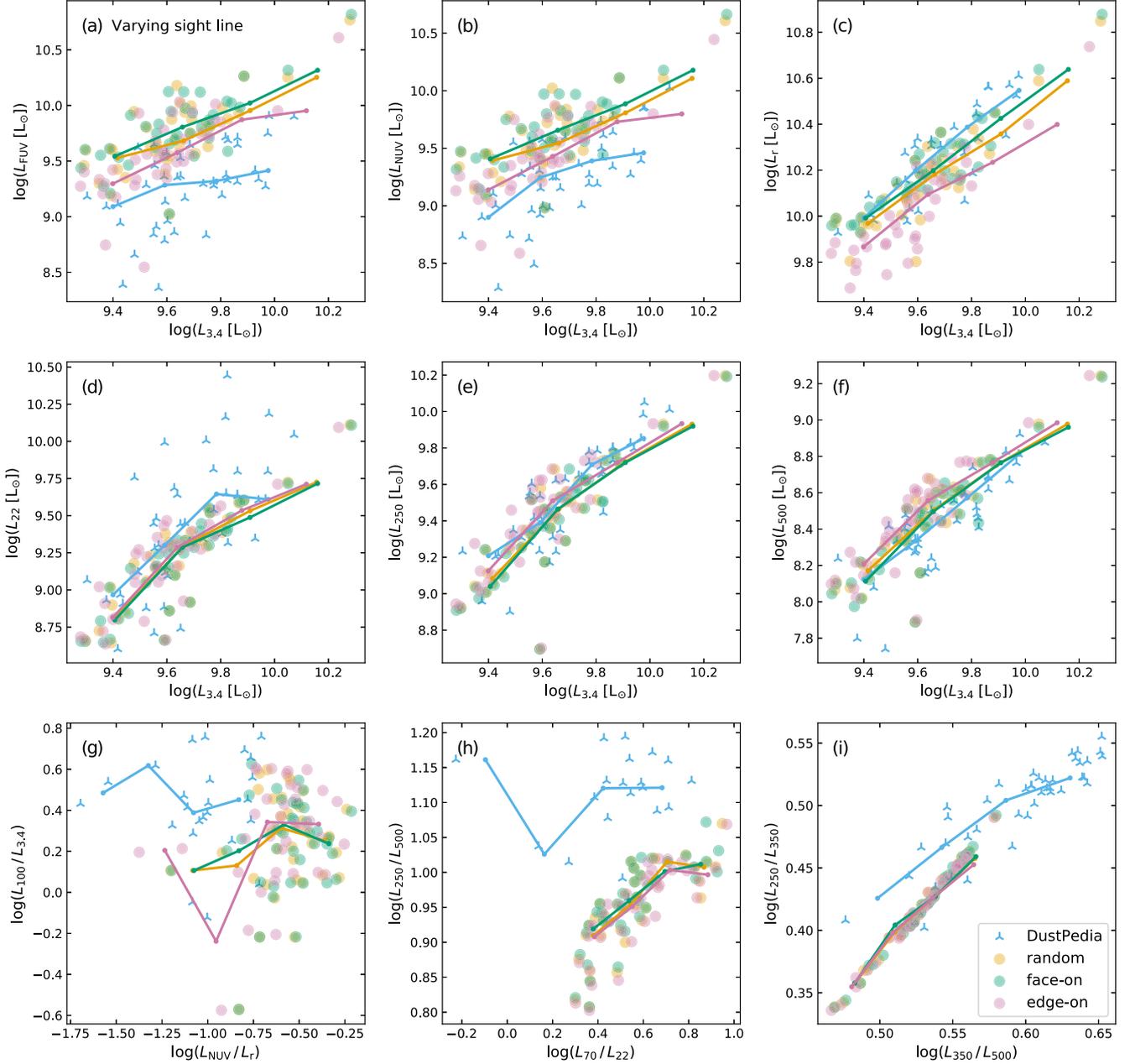
**Figure B1.** The same scaling relations as in Fig. 2, now showing ARTEMIS luminosities calculated for three different sightlines (random – orange, face-on – green, edge-on – purple) using the $^{\rm s}$K21/$^{\rm d}$T12 recipe at optimal dust fraction $f_{\rm dust}$ (see Table 1) in comparison with observed DustPedia luminosities (blue).

up to 0.37 dex in the *r* band and up to 0.48 dex in the FUV band, while the effect is minimal at longer wavelengths. We mitigate the sight line factor by using a random viewing angle for the ARTEMIS galaxies in our calibration process, which probably corresponds most closely to the observed data set.

We need to quantify how well a given set of ARTEMIS data points corresponds to the observed DustPedia data points in the scaling relations. Following Camps et al. (2016) and Kapoor et al. (2021), we employ a generalization of the Kolmogorov–Smirnov test (K–S test, Kolmogorov 1933; Smirnov 1948) to two-dimensional distributions (Peacock 1983; Fasano & Franceschini 1987; Press et al. 2002). The 2D K–S test computes a metric which can be interpreted as a measure of the 'distance' between two sets of two-dimensional data points. The metric may not be perfectly suited for our purposes because, for

example, a shift away from the scaling relation is generally penalised in the same way as a more acceptable shift along the scaling relation. More importantly, to obtain a single overall measure for a given recipe, the metric for each of the relations must be aggregated. The final measure, and thus the ranking of different recipes, depends on the selection of scaling relations included in the metric and on the relative weights assigned to them.

As an example that is representative for the three recipes described in Section 3.2.4, Fig. B2 shows our scaling relations for results using the $^{\rm s}$K21/$^{\rm d}$T12 recipe with different dust fraction values: $f_{\rm dust} = 0.250$, 0.275, and 0.300. The shifts in the scaling relations are much smaller than those shown for different sightlines in Fig. B1. Also, the shift can be toward or away from the observed DustPedia relation depending on the wavelength regime. For example, increasing the dust fraction
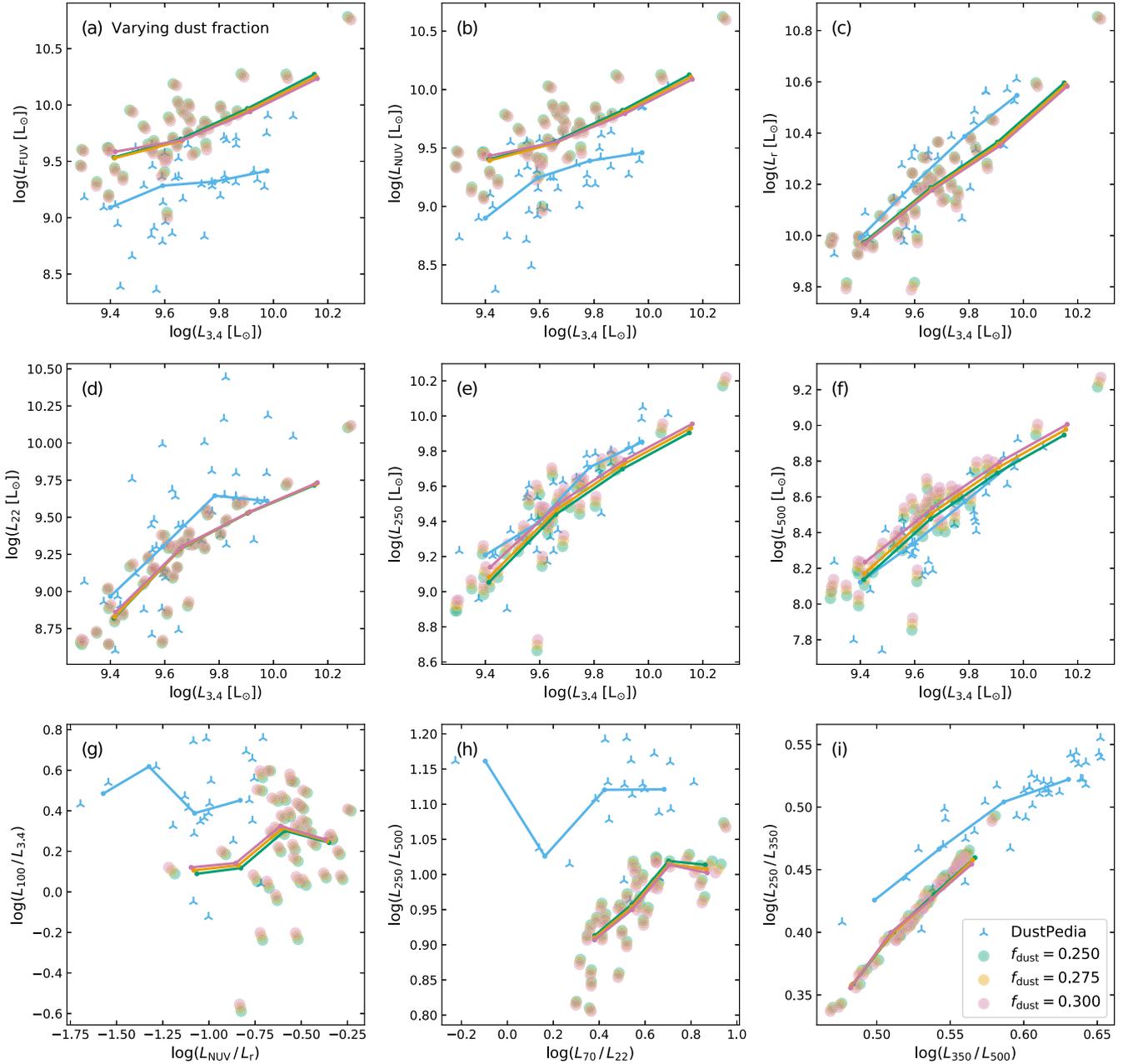
**Figure B2.** The same scaling relations as in Fig. 2, now showing ARTEMIS luminosities calculated for a random viewing angle using the $^{s}$K21/$^{d}$T12 recipe with three values for the dust fraction $f_{dust} = 0.250$ (green), 0.275 (orange), and 0.300 (purple) near or at the optimal value (see Table 1) in comparison with observed DustPedia luminosities (blue).

and thus the attenuation at shorter wavelengths improves the tension with observations in the UV (panels a and b) but worsens it in the optical (panel c). Similarly, a shift towards colder dust temperatures (panel i) causes opposing shifts in the infrared relations (e.g. panel e versus panel f). Any aggregated metric based on these relations therefore depends significantly on the employed relative weights.

We experimented with various averaging schemes, always mixing results in UV, optical, and infrared wavelength regimes, and came to the following conclusions (see Table 1). For the $^{s}$C16/$^{d}$C16 recipe, $f_{dust} = 0.300$ is a robust optimal value, i.e. it comes out on top regardless of the averaging scheme. This corresponds to the optimal value found by Camps et al. (2016) for the EAGLE simulations, which is comforting because both the hydrodynamical simulation

and RT post-processing recipes are very similar. By the same token, for the $^{s}$C16/$^{d}$T12 recipe, $f_{dust} = 0.275$ is a robust optimal value. The decrease in dust-to-metal ratio compared to the $^{s}$C16/$^{d}$C16 recipe can be understood by noting that the $^{d}$T12 scheme assigns dust to a broader set of gas particles than the $^{d}$C16 scheme (see Section 3.2.3). For the ARTEMIS galaxies, this results in a 'dusty' ISM mass that is 5 to 15 per cent higher. This increase is roughly compensated by the 9 per cent decrease in dust fraction. We do note, however, that the emitted radiation will also vary between the two recipes because of the changed relative dust/star geometry.

For the $^{s}$K21/$^{d}$T12 recipe, the values $f_{dust} = 0.275$ and 0.300 result in very comparable statistics so that the 'optimal' value sensitively depends on the chosen averaging scheme. We take this to mean

that the actual optimal value lies between 0.275 and 0.300. This slight increase from the $^s$C16/$^d$T12 recipe must be related to the different handling of SF regions, which affects the dust-related emission modelled as part of the MAPPINGS III SED templates (see Sections 3.2.1 and 3.2.2). We choose to employ $f_{dust} = 0.275$ as the 'optimal' value because the differences are small (see Fig. B2) and we then have a consistent value for the $^d$T12 dust allocation scheme regardless of the recipe for handling SF regions. Also, Kapoor et al. (2021) find a fairly large difference in optimal dust fraction between the two dust allocation schemes (see Table 1).

This paper has been typeset from a TeX/LaTeX file prepared by the author.