

# Constraining the Milky Way’s ultraviolet-to-infrared SED with Gaussian process regression

Catherine E. Fielder<sup>1,2</sup>★, Jeffrey A. Newman<sup>1,2</sup>, Brett H. Andrews<sup>1,2</sup>, Gail Zasowski<sup>3</sup>,  
Nicholas F. Boardman<sup>3</sup>, Tim Licquia<sup>1,4</sup>, Karen L. Masters<sup>5</sup> and Samir Salim<sup>6</sup>

<sup>1</sup>Department of Physics and Astronomy, University of Pittsburgh, Pittsburgh, PA 15260, USA

<sup>2</sup>Pittsburgh Particle Physics, Astrophysics, and Cosmology Center (PITT PACC), University of Pittsburgh, Pittsburgh, PA 15260, USA

<sup>3</sup>Department of Physics and Astronomy, University of Utah, Salt Lake City, UT 84112, USA

<sup>4</sup>Core R&D: Information Research, Dow Inc., 633 Washington Street, 1776 Building, Midland, MI 48667, USA

<sup>5</sup>Departments of Physics and Astronomy, Haverford College, 370 Lancaster Avenue, Haverford, PA 19041, USA

<sup>6</sup>Department of Astronomy, Indiana University, Bloomington, IN 47405, USA

Accepted 2021 September 9. Received 2021 September 1; in original form 2021 June 29

## ABSTRACT

Improving our knowledge of global Milky Way (MW) properties is critical for connecting the detailed measurements only possible from within our Galaxy to our understanding of the broader galaxy population. We here train Gaussian process regression (GPR) models on SDSS (Sloan Digital Sky Survey) galaxies to map from galaxy properties (stellar mass, apparent axial ratio, star formation rate, bulge-to-total ratio, disc scale length, and bar vote fraction) to ultraviolet (UV; *GALEX FUV/NUV*), optical (SDSS *ugriz*), and infrared (IR; 2MASS *JHKs* and *WISE W1/W2/W3/W4*) fluxes and uncertainties. With these models, we estimate the photometric properties of the MW, resulting in a full UV-to-IR spectral energy distribution (SED) as it would be measured externally, viewed face-on. We confirm that the MW lies in the green valley in optical diagnostic diagrams, but show for the first time that the MW is in the star-forming region in standard UV and IR diagnostics – characteristic of the population of red spiral galaxies. Although our GPR method predicts one band at a time, the resulting MW UV–IR SED is consistent with SEDs of local spirals with characteristics broadly similar to the MW, suggesting that these independent predictions can be combined reliably. Our UV–IR SED will be invaluable for reconstructing the MW’s star formation history using the same tools employed for external galaxies, allowing comparisons of results from *in situ* measurements to those from the methods used for extragalactic objects.

**Key words:** Galaxy: fundamental parameters – Galaxy: general – Galaxy: structure.

## 1 INTRODUCTION

Within the Milky Way (MW), we have a unique opportunity to study the nuances of galactic properties, allowing us to test galaxy formation and evolution models at an unrivalled level of detail. For example, chemical abundances for hundreds of thousands of stars have been obtained from spectroscopic surveys (Majewski et al. 2017; Martell et al. 2017), and stellar surveys that catalogue distance and dynamical measurements on millions of stars have been performed (Gaia Collaboration 2018). This exhaustive stellar information has helped constrain the MW’s evolutionary history. In turn, high-resolution dynamical simulations have been able to produce galaxies of increased similarity to the MW (Guedes et al. 2011; Sawala et al. 2016; Wetzel et al. 2016), matching fundamental galaxy properties such as dwarf satellite populations and reproducing characteristics of our Galaxy’s gas, dust, and stellar components. Comparisons between MW stellar data and high-resolution hydrodynamical simulations of MW-like galaxies are an increasingly useful way to improve our understanding of galaxy formation.

However, the marriage between observations and models is delicate: Incorrect assumptions on one side can propagate into the other. Simulators must make choices about how to implement crucial parameters that affect the galaxy evolution process, such as the gas density threshold for star formation to occur and the efficiency with which it proceeds; this is sometimes done by attempting to match observed properties of the MW. However, without knowing how our Galaxy fits in among the broader galaxy population, it is difficult to determine whether simulations match the MW because they have the correct physics or because they have incorrectly tuned parameters that match by coincidence or design.

This is complicated by the exceptional difficulty of obtaining a global picture of the MW, given our location in the disc and the obscuration caused by interstellar dust. As a result, there are properties that we can easily measure in external galaxies that are impossible to measure directly within our own, making it difficult to determine where we fit within the broader galaxy population.

Creating an outside-in picture of the MW that spans a multitude of broad-band wavelengths will enable simulators to more accurately tune their physics assumptions, as it will then be possible to test whether quantities that can only be determined from large-scale stellar surveys and those that can be measured directly only for

\* E-mail: [cef41@pitt.edu](mailto:cef41@pitt.edu)

extragalactic objects are reproduced. The most basic, easiest-to-measure intrinsic quantities we can use to study galaxies are their luminosities and colours; once redshift is known, these can be inferred from broad-band photometry. Hence, the focus of this paper will be in determining these properties for the MW. This will enable our Galaxy to be placed on standard colour–magnitude and colour–colour diagrams and result in a multiwavelength spectral energy distribution (SED) for the MW.

Astronomers have found that galaxies in the local Universe predominantly fall into two populations: passively evolving red galaxies with older stellar populations or blue galaxies that are still forming stars. In the optical colour–magnitude diagram (CMD), these two galaxy populations are commonly referred to as the ‘red sequence’ and the ‘blue cloud’. The colour bimodality of galaxies has been observed at both low redshift ( $z \sim 0.1$ ; e.g. Strateva et al. 2001; Baldry et al. 2004) and up to a redshift of  $z \sim 1$  (e.g. Bell et al. 2004; Weiner et al. 2005). The region of the CMD between these two distinct galaxy populations is often referred to as the ‘green valley’. This locus is thought to contain a transitional population of galaxies that are ‘passively’ evolving in the sense that no new star formation is occurring (e.g. Bell et al. 2004; Faber et al. 2007), though they still can contain some younger stars. The increase in the fraction of red galaxies over time has led many astronomers to conclude that a galaxy first lives in the star-forming blue cloud and then transitions into the green valley and ultimately into the red sequence in complete quiescence, with the galaxy growing more and more red over time due to the ageing stellar population. Green valley galaxies are presumed to be undergoing some form of quenching of their star formation (Salim et al. 2007; Schawinski et al. 2014; Smethurst et al. 2015) – either late-type galaxies that are gradually running out of their cold gas reservoir or having their star formation suppressed, or early-type galaxies that had their gas reservoirs rapidly destroyed. Objects found within the green valley may also simply be in the tails of the blue cloud or red sequence, rather than being in a transitional state (Taylor et al. 2015). While the precise details of quenching processes and the origin of the galaxy colour bimodality have yet to be determined, a galaxy’s location on the CMD remains a very useful tool for determining how a galaxy fits into the broader population.

The radiation emitted by a galaxy is characterized by its SED, or flux as a function of wavelength. Galaxy SEDs contain the imprints of the physical processes occurring within – the stellar population’s ages and abundances (i.e. the star formation history and metallicity of the galaxy), the dust and gas content, and the chemistry and physical state of the interstellar medium (ISM). Because different sources dominate the emission at different wavelengths, long-wavelength-baseline SEDs allow one to disentangle the contributing effects. This makes SEDs one of the best direct probes for studying galaxy formation and evolution from both an observational and theoretical modelling perspective.

However, comparing colours and luminosities of the MW to external galaxies is not trivial, regardless of whether we compare to observed galaxies or to mock images from high-resolution hydrodynamical simulations (such as Eris, Guedes et al. 2011; APOSTLE, Sawala et al. 2016; and Latte, Wetzel et al. 2016). Much of our view of the Galaxy is obscured by interstellar dust, especially at ultraviolet (UV) and optical wavelengths (e.g. Cardelli, Clayton & Mathis 1989; Schlegel, Finkbeiner & Davis 1998). Stars outside of the local solar region are reddened as a result of the dust obscuration. Determining the integrated light of stellar populations in the MW is challenging due to the spread of stars over large and varying distances, with correspondingly large and varying dust extinction along lines of sight to the Earth. This makes the study of any portions

of the Galactic disc beyond the solar neighbourhood exceptionally difficult, and results in a fragmented picture of the MW. Integrated properties that are relatively painless to obtain in external galaxies (though dust obscuration can affect these observations as well; see e.g. Masters, Giovanelli & Haynes 2003; Masters et al. 2010a) are impossible to obtain directly within our own Galaxy (e.g. Mutch, Croton & Poole 2011). As a result, simulators often must resort to comparing their simulated MWs to very general galaxy populations (such as sets of Sbc or late-type galaxies) that, while superficially resembling the MW, have a wide range of other global properties (e.g. Guedes et al. 2011).

In an effort to circumvent our limited view of the MW, we can study galaxies that mimic the properties of our Galaxy but can be observed from outside, which we label Milky Way Analogues (MWAs). This method hinges on the Copernican assumption that the MW should not be extraordinary among a galaxy population that shares some key properties with it. These comparisons are enabled by working within volume-limited subsets of large surveys, which ensure that the observed objects constitute a representative population. Previous work suggests that galaxies with similar stellar mass and star formation rates (SFRs) are also similar in other properties, as the observed galaxy population is well matched by models that parametrize galactic star formation histories with a limited collection of curves (Behroozi et al. 2013; Gladders et al. 2013; Abramson et al. 2014; Kelson et al. 2014). Even further, Bell & de Jong (2001) showed that mass and SFR are strongly correlated with the photometric properties of a galaxy. Therefore, we can exploit the fact that two galaxies of identical mass and SFR should have similar luminosities and colours, with some scatter given the range of galaxy photometric properties at fixed physical parameters. Licquia, Newman & Brinchmann (2015), hereafter **LNB15**, utilized this to constrain the MW’s optical colours and magnitudes based on the range of observed properties of MWAs that were matched in stellar mass and SFR. MWAs also allow direct comparison of properties of our Galaxy to its closest peers (e.g. Licquia & Newman 2016; Licquia, Newman & Bershadsky 2016; Fraser-McKelvie, Merrifield & Aragón-Salamanca 2019; Boardman et al. 2020a; Krishnarao et al. 2020) and have been a successful tool for improving our understanding of the MW in an extragalactic context.

The MW, however, has some characteristics that are atypical (at the  $<2\sigma$  level) among its peers – e.g. the MW has an unusually compact disc (i.e. a small disc scale length; Bovy & Rix 2013; Bland-Hawthorn & Gerhard 2016; Licquia & Newman 2016), and an unusually quiescent merger history [from observation (Unavane, Wyse & Gilmore 1996; Ruchti et al. 2015) and simulation (e.g. Fielder et al. 2019; Carlesi et al. 2020)]. The deviations of the MW from the average suggest that we should consider parameters beyond just stellar mass and SFR in order to identify samples of objects that more closely resemble the MW.

Galaxy morphological characteristics such as disc scale length ( $R_d$ ) and bulge-to-total ratio ( $B/T$ ) are tied to a galaxy’s evolutionary history and therefore should connect to its photometric properties (Cappellari 2016; Saha & Cortesi 2018) as well as to the ways in which the MW is atypical. We would therefore wish to incorporate these properties in addition to stellar mass and SFR in defining an MWA. However, as the number of parameters required to match the MW increases, the number of MWAs correspondingly reduces dramatically. For example, Fraser-McKelvie et al. (2019) only found 179 analogues when selecting on stellar mass, bulge-to-total mass ratio, and morphology; Boardman et al. (2020a, b) found no MWAs within  $1\sigma$  of the MW when selecting on stellar mass, SFR, bulge-to-total ratio, and disc scale length in either the Sloan Digital Sky Survey IV (SDSS-IV) MaNGA survey (Bundy et al. 2015) or a larger

photometric sample drawn from the *GALEX*–SDSS–*WISE* Legacy catalogue (GSWLC; Salim et al. 2016), respectively.

LNB15 found that the colour of the MW is consistent the green valley region of the CMD as it has been defined using purely optical passbands. Characterizing the UV and infrared (IR) colours of the MW can provide more sensitive probes of whether it would be classified as in the process of quenching if seen from outside, allowing us to better understand what type of population the MW may belong to. LNB15 speculated that the MW might belong to the population of massive ‘red spiral’ galaxies, which are characterized by their red optical colours despite ongoing star formation (Masters et al. 2010b; Cortese 2012). Galaxies within this population may be moving into the green valley due to slow quenching (cf. Schawinski et al. 2014). This conjecture can only be fully tested by examining wavelengths outside of the optical range; in  $g - r$ , the colours of massive spiral galaxies on the star-forming main sequence (a population that should include the Galaxy) overlap with both the red sequence and the blue cloud (Cortese 2012; Salim 2014). However, samples of MWs that have high-quality photometry over a broader wavelength range will have reduced numbers due to the limited coverage of sufficiently deep photometry in *GALEX*.

To address the lack of analogues when multidimensional parameter spaces are used, and the smaller overall sample size resulting from the increase in wavelength coverage, we introduce a Gaussian process regression (GPR) approach in this work. GPR is an emergent tool in astrophysics. For example, Bocquet et al. (2020) employed GPR to emulate results of cosmological simulations, while Gordon, Agol & Foreman-Mackey (2020) used GPR to detect and classify exoplanets. We can use GPR to leverage information from a wider variety of galaxies, instead of just the closest MW analogues, in order to extract information from large-scale trends between galaxy physical and photometric properties. Thanks to the probabilistic framework that underlies GPR, we obtain uncertainty estimates for all predicted quantities for free. The primary result from this paper will be a UV to IR SED of the MW as viewed face-on, determined via GPR based on star formation history and structural parameters (i.e. galaxy physical parameters) that have been measured well for both the MW and galaxies from the SDSS (Aihara et al. 2011).

The paper is organized in the following manner: In Section 2, we describe the observational data used, including the external galaxy data in Sections 2.1 and 2.2, and estimated properties of the MW in Section 2.3. Section 3 details our new GPR-based methodology. In Section 4, we compare the luminosity and colours of the MW at multiple wavelengths and the MW’s predicted SED to properties of other galaxies. Finally, we summarize our results, and discuss implications and future work in Section 5. We provide supplementary appendices digitally. Within this document, Appendix A provides a summary of the galaxy parameters and tables that list predicted photometry for the MW. Appendix B describes tests of the accuracy of the GPR procedures used here, and Appendix C describes how we address the systematic corrections needed for  $k$ -corrections and Eddington bias.

In this paper, all magnitudes are reported in the AB system, except for the Johnson–Cousins UBVRI magnitudes that are presented in the Vega system. Absolute magnitudes are derived using a Hubble constant  $H_0 = 100 \text{ km s}^{-1} \text{ Mpc}^{-1}$ , so they are equivalent to  $M_y - 5 \log h$  (where  $M_y$  is the  $y$ -band absolute magnitude and  $h = H_0/100$ ) for other values of  $h$ . For other properties in which measurements for the MW are compared to extragalactic galaxy measurements, we assume  $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$  ( $h = 0.7$ ) in accordance with Licquia et al. (2015), for a standard flat Lambda cold dark matter cosmology with  $\Omega_m = 0.3$ . Parameters such as log stellar mass and log SFR can

be modified for different  $h$  values by subtracting  $2 \log h/0.7$ . We do this to avoid confusion and to allow for potential updates to future  $h$  measurements.

## 2 OBSERVATIONAL DATA

In this section, we describe the many galaxy catalogues utilized in this work. We break this up by photometry (Section 2.1) and inferred galaxy properties (Section 2.2), with the MW measurements included in the final subsection (Section 2.3).

### 2.1 Photometry

#### 2.1.1 SDSS galaxies

The sample of galaxies that we use as a starting point originates from the eighth data release (DR8; Aihara et al. (2011)) of the SDSS-III (York et al. 2000). DR8 provides both images and photometry of thousands (almost  $10^6$ ) of local galaxies. The optical broad-band passbands,  $u$ ,  $g$ ,  $r$ ,  $i$ , and  $z$  were the subjects of previous MW analogue work by LNB15 and are used in this study in addition to bands outside of the optical range.

We make use of both the ‘model’ and ‘cmodel’ magnitudes from SDSS. The former refers to magnitudes derived from the better of either a de Vaucouleurs or an exponential profile fit to the galaxy surface brightness distribution. These types of magnitudes are expected to produce the highest signal-to-noise estimate of galaxy colours; thus, when we refer to galaxy colours derived from SDSS we will be using model magnitudes for the calculations. Alternatively, cmodel magnitudes are derived from the best fit to a linear combination of a de Vaucouleurs and exponential profile. These magnitudes provide the best estimate of the total flux of a galaxy in each passband. When we refer to galaxy absolute magnitudes for SDSS bands, we will use ‘cmodel’ magnitudes.

$k$ -corrections on these passbands to rest-frame  $z = 0$  were calculated via the KCORRECT V4.2 software (Blanton & Roweis 2007), as described in LNB15. This provided AB absolute magnitudes for the SDSS *ugriz* photometry. Additionally, KCORRECT was used to convert the SDSS *ugriz* photometry to rest-frame Johnsons–Cousins *UVBRI* Vega magnitudes in order to make easy comparisons to literature values. Results are presented with the adoption of the Blanton & Roweis (2007) and LNB15 notation, where an absolute magnitude of passband  $y$  at redshift  $z$  is denoted as  $^z M_y$ .

Our main galaxy sample is derived from the volume-limited sample presented in LNB15. A volume-limited sample is required for accurate results from MW analogues in order to alleviate a radial selection effect known as Malmquist bias, i.e. the preferential inclusion of intrinsically bright galaxies. At higher redshifts within the main SDSS sample (Strauss et al. 2002), only the most luminous galaxies will be brighter than the sample magnitude limit and followed-up spectroscopically. By using a volume-limited sample, we ensure that galaxies within the range of the MW’s parameters are included equally at all distances considered. LNB15 determined the limits for their volume-limited sample from an initial draw of MW analogues from the full SDSS DR8 parent catalogue without any redshift cuts. Then, in  $^0(g - r)$  versus  $^0 M_r$  (i.e. rest-frame  $g - r$  colour derived using  $z = 0$  passbands versus  $r$ -band absolute magnitude, again evaluated with the  $z = 0$  passband) colour–magnitude space a maximum redshift was chosen such that all objects as low in luminosity as the faintest MW analogues would still be included at that  $z$ . A minimum redshift was also applied to limit the impact of the finite SDSS fibre aperture on measured galaxy properties. The



resulting volume-limited sample contains a total of 124 232 target galaxies within the redshift range of  $0.03 < z < 0.09$ . Some initial cuts on SDSS quality flags were also employed; for further details on the construction of this volume-limited sample, refer to LNB15, section 3.1. All cross-matches from SDSS to other catalogues presented here were constructed only using the volume-limited sample. Both the SDSS sample used here and the cross-matched catalogues presented in this paper are available at our catalogue GitHub repository.<sup>1</sup>

### 2.1.2 GALEX–SDSS–WISE Legacy catalogue

Photometry in UV and IR wavelengths used in this work comes from the GSWLC (Salim et al. 2016; Salim, Boquien & Lee 2018). We use GSWLC-M2, the medium-deep catalogue of GSWLC-2, which covers 49 per cent of the SDSS DR10 footprint. While this reduces the number of targets for study, the improved signal-to-noise in the UV-imaging over the shallow catalogue enables tighter results. SDSS photometry between DR7/DR8 and DR10 is the same, so no issue arises between cross-matches of the SDSS volume-limited sample and the GSWLC-M2 sample. In order to account for any differences in astrometry, we consider matches to be separations within 1.5 arcsec.

The UV sample in the GSWLC catalogue originates from the GALEX survey (*FUV* and *NUV*), and has been corrected for galactic reddening and calibration errors (Martin & GALEX Team 2005). UV detections are available for 74 per cent of SDSS targets in the GSWLC-M2 catalogue. The IR sample in the GSWLC catalogue originated from the 2MASS and WISE surveys. The 2MASS photometry (*JHKs*) is available for 48 per cent of SDSS targets and WISE photometry (*W1*, *W2*, *W3*, and *W4*) is available for 41 per cent of SDSS targets.

### 2.1.3 DESI Legacy imaging surveys

DR8 of the DESI (Dark Energy Spectroscopic Instrument) Legacy imaging surveys includes *g*, *r*, *z*, and WISE photometry of sources detected in DECam or BASS/MzLS imaging (Dey et al. 2019). We use these catalogues for all WISE-band photometry presented here, as the matched-model measurements from the Legacy Surveys Tractor catalogues go substantially deeper and have lower errors than other public WISE data products. We also calculate *k*-corrections for the WISE bands using this photometry; we discuss our procedures (inspired by those of Beare et al. 2014) for this, which allow accurate *k*-corrections in the IR without making any assumptions about SED templates at those wavelengths, briefly in supplementary Appendix C1 and more extensively in a forthcoming paper (Fielder, Andrews & Newman, in preparation).

DESI Legacy and SDSS contain photometry that comes from differing filter sets and detectors. DESI Legacy North uses BASS and MOSAIC filters while DESI Legacy South used DECam filters (Dey et al. 2019). Our corrections for offsets between the DESI Legacy Survey and SDSS photometry are described in supplementary Appendix C2.

## 2.2 SDSS-based properties

Only one of the galaxy properties that we use to create SEDs is recorded in the standard SDSS photometric catalogues and required no further analysis, specifically, the best-fitting axial ratio (*b/a*) of a

galaxy. We use the *b/a* value determined from an exponential fit to the galaxy’s surface brightness density in the *r* band throughout this work.

Axial ratio is used here as a proxy for galaxy inclination. Disc galaxies with very small axial ratios are likely heavily tilted from our perspective, in contrast to face-on discs that would typically have axial ratios of  $b/a \sim 0.9$  (Cho & Park 2009; Maller et al. 2009). Highly inclined star-forming galaxies suffer greatly from dust obscuration that can affect their colours and magnitudes by making them appear redder and fainter than their intrinsic properties (e.g. Conroy, Schiminovich & Blanton 2010; Masters et al. 2010a; Morselli et al. 2016; Kourkchi et al. 2019). This effect is important to consider when we are making predictions for a disc galaxy like the MW.

### 2.2.1 MPA-JHU masses and SFRs

The MPA-JHU galaxy property catalogue provides total mass and SFR estimates for galaxies within SDSS DR7. LNB15 developed an updated version of this catalogue using the SDSS DR8 photometry. To summarize, the masses are calculated by fitting stellar population synthesis model to the galaxy’s photometry [instead of spectral features, as in Kauffmann et al. (2003) and Gallazzi et al. (2005)], similar to Salim et al. (2007). The SFRs are calculated by emission-line modelling based on Brinchmann et al. (2004) with some updates. In order to account for aperture bias, LNB15 follow the method of Salim et al. (2007) in calculating photometry for the light that falls outside of the fibre and fitting stochastic stellar population synthesis models to it. Thus, each galaxy SFR measurement consists of a combination of the SFR measured from inside and outside of the SDSS fibre. For more details on these calculations, refer to LNB15, section 2.2.2, and Brinchmann et al. (2004). In both cases, a Kroupa initial mass function (broken power law) is assumed (Kroupa & Weidner 2003).

Each galaxy in our volume-limited sample is assigned a posterior probability distribution function (PDF) of log stellar mass and log SFR, as well as a corresponding cumulative distribution function (CDF) determined by the posterior. Nominal values used for the log  $\dot{M}_*$  and log SFR of each object are taken to be the mean value of the posterior. The standard deviation ( $\sigma$ ) on these values is calculated from percentiles (*P*) in the CDF provided in the MPA-JHU measurements; we take  $\sigma = (P_{84} - P_{16})/2$ , i.e. half the difference between the 16th and 84th percentile values. This gives us estimates of the stellar mass and SFRs as well as their associated errors for each galaxy in the sample.

The GSWLC-2 catalogues that we use for photometry also provide computations of stellar masses and SFRs. We chose to use the MPA-JHU results instead in order to avoid any systematic effects that may result from using the same measurements for our galaxy properties that were used to determine these derived quantities. We have tested how our results change if we use GSWLC-2 masses and SFRs and find minimal differences, as summarized in Section 4.2.4.

### 2.2.2 Simard et al. bulge and disc decompositions

In order to characterize the light-weighted bulge-to-total ratios (*B/T*) and disc scale lengths (*R<sub>d</sub>*) for SDSS galaxies, we use the catalogue of Simard et al. (2011). This work performed galaxy image decompositions for objects within the Legacy area in SDSS via the GIM2D software package (Simard et al. 2002).

<sup>1</sup><https://github.com/cfielder/Catalogs>

Specifically, we use the catalogue of fits based on composites of a Sersic  $n = 4$  bulge and a pure exponential disc (Sersic  $n = 1$ ). For our analysis, we use the  $B/T$  computed in the  $r$  band, as it is expected to be more stable than  $g$ -based measurements. As a check, we have performed our analyses using  $B/T_g$  instead of  $B/T_r$  and found minimal differences (cf. Section 4.2.4). It is worth noting that Kruk et al. (2018) find that these bulge + disc decompositions can be somewhat inaccurate when applied to strongly barred galaxies, which may lead to some biases within our galaxy sample. However, fits optimized for barred galaxies have been performed for only samples of a few thousand objects, inadequate for our purposes, so we rely on the Simard et al. (2011) results here. The Simard et al. (2002) bulge and disc decompositions are derived using an  $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$ .

### 2.2.3 Galaxy Zoo 2 bar presence

The presence of bars has been speculated to be correlated to the star formation history within a galaxy. However, the sense of the effect is unknown; a bar may be related to an increase in star formation (e.g. Hawarden et al. 1986; Devereux 1987; Hummel et al. 1990; Hawarden, Huang & Gu 1996), no effect (Pompea & Rieke 1990; Martinet & Friedli 1997; Chapelon, Contini & Davoust 1999), or decreased star formation (e.g. Vera, Alonso & Coldwell 2016; Díaz-García et al. 2020; Fraser-McKelvie et al. 2020), along with other potential effects on colour and metallicity (see e.g. Masters et al. 2011). Regardless of the source of any correlations, we wish to incorporate any possible differences having a bar may cause when we determine the MW's SED, since our Galaxy does exhibit clear evidence of a bar (e.g. Blitz & Spergel 1991). The Galaxy Zoo 2 (GZ2) catalogue (Willett et al. 2013; Hart et al. 2016) contains identifications of detailed morphological features in disc galaxies, such as spiral arms and bar presence. The galaxies classified in GZ2 are a subsample of the brightest and largest galaxies in the SDSS Main Galaxy Sample.

The Galaxy Zoo projects are open public projects in which members of the community identify whether they found a variety of galaxy features in the images provided. In the GZ2 catalogue, the number of raw votes for each morphological feature is weighted (to account for user consistency) and adjusted to mitigate the impact of redshift-dependent biases, yielding a corrected fraction of volunteers who identified a given morphological feature for each galaxy. In our case, we focus on the debiased fraction of volunteers who identified a galaxy as having a bar, which we denote by  $p_{\text{bar}}$  (following Galaxy Zoo labelling conventions; however, this is a fraction of votes, not a probability). Above some threshold in  $p_{\text{bar}}$  (after cuts in related parameters), we can be confident that a given galaxy indeed hosts a bar. Willett et al. (2013) developed the initial version of GZ2 bar thresholds; more conservative thresholds were later defined by Galloway et al. (2015).

Often, when one uses Galaxy Zoo results it is necessary to consider responses to previous questions that influence whether the question of interest is even presented to the volunteers. For example, as described in Willett et al. (2013), a voter is only asked ‘Is there a sign of a bar feature through the centre of the galaxy?’ if they first selected that the object has ‘features or disc’ when asked ‘Is the galaxy simply smooth and rounded, with no sign of a disc?’, and then responded in the negative when asked ‘Could this be a disc viewed edge-on?’. One would then identify galaxies that have received a large fraction of ‘yes’ votes as containing bars. It is worth noting that Willett et al. (2013) state galaxies that receive fewer than 10 net votes for a given question may not have a reliable classification.

Therefore, to construct a bar sample using the minimum allowances of Willett et al. (2013), using the debiased vote fractions from Hart et al. (2016) [which was an improvement to the debiasing of Willett et al. (2013)], one would use the cuts in the second and third rows of the third column of table 3 in Willett et al. (2013) with  $p_{\text{bar}} \geq 0.3$  as recommended by Galloway et al. (2015) and additional voting number thresholds  $N_{\text{not edge on}} \geq 10$ , and  $N_{\text{bar}} \geq 10$ .

However, here we do not want to consider only information on the relationship between galaxy star formation history, structural properties, and photometric properties that comes from those galaxies that most definitively host bars. Rather, it is desirable for the training sample for our GPR model to include objects spanning a broad range of parameter space: This yielded smaller net prediction errors in MW photometric parameters than when small, restricted samples are used for training. For additional discussion on this choice, refer to Section 4.2.4-3.

To summarize, the primary set of data that we employ in this paper consists of a cross-match between the original SDSS DR8 volume-limited sample reported in LNB15, an updated version of the MPA-JHU catalogue (Brinchmann et al. 2004) of masses and SFRs in SDSS, the Simard et al. (2011) morphological bulge–disc decomposition catalogue, the GSWLC-2 medium-deep photometry catalogue (Salim et al. 2016, 2018), the GZ2 catalogue (Willett et al. 2013; Hart et al. 2016), and the DESI Legacy Survey DR8 catalogue (Dey et al. 2019). These cross-matches were performed using the *astropy* coordinates package. We required that objects be separated by less than 1.5 arcsec from a counterpart in the SDSS volume-limited sample to be considered a match, and discard any galaxies that are not included in all catalogues considered.

Our sample is smaller than the original volume-limited sample as a result. The MPA-JHU catalogue contains all objects from the volume-limited sample, so we begin with the same set of 124 232 galaxies utilized by LNB15. After matching to Simard et al. (2011), 123 167 galaxies remain; some objects are lost due to minor differences between the DR7 and DR8 SDSS catalogues. When we require GSWLC-M2 measurements, we are left with only 60 857 galaxies, as roughly half of the SDSS footprint is covered by medium-deep *GALEX*, 2MASS, and *WISE* photometry (see Section 2.1.2). After matching to GZ2, 29 836 galaxies remain, a consequence of the brighter magnitude limit used to select GZ2 objects compared to the original SDSS Main Galaxy Sample (see Section 2.2.3). We note that due to the brighter magnitude limit of GZ2, our final sample is no longer volume limited. This would yield biased results if we were measuring aggregate properties of MW analogues. However, for our GPR methodology this only modulates the density of our training set within parameter space, causing larger prediction errors due to the sparser sampling, but not leading to a bias.

Finally, matching to DESI Legacy leads to only a minor reduction in galaxy number as it covers a superset of the SDSS area with deeper photometry (but different reduction algorithms). After these cross-matches, we then remove objects with photometric values of ‘NaN’, infinity, or –99, which all indicate missing photometry in one catalogue or another. This is only done when necessary for the evaluation of the GPR. For our *WISE*  $k$ -correction calculations (see supplementary Appendix C1), we will exclude objects with large *WISE* photometry errors in a given band, which we do not propagate into our main galaxy sample.

The final data sample consists of 29 588 galaxies in total from redshift  $0.03 < z < 0.09$ , which is publicly available at our catalogue GitHub. The parameters that we will use to predict photometric properties are stellar mass ( $M_*$ ), SFR, galaxy axial ratio as a proxy

for inclination ( $b/a$ ), bulge-to-total ratio ( $B/T$ ), disc scale length ( $R_d$ ), and corrected bar vote fraction ( $p_{\text{bar}}$ ). Covariances among these parameters are minimal, as discussed further in supplementary Appendix A1; we show joint distributions for these parameters in Fig. A1.

### 2.3 MW properties

A number of the properties of the MW used in this study have been derived using the Bayesian mixture model meta-analysis method first presented in Licquia & Newman (2015). This technique combines information from multiple measurements in order to obtain aggregate constraints on the MW's properties, taking into account the possibility that individual measurements could be incorrect or have their errors miss-estimated. This Bayesian method is combined with Monte Carlo simulations in order to account for uncertainties on the Sun's measured Galactocentric radius,  $R_0$ ; the Galactic exponential disc scale length; and uncertainties in the local surface density of stellar mass. The inferred mass of the MW's stellar disc depends on all three of these parameters.

Gravity Collaboration (2019) recently obtained a greatly improved geometric measurement of our distance from the centre of the MW,  $R_0 = 8.178 \pm 0.026$  kpc [Licquia & Newman (2015) used the value  $8.3 \pm 0.35$  kpc from Gillessen et al. (2009)]. We have rerun the Bayesian mixture model inference from Licquia et al. (2015) and Licquia & Newman (2016) with this updated measurement of the Galactocentric distance to obtain updated measurements of the MW's mass, bulge-to-total mass ratio, and disc scale length; see those papers for all details of the data sets used and the calculations [the SFR estimate from Licquia & Newman (2015) is not affected by the value of  $R_0$ , so we adopt it unchanged]. The resulting MW parameters used in this study are as follows:

- (i) Bulge mass  $M_*^B = (0.90 \pm 0.06) \times 10^{10} M_\odot$ ;
- (ii) Disc mass  $M_*^D = 4.58^{+1.18}_{-0.94} \times 10^{10} M_\odot$ ;
- (iii) Total stellar mass  $M_* = 5.48^{+1.18}_{-0.94} \times 10^{10} M_\odot$ ;
- (iv) SFR =  $1.65 \pm 0.19 M_\odot \text{ yr}^{-1}$  (Licquia & Newman 2015);
- (v) Log specific SFR  $\log \frac{\text{SFR}}{M_*} = -10.52 \pm 0.10$ ;
- (vi) Bulge-to-total mass ratio  $B/T = 0.16 \pm 0.03$ ;
- (vii) Disc scale length  $R_d = 2.48^{+0.14}_{-0.15}$  kpc.

The revised stellar mass estimate for the MW is smaller than the previous estimate, but by an amount that is significantly smaller than the previously estimated uncertainties.

By design, these physical parameters are constructed such that they can be directly compared to the extragalactic catalogues used to predict photometric properties. For example, the stellar mass and SFR estimates for the MW are constructed assuming a Kroupa initial mass function (Kroupa & Weidner 2003) and an exponential disc model, which is the same way in which mass and SFRs were calculated in the MPA-JHU catalogue (Brinchmann et al. 2004) that we use in this analysis. The only parameter where this is not fully the case is the bulge-to-total ratio,  $B/T$ . For the MW, we can securely estimate only a mass-weighted value for this quantity. In external galaxies, however, the mass-weighted  $B/T$  is much more difficult to obtain, and light-weighted measurements tend to be more reliable. However, our predicted MW properties do not change significantly when we switch from  $r$ -band-based to  $g$ -band-based  $B/T$  values, even though mass-to-light ratios differ significantly between these bands, suggesting that this is not a major issue.

Our analysis also depends on two additional parameters that are determined independently of any Bayesian mixture model meta-analysis. These quantities have values assigned to them based on

their meaning in their respective catalogues and our understanding of the Milky Way:

- (i) Axial ratio (inclination proxy)  $b/a = 0.9 \pm 0.1$ ;
- (ii) Bar vote fraction  $p_{\text{bar}} = 0.45 \pm 0.15$ .

Galaxy inclination has a strong effect on colour and luminosity measurements for disc galaxies. As mentioned in Section 2.2, dust alters the observed colours and magnitudes of star-forming galaxies that are highly inclined or edge-on. Our perspective within the MW makes it somewhat equivalent to being edge-on to us (though our position within the Galaxy, rather than outside, does cause some differences). However, photometric properties are most cleanly determined for those objects that are observed face-on. Therefore, we predict the SED that would be observed for the MW for axial ratio values drawn from a uniform distribution spanning from  $b/a = 0.8$  to  $1.0$ , consistent with the intrinsic axial ratios of spiral galaxy discs as described by Maller et al. (2009) and in Section 2.2. Our results should therefore correspond to the properties of our Galaxy if it were observed face-on. While axial ratio is a good proxy for inclination, it is not a perfect substitute. For example, van de Sande et al. (2018) found that disc galaxies that are rounder ( $b/a \sim 1$ ) tend to be older and therefore intrinsically redder. This means that small biases could result from treating the MW as having a face-on axial ratio. However, there is no straightforward way to avoid this, and the effect should be small compared to other sources of error.

The MW exhibits clear evidence that it contains a bar (e.g. Blitz & Spergel 1991; Shen & Zheng 2020). However, very few GZ2 galaxies have  $p_{\text{bar}} = 1.0$ , and those galaxies with the highest bar vote fractions are expected to have *very strong* bars, which may not match our Galaxy. Therefore, we assume that in GZ2 the MW would have a vote fraction above the threshold for defining a bar, but not a value higher than the bulk of barred galaxies. Galloway et al. (2015) and Willett et al. (2013) find that  $p_{\text{bar}} \geq 0.3$  serves as a reliable threshold between bar presence and lack thereof. Galaxies with  $0.3 \leq p_{\text{bar}} < 0.5$  likely have weaker bars while galaxies with  $p_{\text{bar}} > 0.5$  likely have stronger bars. Because the bar strength of the MW's bar as it would be determined from outside our Galaxy is not well constrained, we treat the bar vote fraction for the MW as uniformly distributed between  $p_{\text{bar}} = 0.3$  and  $0.6$ . Choosing a larger mean vote fraction has small effect on our results, given the large range of fractions considered (compared to the distribution of  $p_{\text{bar}}$  in GZ2).

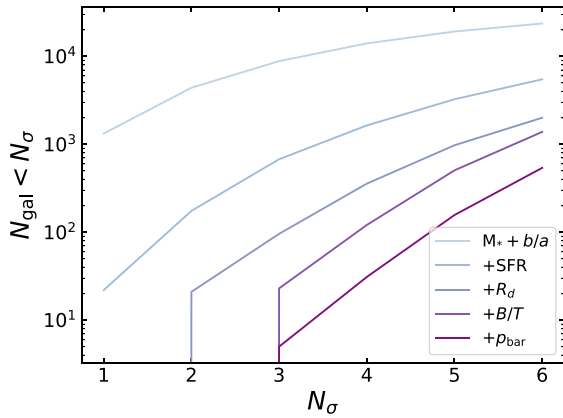
## 3 GPR FOR PREDICTING MW PHOTOMETRY

In this section, we describe how GPR can be used to estimate photometric properties for the MW. First, we explain the need to transition from MW analogue-based methods to GPR when we consider higher dimensionality parameter spaces in Section 3.1. In Section 3.2, we explain the basic concepts behind GPR, and in Section 3.2.3 we highlight the fundamental differences between GPR and analogue galaxy methods. Section 3.2.1 describes the kernel used to set up our GPR, which guides how information is propagated from training objects to predictions. We briefly describe the computational limitations of the GPR implementation we are using in Section 3.2.2. Lastly, Section 3.2.4 investigates the contributions of various sources of uncertainty to our GPR predictions.

### 3.1 Limitations of using analogue galaxies

Using MW analogues to predict the photometric properties of the MW, as was done in Licquia et al. (2015), has been a very useful methodology but also has limitations. Of particular concern is the





**Figure 1.** Total number of galaxies within an allowed tolerance ( $N_{\text{gal}} < N_{\sigma}$ ) as a function of the maximum deviation away from the Galactic values allowed for each parameter, in units of the uncertainty in the MW value of that parameter,  $N_{\sigma}$ . The lightest shade of blue denotes galaxy counts when analogues are selected using only stellar mass ( $M_*$ ) and axial ratio ( $b/a$ ). Consecutive parameter additions are cumulative. Thus, the darkest purple line incorporates all six parameters of stellar mass, axial ratio, SFR, disc scale length ( $R_d$ ), bulge-to-total ratio ( $B/T$ ), and bar vote fraction ( $p_{\text{bar}}$ ). The  $\sigma$  values used for each parameter are defined in Section 2.3, except in the case of bar vote fraction, for which we use a smaller error ( $\sigma_{p_{\text{bar}}} = 0.05$  instead of 0.15) for illustrative purposes. Note that there are very few galaxies even within large tolerance windows when using several parameters, e.g. only  $\sim 10$  MWAs within  $3\sigma$  for five parameters simultaneously, and no objects that are close to the MW in all aspects, since the MWA method scales poorly with increased dimensionality.

dramatic reduction in MWA sample size that occurs as the number of parameters that must be matched increases; we would like to move from the two parameters considered by Licquia et al. (2015) ( $M_*$  and SFR) to a total of six, adding  $b/a$ ,  $B/T$ ,  $R_d$ , and  $p_{\text{bar}}$ . Requiring that analogues be MW-like in more ways should reduce the spread in photometric properties of the resulting sample, potentially enabling stronger constraints. There are only limited correlations between these six parameters (cf. Fig. A1 in supplementary Appendix A1), so degeneracies between them are minimal: They each add new information. For instance, galaxies with stellar mass and SFR matching the MW exhibit bulge-to-total ratios ranging from 0 to 1: Structural and star formation history parameters carry distinct information. However, while matching on additional galaxy parameters produces a population of analogues that must each be closer in properties to the MW, the resulting MWA sample becomes much smaller (with as few as  $\sim 5$  analogues in the sample that are within  $3\sigma$  of the MW in all of the properties considered, and none within  $2\sigma$  for every aspect).

The reduction in the size of analogue galaxy samples as more parameters are considered is illustrated in Fig. 1. We plot the total number of galaxies that are within a given number of  $\sigma$  for every MW parameter considered (where  $\sigma$  represents the uncertainty in the MW value for a given property) as a function of the number of  $\sigma$  used as a threshold. The lightest shade of blue denotes the number of analogues within a given threshold when only considering stellar mass ( $M_*$ ) and axial ratio ( $b/a$ ). The consecutive additions indicated in the legend represent the inclusion of the listed parameter in addition to all previous ones; i.e. we consecutively incorporate SFR, disc scale length, bulge-to-total ratio, and finally bar vote fraction. Hence, the darkest purple line shows the number of analogues when using all six parameters (which we have used in order of decreasing constraining power on MW colours). The  $\sigma$  tolerances used for each parameter are the same MW measurement errors defined in Section 2.3, *except*

for  $p_{\text{bar}}$ . The significant uncertainty we fiducially ascribe to bar vote fraction would cause the five parameter and six parameter lines to be degenerate with one another; to avoid confusion, we use  $\sigma_{p_{\text{bar}}} = 0.05$  instead of 0.15 when constructing this plot.

The previous work by LNB15 would most closely correspond to the +SFR (three parameters), medium blue line. If one were to restrict to objects within  $\pm 2\sigma$  of the MW value for all parameters employed, there would be a total of zero MW analogue galaxies when using five or more parameters. In the six-dimensional space that we employ below, there are only 200 galaxies that are within even  $\pm 6\sigma$  of the MW value in all six properties; at that extreme, analogue samples would be selecting objects that are not very close to the MW at all. The lack of close analogues in high-dimensional parameter spaces makes constraints on MW properties from the MWA method weak in that limit, with correspondingly large uncertainties.

### 3.2 GPR: a powerful method for interpolation and prediction

To address the lack of MW analogue galaxies in our multidimensional parameter space, we have developed alternative methods for predicting MW properties based on GPR. In this subsection, we summarize the basic properties of GPR relevant for this work. For in-depth discussion, we refer the reader to Rasmussen & Williams (2006) and Görtler, Kehlbeck & Deussen (2019).

GPR (sometimes called kriging) is effectively a method of interpolation where information from training data is accounted for by a smooth and continuous weighting function, called a ‘kernel’ or covariance function. The joint probability distribution of the values of a Gaussian process at any finite set of points in parameter space will be a multivariate Gaussian (with a number of dimensions set by the number of points in the set); the kernel specifies how the covariance between points depends on their separation. The kernel should be a smooth and continuous function, with a length-scale (which governs how far information propagates from a given point) that is optimized by training on the observed data. We can then *predict* what the value and the uncertainty of the desired quantity would be at any arbitrary point in space by applying this kernel to the training data. This is in contrast to other supervised learning algorithms that typically make single-valued, ‘point’ predictions rather than predicting PDFs. It can be shown that GPR yields the minimum variance out of any unbiased interpolation method that depends only linearly on the training data; this makes GPR an optimal interpolation algorithm.

For our application of GPR, the galaxy sample described in Section 2 will serve as the training set. The six parameters we defined in Section 2.2 (stellar mass, SFR, etc.) serve as the features we will use for prediction. Our goal is to determine an optimized mapping from these physical parameters to a single output photometric parameter in our catalogue, e.g. the  $r$ -band absolute magnitude,  $^0M_r$ .

Once our training data are selected, we then go into the model-selection phase of GPR, during which the mean function and covariance function (or kernel) used for GPR are selected and tuned. We detail our selection of the covariance function in Section 3.2.1. Effectively, the kernel determines how information from a given training point will be propagated to make predictions at other points in parameter space. Hyperparameters describing the kernel are tuned at this step to maximize the log-marginal likelihood of the training data. After this step, we consider the model to be ‘fit’.

Finally, we enter the inference phase of GPR. At any point in our six-dimensional parameter space, we can now determine the posterior probability distribution for the parameter of interest by applying the kernel to the training data. When evaluated at a single point in parameter space, a Gaussian process corresponds to a 1D

Gaussian; we thus obtain both a predicted mean for the property of interest (e.g.  $^0M_r$ ) and the standard deviation of the Gaussian describing its uncertainty. In our example, we would pass in a set of physical parameters measured for the MW and obtain a predicted value for the  $^0M_r$  of the MW, as well as the uncertainty in that value.

In reality, we do not only query the GPR at the mean measured MW properties presented in Section 2.3; rather, we perform random draws from the PDFs describing  $M_*$ , SFR,  $b/a$ ,  $B/T$ ,  $R_d$ , and  $p_{\text{bar}}$  in order to incorporate the uncertainties in the MW’s measured properties into our analysis. For  $\log M_*$ ,  $\log \text{SFR}$ ,  $B/T$ , and  $R_d$ , we assume a normal distribution. For  $b/a$  and  $p_{\text{bar}}$ , we draw from uniform distributions, as described in Section 2.3.

We perform these random draws 1000 times (so that we have 1000 full sets of MW parameters). We then evaluate the GPR predictions for MW photometric properties at each of these points in parameter space. This gives us a prediction and error estimate corresponding to each draw from the PDFs of MW characteristics. Thus, we end up with 1000 total predictions for each MW photometric property. Our mean prediction for the MW in a given photometric band corresponds to the arithmetic mean of all these predictions.

In this work, all GPR calculations have been done the PYTHON SCIKIT-LEARN Gaussian Process module, *sklearn.Gaussian\_process* (Pedregosa et al. 2011). For details on the implementation of this module, refer to Pedregosa et al. (2011) and the SCIKIT-LEARN documentation.

The following subsections provide more details on some aspects of our GPR methods and their advantages.

### 3.2.1 Choice of kernel for GPR

In this work, we use a combination of two kernels for GPR: a Radial Basis Function (RBF) kernel and a white noise kernel. The RBF kernel decreases proportionally to  $\exp - \gamma D^2$ , where  $\gamma$  is a free parameter and  $D$  is the Euclidean distance between points; this kernel will cause the covariance between the predicted values from GPR at different points in parameter space to decrease as a Gaussian in distance as the separation between those points increases.

However, there is also scatter in galaxy photometric properties even for objects measured to have the same physical properties. In order to capture that, we also incorporate a white noise kernel, which models the spread in values for the predicted property at a fixed point in parameter space with normally distributed noise (Rasmussen & Williams 2006). The net covariance used for the GPR is then the sum of the distance-dependent covariance from the RBF kernel and the (diagonal) covariance matrix corresponding to the white noise kernel.

For a given set of training data, there are a nearly endless number of functions that can fit the given data points, each one a realization of the Gaussian process. The kernel creates a prior on the GP to constrain which functions from that set are most likely to describe the parameter space. The posterior is then determined using the training data values. Due to this probabilistic approach, the Gaussian process provides both predicted values and uncertainties at any points within the parameter space. Uncertainties due to the finite training sample size and its distribution in parameter space and those corresponding to intrinsic variation between training objects that have the same physical parameters are both captured. In regions of parameter space that are poorly constrained by the training data, the prediction uncertainties are correspondingly larger.

The kernels we use in this work are available in the *sklearn.Gaussian\_process.kernels* base class. For the

white noise kernel (*WhiteKernel* in SCIKIT-LEARN), we initialize the noise level to be 1; similarly, we initialize the length-scale for the RBF kernel to be 1 for each parameter. We opt to normalize the output photometric property to have mean zero and variance one across the training set, which helps to ensure that these initial guesses will have the right order of magnitude. The noise level and length-scales are then optimized and the regression model is built via the *sklearn.Gaussian\_process.GaussianProcessRegressor* class. In our model, we allow the optimizer to restart 10 times in order to find the kernel parameters that maximize the likelihood without being trapped in a false maximum.

### 3.2.2 Optimizing training samples

The computation time and required memory for the SCIKIT-LEARN implementation of GPR scale as the number of data points used to train the model squared and cubed, respectively. As a result, we find that the maximum training sample size we can use without running out of memory on the computers used for this work is  $\sim 6000$ ; it is infeasible to train from our entire catalogue when using this GPR implementation.

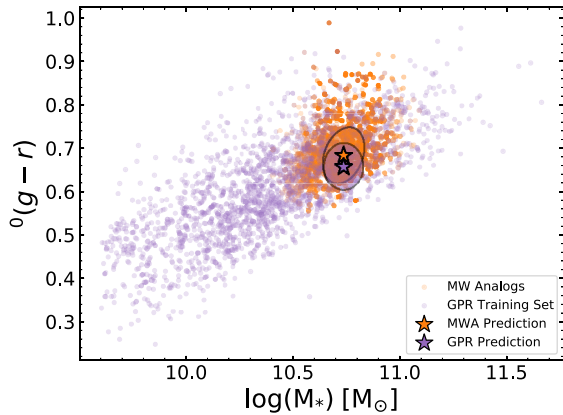
We have therefore tested the effects of either restricting to objects with physical parameters within some tolerance of the MW fiducial values or randomly selecting a subset of objects in order to reduce the training set size. We have focused on the root mean squared error (RMSE) of predicted MW photometry for the *NUV*, *r*, and *J* bands for this optimization. We use fivefold cross-validation for all the tests; i.e. we always train with 80 per cent of the data and test with 20 per cent, but rotate what objects are used for training and testing through the whole data set, and only retain the values for an object when it was in the test set. This provides unbiased estimates of the RMSE for a training set 80 per cent as large as the one we actually have. We find that the combination that offered the lowest RMSE across all bands while keeping computational time manageable was to randomly select 2000 galaxies out of the set of objects that are within  $12\sigma$  of the MW for every parameter of interest. We therefore adopted this training strategy for all results below.

### 3.2.3 Comparison to results from analogue samples

GPR can provide more accurate predictions than many other techniques thanks to its ability to leverage information from both nearby objects in the training set as well as from more distant objects that characterize larger scale trends. In our application, this allows the GPR to map from the MW’s physical properties to its photometric properties much more accurately than if we had only used the few objects that are similar to the MW in all respects (i.e. those that would be classified as MWAs) to inform the mapping.

Fig. 2 illustrates the fundamental difference between how properties are constrained by the MW analogue selection method versus GPR. For simplicity’s sake, we perform this comparison based on only three parameters (stellar mass, SFR, and axial ratio, the same ones utilized in LNB15), as the analogue method starts to break down when more parameters are included. We also do not correct for Eddington bias in either measurement (q.v. supplementary Appendix D) for simplicity. Objects included in a set of MWAs based on 5000 samples from the distribution of possible MW properties via methods equivalent to those from Licquia et al. (2015) are depicted by the orange points. The analogues fall within a narrow range of stellar mass, limiting the set of objects that contribute information. The





**Figure 2.** A comparison of galaxy samples and results from the MW analogue method versus GPR. For this analysis, analogues were selected based on stellar mass, SFR, and axial ratio, and we use the same parameters to predict colour via GPR. Orange points represent MW analogues selected by methods equivalent to Licquia et al. (2015), while purple points indicate the galaxies used to train the GPR. The predicted mean MW  ${}^0(g-r)$  colour is denoted by a star in the corresponding colour, and ellipses depict a  $1\sigma$  joint confidence region for colour and stellar mass. Both methods yield a similar predicted  ${}^0(g-r)$  colour for the MW when we have large numbers of analogues. In contrast to the small window that the MW analogues lie in, the GPR utilizes a wider variety of galaxies to capture larger scale trends.

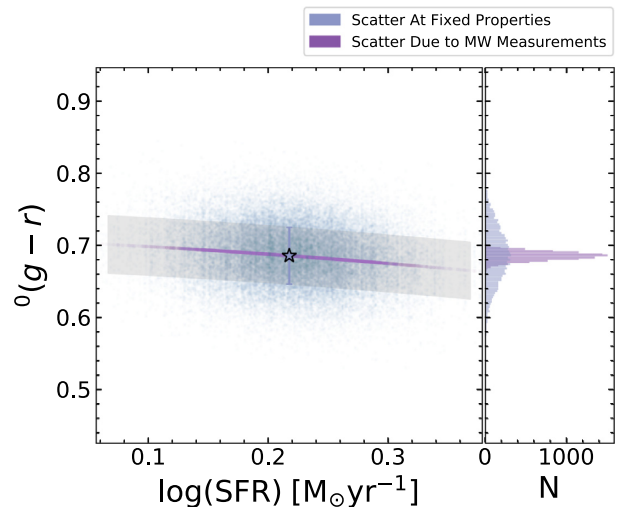
orange star represents the mean prediction for the MW’s  ${}^0(g-r)$  colour [ ${}^0(g-r) = 0.682$ ] resulting from this set (derived via the Hodges–Lehmann robust estimator; Hodges & Lehmann 1963). The orange ellipse depicts the  $1\sigma$  confidence region for  ${}^0(g-r)$  colour and stellar mass. In contrast to the MW analogues, the sample of galaxies used to train the GPR is shown by purple points. These cover a much broader range of parameter space than the MWAs. Similarly, the prediction for the MW’s  ${}^0(g-r)$  colour using GPR is shown by a purple star [with  ${}^0(g-r) = 0.668$ ], along with the corresponding  $1\sigma$  confidence region that is shaded in purple. When we have many analogues, both techniques yield very similar results, but unlike MWAs the GPR technique still provides strong constraints when we consider many parameters at once.

### 3.2.4 Characterizing sources of uncertainty

We can quantify the contributions of different sources of uncertainty to our GPR estimates by changing how we perform the regression. We illustrate our methods by evaluating how the prediction from a six-parameter GPR fit changes as a single parameter varies. The left-hand panel of Fig. 3 shows one of the physical parameters being regressed from, SFR, on the x-axis and the target value,  ${}^0(g-r)$  colour, on the y-axis. We chose this pair as galaxy SFR is expected to correlate with galaxy colour well at fixed stellar mass.

First, we isolate the scatter in colour at fixed properties; this corresponds to the contribution of the white noise kernel to the covariance function of the GPR. To determine the magnitude of this scatter, we query the GPR at the MW’s fiducial physical properties to obtain a predicted PDF of  ${}^0(g-r)$  at this point in parameter space from which we can draw samples. The standard deviation of the colour of these samples corresponds to the scatter encoded in the white noise kernel.

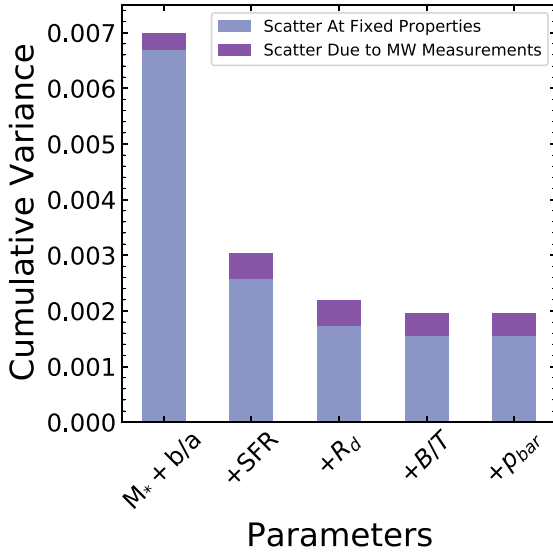
In the panel at the right of Fig. 3, we plot a histogram of 10 000 possible  ${}^0(g-r)$  colours drawn from the GP’s predicted PDF, evaluating it at the fiducial value of the MW’s SFR. In the left-hand



**Figure 3.** A breakdown of uncertainty due to scatter at fixed properties and scatter due to MW measurement uncertainty. We plot results from a six-parameter GPR trained to predict  ${}^0(g-r)$  colour. For this example, we vary only one of the input parameters, the SFR. The light violet colour corresponds to results when we use a fixed training set and evaluate the GPR at fixed MW properties; this isolates the variation attributable to the white noise kernel, corresponding to the scatter in colour at fixed galaxy properties. The dark purple colour corresponds to results when the training set remains fixed but the SFR value is drawn randomly from the fiducial MW PDF, so that the contribution to the colour error attributable to the uncertainty in the SFR of the MW can be evaluated (the other five parameters used to train the GPR are held constant for simplicity for this example). In the left-hand panel, the shaded region depicts the  $\pm 1\sigma$  range predicted by the GPR fit, plotted out to  $\pm 3\sigma$  of the MW’s SFR. The predicted  ${}^0(g-r)$  colour and  $1\sigma$  error for the MW’s fiducial SFR (and hence only including error at fixed properties) are shown by the star-shaped point and error bar in light purple. The predicted  ${}^0(g-r)$  colour from each of 5000 draws from the SFR PDF is shown by the purple points, which accounts for scatter due to the SFR measurements. These predictions perfectly trace the GPR fit and tend to fall within  $\sim 3\sigma$  of the MW’s SFR, by construction. For reference, we also show 10 samples from the GPR-predicted PDF for each of the 5000 random SFR values as faint blue points. The distribution of these points reflects both the scatter of galaxy colours at fixed properties and the uncertainty in the MW measurements. Histograms of the colours for each sample, whose distributions correspond to the scatter at fixed properties and the scatter due to MW SFR measurement uncertainties, are shown in the right-hand panel. It is evident that the spread in GPR predictions at fixed properties is much larger than the scatter that results from uncertainties in the MW’s SFR.

panel, the lavender star denotes the mean predicted  ${}^0(g-r)$  for the MW with this model, and the error bar corresponds to the standard deviation of the sample values. This error bar therefore corresponds to the  $1\sigma$  scatter in colour at fixed properties. The grey shaded area corresponds to the  $\pm 1\sigma$  band for the GPR prediction of colour for a range of  $\log(\text{SFR})$  values. By construction, the half-width of the band must match the standard deviation of samples from the PDF at fixed properties.

To instead isolate the contribution to errors resulting from the uncertainty in MW properties, we determine the distribution of mean predicted colours evaluated at varying values of SFR drawn from the PDF for the MW. We perform 1000 draws from the fiducial MW log SFR PDF in total and evaluate the GPR mean predicted  ${}^0(g-r)$  for each. In Fig. 3, the dark purple points in the left-hand panel show the resulting predictions, which all fall on a continuous curve by construction. The panel to the right shows the histogram of this set of predictions in purple. We quantify the scatter in colour attributable to



**Figure 4.** Contributions to the variance in rest-frame  $^0(g-r)$  colour for GPR employing varying sets of galaxy physical parameters. The x-axis shows the set of parameters used to predict colour; they increase cumulatively as we go from left to right, in order from the most constraining to the least constraining parameter. The variance decreases monotonically as the number of parameters increases. In every case, the scatter in colour at fixed properties dominates errors; uncertainties in the physical properties of the MW are subdominant. Contributions to uncertainties due to having a finite training set are small enough to be considered negligible, enough so that they would not be visible on this plot if we included them.

uncertainties in the MW’s measurements via the standard deviation of the  $^0(g-r)$  values at these 1000 points.

To illustrate the full range of values obtained via GPR, we plot 10 samples from the distribution of predictions for each of the thousand MW SFR draws as faint blue points in Fig. 3. These samples vary in colour both due to the scatter in colour at fixed properties and due to the uncertainty in MW properties.

We can extend these same ideas in 1D to evaluate the relative contributions of uncertainties in MW properties and of the scatter in properties at fixed colour to the error in MW  $^0(g-r)$  colour, for any GPR model of interest. The key difference from Fig. 3 is that, in order to quantify the full scatter due to uncertainties in MW properties, we allow all the parameters to vary, not only SFR. We present the results of this analysis for GPR models based on two to six physical parameters in Fig. 4. We use a stacked bar plot to display the contribution to the variance from each error source, where the x-axis is labelled according to the physical parameters used to train the GPR (where additions are all cumulative, so entries further to the right incorporate more parameters). Since independent errors will add in quadrature, the contribution to the net variance from each factor is proportional to the height of its bar. The variance due to the scatter at fixed properties is shown in a lighter violet shade, while that resulting from the uncertainty in MW properties is shown in dark purple. The scatter at fixed properties contributes to the majority of the error in the GPR for  $^0(g-r)$  colour.

We have also evaluated the contribution to uncertainties resulting from the finite size of the training sets used. This scatter is isolated by varying the randomly selected training sample 100 times. For each training sample, we evaluate the mean predicted value from GPR at the MW’s fiducial physical parameters. The contribution to uncertainties from the finite size of training samples is obtained by calculating the variance of the GPR predictions across all of the

training sets. These values are minute: The variances are an order of magnitude smaller than error attributed to the scatter due to MW measurements. Thus, any contribution to errors resulting from the finite training sample size is negligible.

We have performed the same error budget test for colours in the UV, near-IR, and mid-IR. The results mirror those presented in Fig. 4: The errors are dominated by scatter at fixed properties, followed by scatter from the MW measurements. In all cases, errors attributable to finite training set size are negligible compared to other sources. While the cumulative variance decreases for every parameter added when we predict colours, this is not the case for absolute magnitudes. In that case, the cumulative variance decreases as we add parameters until the sixth parameter,  $p_{\text{bar}}$ , is incorporated. At that point, the variance increases and the scatter due to finite training becomes more important. For this reason, all absolute magnitude predictions within this paper are performed using only five parameters, excluding  $p_{\text{bar}}$ .

While uncertainties in MW characteristics will contribute to the random errors in the derived photometric properties of our Galaxy, uncertainties in the physical parameters of the *training* galaxies can cause *systematic* errors. If the density of objects in parameter space varies quickly (with non-negligible second or higher derivatives), objects will more often scatter from well-populated regions of parameter space into sparser regions than vice versa. The resulting systematic shift in the measured distribution of parameters compared to the underlying distribution with no scatter is known as Eddington bias.

In the context of this work, Eddington bias will lead to shifts in the colour and luminosity predicted for the MW. We derive corrections for Eddington bias using methods similar to those of LNB15; we detail our procedures in supplementary Appendix D. The estimated Eddington bias is subtracted off from the GPR-predicted colours and luminosities for the MW to produce our final estimates for the Galaxy’s photometric properties and likewise the uncertainty on the Eddington bias calculations is propagated into our final error estimates. In general Eddington bias has small but non-zero effects on our results ( $< 1\sigma$  for almost all parameters, as listed in supplementary Appendix A).

In the following sections, the errors on GPR results presented include the contributions from scatter at fixed properties, uncertainties in MW properties, and uncertainty from Eddington bias.

### 3.2.5 Summary of the GPR algorithm for determining MW photometric properties

Here, we summarize the steps taken to predict MW photometric properties via GPR. Our method proceeds as follows:

- (i) Construct the training sample by restricting to objects within  $12\sigma$  of the MW in all physical parameters considered and then randomly down-sampling to 2000 objects.
- (ii) Adopt the combination of an RBF kernel and a white noise kernel as the covariance function to be used for GPR.
- (iii) Train the GPR using a single photometric property (normalized to have mean zero and variance one) as the output or ‘y’ value and the physical galaxy parameters as the ‘x’ values. This training will tune the hyperparameters of the kernel.
- (iv) Perform 1000 random draws from the PDFs that describe the fiducial MW’s properties. This will allow us to incorporate uncertainties in the MW measurements into our results.
- (v) Use GPR to apply the optimized kernel to the training set and predict the photometric property of interest. For each randomly drawn set of physical properties for the MW, we obtain the mean

prediction, predicted variance, and a set of 1000 values drawn from the GPR-predicted PDF corresponding to that position in physical parameter space (which we refer to as a set of samples).

(vi) The mean photometric prediction for the MW is then calculated as the mean of the set of GPR output means at the position of each MW draw. The error on the prediction is calculated as the standard deviation of the values from the complete set of samples generated, allowing us to incorporate both uncertainties associated with the scatter at fixed properties and errors resulting from the uncertainties in MW properties.

The code used to construct the GPR is provided on our GP GitHub page for public use here.<sup>2</sup> At this site, we provide sample code for determining photometry estimates, addressing systematics, and constructing an SED.

## 4 RESULTS

Via GPR predictions for MW photometric properties across the spectrum, we can produce a comprehensive outside-in portrait of the MW SED, allowing comparisons to the colours and luminosities of other galaxies. In this section, we apply a variety of diagnostics from the literature, such as colour–luminosity, colour–mass, and colour–colour diagrams, in order to assess how the MW compares to the broader population. We also construct a multiwavelength SED for the MW and compare our results to templates from the literature.

### 4.1 The MW compared to the broader galaxy population

As discussed in Section 3, we have predicted the MW colours and luminosities based on the six parameters of stellar mass ( $M_*$ ), SFR, axial ratio ( $b/a$ ), bulge-to-total ratio ( $B/T$ ), disc scale length ( $R_d$ ), and bar vote fraction ( $p_{\text{bar}}$ ). In the following colour diagrams, all magnitudes and colours are presented as rest-frame AB magnitudes (evaluating all passbands at redshift zero).

Our quantitative results are summarized in Tables A1–A3. The values provided correspond to the mean rest-frame predictions based on the GPR derived via the methods presented in Section 3, and have been corrected for Eddington bias as described in supplementary Appendix D. Colours and magnitudes are all calculated independently of one another. For example, we use GPR to predict  $^0(g-r)$  galaxy colour directly, as opposed to deriving this value by subtracting the predicted  $^0M_r$  from the predicted  $^0M_g$ . For SDSS photometry, our derived colours are based on `model` magnitudes, as these yield the most accurate colour estimates for SDSS galaxies; however, the absolute magnitudes provided are based on `cmodel` magnitudes, as those most accurately represent the total brightness of an object.

Log-spaced density contours corresponding to the cross-matched galaxy sample of 29 836 galaxies described in Sections 2.1 and 2.2 are plotted in greyscale on all of the following colour-based diagrams. We also overlay red and blue ellipses that denote the rough locus of the red sequence and blue cloud, respectively, in each plot. These shadings are intended to guide the eye and should not be interpreted in a quantitative manner. In a corner of each plot, we provide error bars that correspond to the mean uncertainties in each galaxy property being plotted for the training set.

In each diagram, we also show the locations of the 36 red spiral galaxies selected in Masters et al. (2010b) that overlap with our cross-matched sample (out of 294 in the original catalogue). This

sample of objects was selected based on their colour, presence of spiral features, and shape/structural parameters from SDSS. They are required to have colour  $^0(g-r) > 0.63 - 0.02(^0M_r + 20)$ , overlapping the blue edge of the red sequence. They are also selected to have a spiral likelihood  $p_{\text{spiral}} \geq 0.8$  in the prescription of Bamford et al. (2009), and are required to have visible arms in Galaxy Zoo 1  $p_{\text{CW}} > 0.8$  or  $p_{\text{ACW}} > 0.8$  (Lintott et al. 2011) in order to ensure they have spiral morphology. These objects are also selected to be approximately face-on (equivalent to an axial ratio requirement  $b/a > 0.63$ ), as dust reddening is expected to have a substantial impact on the apparent colours of spirals (Masters et al. 2010a). However, in that paper the axial ratio values were calculated via  $r$ -band isophotal measurements, while ours are determined from an exponential profile fit. Therefore, we apply a profile-fit-based cut of  $b/a > 0.6$  to this sample to enable a more direct comparison to our face-on results for the MW. Finally, Masters et al. (2010b) require that the red spiral sample contains galaxies with an SDSS  $f_{\text{dev}} \leq 0.5$ , where  $f_{\text{dev}}$  is defined as the weight of the de Vaucouleurs profile in the best-fitting linear combination with the exponential profile matched to the object’s image. This ensures that S0 galaxies do not contaminate the sample, although they are already only a small percentage of the GZ1 sample.

The resulting red spiral sample is represented by red points in our plot. We overlay the positions of these objects in each parameter space to help assess the consistency of the inferred properties of the MW with this population. Two objects whose  $^0(g-r)$  colours in the cross-matched catalogue differed by  $> 0.1$  mag from the photometry used in Masters et al. (2010b) due to changes in SDSS pipelines were excluded.

#### 4.1.1 Optical colours

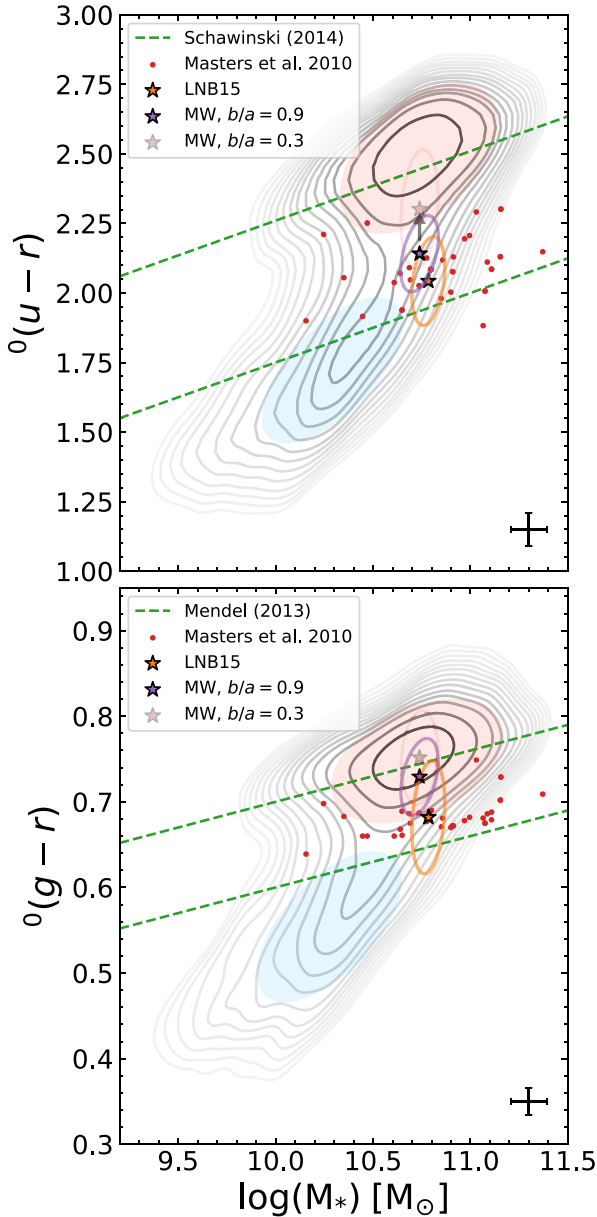
We first present results at optical wavelengths, as they allow us to compare directly to previous work done with MW analogues in LNB15. We focus on the SDSS *ugriz* bands (cf. Section 2.1.1). Fig. 5 presents predictions for MW optical colours as a function of stellar mass ( $M_*$ ) in solar mass units.

The upper panel shows  $^0(u-r)$  colour and the lower panel shows  $^0(g-r)$  colour versus mass. Both panels have overlaid dashed reference lines that can be used to distinguish general regions of the diagrams. The top portion contains galaxies that are on the red sequence, while the middle portion contains green valley galaxies, and the lower portion corresponds to the blue cloud. The dashed lines bracketing the green valley in the upper panel correspond to  $^0(u-r) = -0.24 + 0.25M_*$  and  $^0(u-r) = -0.75 + 0.25M_*$  (Schawinski et al. 2014). In the lower panel, the plotted lines correspond to  $^0(g-r) = 0.6 + 0.06(M_* - 10)$  and  $^0(g-r) = (0.6 + 0.06(M_* - 10)) + 0.1$  (Mendel et al. 2013).

Previous results from LNB15 are plotted in orange. The star represents the mean prediction and the ellipse encompasses the  $1\sigma$  confidence region. We remind the reader that these constraints were determined based only on stellar mass and SFR, along with a cut on axial ratio. In comparison, the results of the six-parameter GPR are plotted in purple. In red, we plot the GPR result evaluated with an axial ratio of  $b/a = 0.3$  rather than 0.9, to illustrate the impact that the assumed inclination has on the inferred SED. The GPR confidence regions are calculated from the covariance between the samples drawn from the regression predictions; the distribution of these samples incorporates both uncertainties in MW properties and scatter in colours at fixed properties (cf. Section 3.2.4). In the lower right corner of each panel, we show the mean error in optical

<sup>2</sup><https://github.com/cfielder/GPR-for-Photometry>





**Figure 5.** Rest-frame optical colour as a function of stellar mass. The greyscale, log-spaced contours depict the density of  $0.03 < z < 0.09$  galaxies in our cross-matched sample, without any limiting magnitude applied. The dashed grey lines correspond to divisions of the galaxy population used in the literature. We expect ‘red sequence’ galaxies to be in the upper portion of each plot, with the ‘blue cloud’ corresponding to the bluest colours. The region between the two lines corresponds to the ‘green valley’ population. For  $^0(u-r)$ , we use the divisions of Schawinski et al. (2014) and in  $^0(g-r)$  we follow Mendel et al. (2013) (cf. Section 4.1.1). The results and  $1\sigma$  confidence region from LNB15 are marked in orange. Our results from applying GPR to all six galaxy physical parameters considered are marked in purple; the  $1\sigma$  region is determined by the covariance between Gaussian process samples. For comparison, in lighter red we show the six-parameter GPR result obtained when setting the axial ratio of the MW to  $b/a = 0.3 \pm 0.1$  rather than 0.9. The stellar masses differ between our prediction and LNB15 due to the updates to the mass estimate for the MW described in Section 2.3. Red points correspond to members of the red spiral galaxy sample of Masters et al. (2010b). In the lower right corner of each panel, we depict error bars representative of the mean uncertainties for galaxies in the comparison sample. Our results are consistent with LNB15 and indicate that at optical wavelengths the MW is redder than the typical star-forming spiral galaxy, in addition to being more massive.

colour and log stellar mass among the galaxies in our final sample. Per-object uncertainties in the optical colours account for roughly half of the total scatter in our GPR colour prediction for a face-on MW. In contrast, the average error in stellar mass in the training sample does not affect the uncertainty in the stellar mass of the MW (it will, however, contribute to Eddington biases, as discussed in supplementary Appendix C). Note that LNB15 did not use the same stellar mass as we do, reflecting our updated estimate for the mass of the MW (cf. Section 2.3).

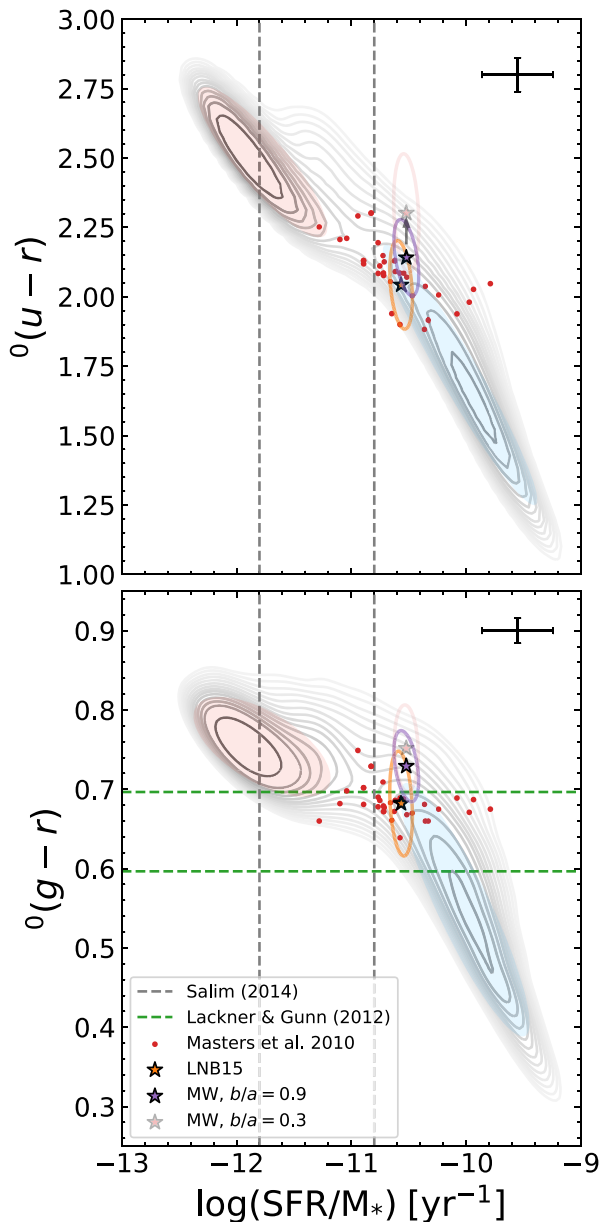
In both  $^0(u-r)$  and  $^0(g-r)$ , our results are consistent with, though marginally redder than, those reported in LNB15. This is no surprise as we do not expect the MW to move far in optical colour space when constraints tighten. Even when we make predictions for a much more inclined MW, our results do not change dramatically in the optical, with shifts well within the uncertainties in both our face-on results and those from LNB15, although the colour does become marginally redder as expected. In the optical, the MW appears to lie in the ‘saddle’ of the galaxy colour bimodality, implying that the MW is redder than the average spiral galaxy in the local Universe in the optical bands. That said, if one were to only consider spiral galaxies of similar mass to our Galaxy, the MW is not as unusually red as it would be if compared to lower mass spiral galaxies.

The green valley (and by extension the galaxy colour bimodality) has been used as a basic tool to distinguish transitional galaxies from the general galaxy population. Transitional galaxies have lower specific SFRs ( $sSFR = SFR/M_*$ ) than a star-forming galaxy of the same mass;  $sSFR$  can be used as a proxy for the evolutionary state of a galaxy and its star forming history. Salim (2014) defines the transitional region in  $sSFR$  space to be below the  $sSFR$  of massive Sbc galaxies, as these Sbc’s are the earliest galaxy type expected to proceed with regular star formation free of quenching, but above the  $sSFR$  at which galaxies appear to no longer be star forming in the UV. As described in that work, this range corresponds to  $-11.8 < \log(SFR/M_*) < -10.8$ . Note that here and throughout this paper log refers to the base 10 logarithm.

In Fig. 6, we plot the same rest-frame colours as in Fig. 5 but as a function of  $\log sSFR$ . The vertical dashed lines denote the transitional region in  $\log SFR/M_*$ , as defined by Salim (2014). The region with  $\log sSFR$  above  $-10.8$  corresponds to galaxies that are actively forming stars while objects with  $\log sSFR$  below  $-11.8$  are quiescent; transitional objects are between them. In the lower panel, the green horizontal lines correspond to the green valley definition of Lackner & Gunn (2012) evaluated with the predicted  $r$ -band absolute magnitude for the MW ( $^0M_r = -20.65$ ). Galaxies residing within this range in  $^0(g-r)$  are expected to reside within the green valley. Galaxies above this designation are expected to lie in or near the red sequence, and galaxies below are expected to lie in or near the blue cloud.

Based on its  $sSFR$ , the MW must lie within the star-forming population, rather than in the transitional range. While one might expect an object that meets optical definitions of the green valley to have a transitional  $sSFR$  this is not necessarily true, as galaxies of different evolutionary states can share the same optical colour (see e.g. Cortese 2012; Salim 2014).

In  $^0(g-r)$ , if we take the green valley to be 0.1 in width, as defined by Mendez et al. (2011), the Galaxy would either fall within or be redder than the green valley, consistent with the results shown in Fig. 5. Despite ongoing star formation, more massive spiral galaxies tend to be redder in the optical than their lower mass counterparts (e.g. Masters et al. 2010b). However, our estimated properties for the MW lie in the middle of the distributions of colours, masses, and  $sSFR$ s of the red spiral sample from Masters et al. (2010b); these



**Figure 6.** Rest-frame optical colour as a function of  $\log \text{sSFR}$  [ $\log (\text{SFR}/M_*)$ ] in units of  $\text{yr}^{-1}$ . As before the greyscale log contours depict the density of the parent cross-matched sample, red points correspond to the red spiral sample of Masters et al. (2010b), the orange star and ellipse correspond to the results of Licquia et al. (2015), and the purple star and ellipse are the results of our GPR analysis. The vertical dashed grey lines designate divisions of galaxy populations according to their sSFRs, following the definitions of Salim (2014): Quiescent galaxies are at left, transitional objects are in the middle, and star-forming objects are at right. The MW lies on the star-forming side of these divisions. The green horizontal lines in the bottom panel correspond to the green valley definition of Lackner & Gunn (2012), evaluated at the  $r$ -band absolute magnitude of the MW. According to the prescription by Mendez et al. (2011), the green valley has a width of 0.1 in  $^0(g-r)$ , leading to the limits shown here. The MW mean value falls above this ‘green valley’ region. While the optical colour of the MW is redder than most star-forming galaxies in the local Universe, based on its sSFR the MW would not be considered a transitional galaxy.

objects are redder in the optical than is typical for even the most massive spirals.

As these plots exemplify, differentiating between galaxy populations based only on optical photometry is challenging. For instance, in the lower panel in Fig. 6 we can see that red sequence, transitional, and star-forming objects can all have colours of  $^0(g-r) \sim 0.7$ . In Fig. 5, the blue cloud becomes difficult to distinguish from the red sequence at high masses as the most massive spirals have lower sSFRs and, therefore, redder colours. In the following subsections, we investigate constraints on the colour of the MW at UV and IR wavelengths where galaxy populations may separate more clearly.

#### 4.1.2 UV colours

We utilize far-UV (FUV) and near-UV (NUV) photometry from GALEX provided in the GSWLC-M2 catalogue (Martin & GALEX Team 2005; Salim et al. 2016, 2018), as discussed in Section 2.1.2. Thermal emission from massive stars with lifetimes  $< 100$  Myr peaks at UV wavelengths, while lower mass, longer lived stars play a larger role at optical wavelengths (Salim 2014; Tuttle & Tonnesen 2020). Because UV radiation is produced by short-lived but high-luminosity stars, it provides a sensitive indicator of *recent* star formation. As a result, UV photometry can more clearly differentiate star-forming from quiescent galaxies than optical measurements can.

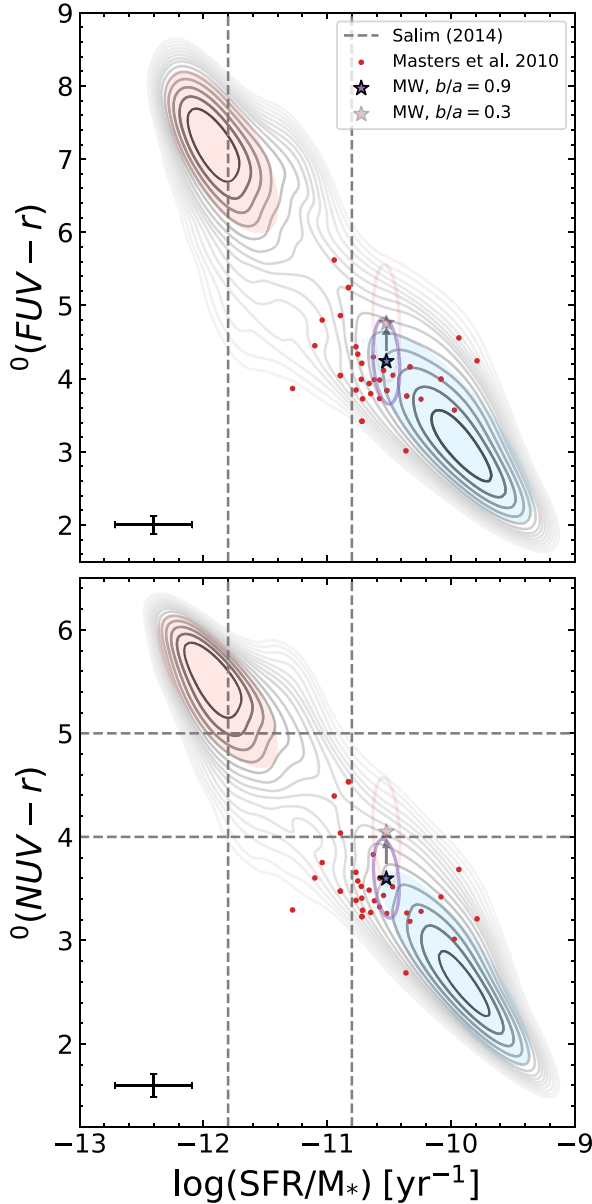
Much as in Section 4.1.1, we can use our GPR results to place the MW on UV-based diagnostic diagrams from the literature. In Fig. 7, we plot  $^0(\text{FUV}-r)$  and  $^0(\text{NUV}-r)$  UV-optical colours versus sSFR. The contours and vertical reference lines shown are defined in the same way as in Fig. 6. For  $^0(\text{NUV}-r)$ , we show horizontal lines corresponding to the ‘green valley’ definition of Salim (2014), bounded at  $4 < ^0(\text{NUV}-r) < 5$ . As before, our face-on MW prediction and the corresponding  $1\sigma$  confidence region are plotted in purple and the inclined MW prediction is plotted in red; unlike in the optical, there are no previous estimates of MW properties in this space that we could plot. Much as in the optical, uncertainties in the UV photometry for individual objects account for roughly half of the total scatter ascribed to our MW UV predictions.

Compared to typical star-forming galaxies in the local Universe, the MW has redder than average UV colours and lower than average sSFR. The MW appears to lie on the blue side of the  $^0(\text{NUV}-r)$  green valley border, in contrast to its location in the optical (cf. Figs 5 and 6). This reflects the limited discriminating power of optical colour; the green valley is only 0.1 mag wide in  $^0(g-r)$ , allowing objects to easily scatter over its borders due to even small photometric errors or inclination effects, but it spans an entire magnitude in  $^0(\text{NUV}-r)$ . A more inclined MW is predicted to be notably redder in the UV than in the optical, so much so that it could be consistent with the UV green valley in colour.

As in the optical, our estimates for the MW in the UV-sSFR plane lie in the middle of the Masters et al. (2010b) red spiral population. Red spiral galaxies tend to lie outside of the UV green valley as they have SFRs comparable to typical blue spirals of the same mass (Cortese 2012), and UV colour is more sensitive to recent SFR than the optical is.

#### 4.1.3 IR/WISE colours

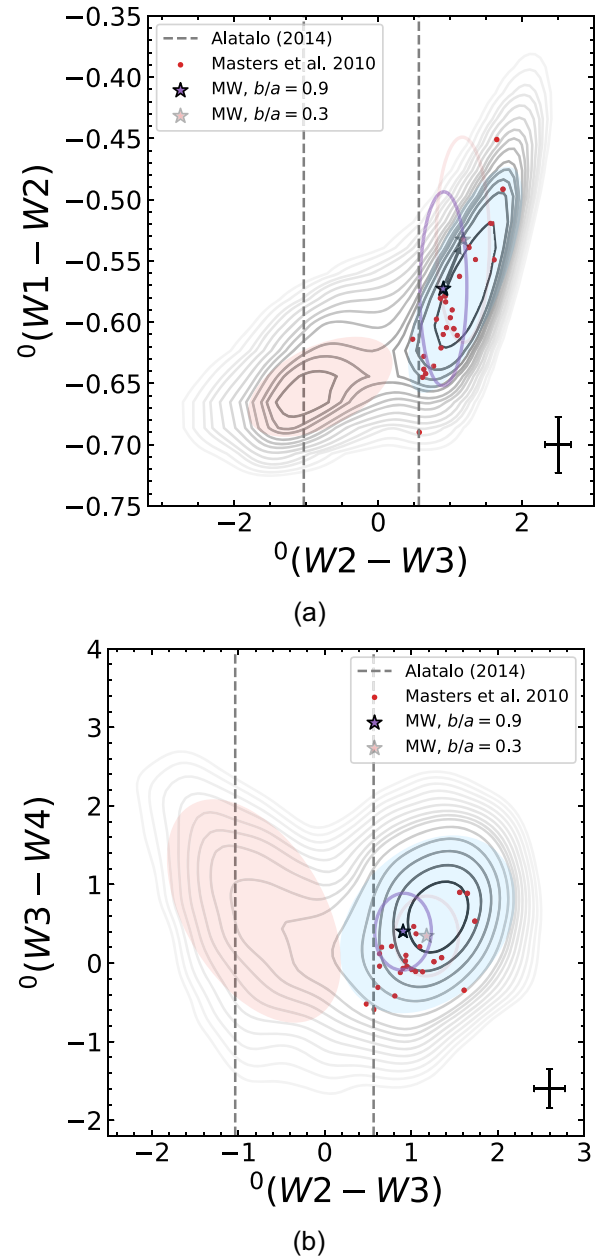
The IR data for our galaxy sample originate from the 2MASS and WISE surveys, as included in the GSWLC-M2 (Salim et al. 2016, 2018) and DESI Legacy catalogues (Dey et al. 2019), respectively (cf. Section 2.1). Similar to in the UV, the IR brightness of a galaxy is



**Figure 7.** As Fig. 6, but for UV-optical colours. The vertical reference lines come from Salim (2014) where objects with  $\log \text{sSFR} > -10.8$  are actively forming stars, objects with  $\log \text{sSFR} < -11.8$  are quiescent, and objects in between are considered transitional. In  $^0(\text{NUV} - r)$ , we also plot the bounds of the UV-optical ‘green valley’ as defined in Salim (2014). Our predictions for the MW show that it lies on the star-forming side of the transitional regions, and is most likely on the blue side of the UV-optical green valley border. Our Galaxy lies in a similar region of these colour spaces to the red spiral sample of Masters et al. (2010b).

sensitive to recent star formation due to re-emission of UV photons absorbed by dust. The IR colours of galaxies also exhibit a colour bimodality, but star-forming galaxies exhibit *redder* IR colours than the passively evolving population, rather than bluer. Instead of the ‘green valley’, the region between the star-forming and quiescent populations in the IR is commonly referred to as the IR transition zone (IRTZ), following Alatalo et al. (2014).

In Figs 8(a) and (b), we plot *WISE* colour-colour diagrams for the cross-matched galaxy sample in addition to the GPR prediction for the MW. The MW colour is poorly constrained in some *WISE* bands



**Figure 8.** *WISE* colour-colour diagrams for both our parent sample ( $\log$  density contours) and the predicted results for the MW from GPR. Reference lines from Alatalo et al. (2014) designate the IRTZ [ $1.035 < ^0(W2 - W3) < 0.565$ ; IRTZ]. The MW appears to lie on the star-forming side of the IRTZ, much closer to the median colours of typical spiral galaxies, in contrast to the UV and IR. Again, our Galaxy lies in a similar region of these colour spaces to the red spiral sample of Masters et al. (2010b).

due to the lower signal-to-noise of these detections. If we compare our covariance ellipse in the  $^0(W1 - W2)$  direction to the average errors in the photometry (lower right error bar), the photometric errors account for a modest fraction of the total uncertainties in our GPR prediction. However, in  $^0(W2 - W3)$  and  $^0(W3 - W4)$  errors in the photometry for individual objects dominate the estimated uncertainties in the MW GPR predictions. The vertical lines in these plots designate the IRTZ from Alatalo et al. (2014); objects with  $^0(W2 - W3) > 0.565$  in AB magnitudes correspond to late-type galaxies, those with  $^0(W2 - W3) < -1.035$  are early-type galaxies, and in between lies the



IRTZ. Magnitudes were converted from Vega to AB magnitudes via the prescription of Jarrett et al. (2011).

As before, we show predictions for the MW if it were approximately face-on or more steeply inclined by evaluating the GPR at different axial ratios. An inclined MW appears to be most notably different in the W3 band, which traces prominent dust emission features, particularly those associated with polycyclic aromatic hydrocarbons (Wright et al. 2010a). It appears that an inclined spiral galaxy would be measured to be more IR bright compared to a face-on counterpart matching it in all other ways; this may represent a systematic effect related to data processing, since the dust emission would be expected to be optically thin. We find similar results in Section 4.2.3.

In both diagrams, the prediction for the MW lies on the star-forming side of the IRTZ. If we compare Fig. 8(a) to the classification scheme in fig. 12 of Wright et al. (2010a) (note that this requires converting our AB magnitudes to Vega magnitudes), the MW lies within the region of colour space they label as typical for spiral galaxies, as would be expected. Similarly, the classification scheme of fig. 11.b of Jarrett et al. (2017) would place the MW in the intermediate disc region, consistent with the expectation for a massive spiral galaxy. Intermediate disc objects are thought to be in transition towards being quenched due to SFRs that are decreasing with time; our estimate for the MW's SFR is slightly below average for a spiral galaxy of the same mass, consistent with this picture.

The IR results mirror what we find from the UV: The MW is still forming enough stars to appear bright in the IR due to re-emission from dust. Galaxies are expected to transition in the optical before they do in the IR (Alatalo et al. 2014; Tuttle & Tonnesen 2020). If we follow the narrative of Alatalo et al. (2014) and Smethurst et al. (2015), the MW may be in the early transition phase from star forming to quiescent. It is brighter and redder in the optical compared to the typical star-forming galaxy. In the UV, the MW is on the blue side of the green valley but near it. In the IR, the MW's inferred colour is more typical for a star-forming galaxy, though uncertainties are substantial. This would track with the expectation that a galaxy transitions in the optical before it does in the IR.

As before, the inferred colours of the MW in the mid-IR are consistent with the range of values for the Masters et al. (2010b) red spiral sample, though some *WISE* bands are not well constraining due to the low signal-to-noise of the underlying measurements used for prediction. Based on this, it remains plausible that the MW is a part of the red spiral population. We discuss how the MW's colours compare to other galaxy populations further in Section 5.2.

## 4.2 The multiwavelength SED of the MW

Thus far, we have focused on predictions for a single MW colour at a time. However, we can assemble colour information across all passbands to construct an SED for the Galaxy. SEDs, which quantify the total energy of emitted photons as a function of wavelength or frequency, are valuable tools in the study of galaxies. Many physical characteristics of galaxies can alter their SEDs – the age of their stellar population, stellar abundances, gas and dust content, ISM chemistry, details of star formation history, and the presence of an AGN can all leave distinct signposts that give observers insight into the formation and evolution of a given galaxy (see e.g. Silva et al. 2011). Because these effects each tends to alter the SED at specific portions of the spectral range, with broad enough wavelength coverage and detailed enough spectral information one can disentangle the dominant processes in a given galaxy.

Detailed modelling of the MW's SED will be the focus of a follow-up paper. In this work, we will present a proof-of-concept for a GPR-

constructed SED for the MW and provide an initial analysis of its properties. In the following subsection, we outline our GPR-based methods for determining the SED of the Galaxy before presenting quantitative results and assessing the effects the galaxy physical parameters used for prediction each has on the SED.

### 4.2.1 Algorithm for calculating the SED for the MW

We work in frequency ( $\nu$ ) space instead of wavelength ( $\lambda$ ) space when calculating the SED of the MW, as SEDs are most typically presented in units of energy per unit frequency. Our algorithm for calculating the SED proceeds as follows:

#### (i) Estimate $^0M_r$ and colours for the MW –

Our SED is calculated in reference to the  $r$  band. Therefore, using the GPR (described in Section 3) we predict the  $r$ -band AB absolute magnitude ( $^0M_r$ ) and all colours with rest-frame  $r$  as the reference band, i.e.  $^0(x-r)$  where  $x$  spans from FUV to W4 [e.g.  $^0(FUV-r), \dots, ^0(W4-r)$ ].

Eddington bias is subtracted off separately from our predicted colours and  $^0M_r$  before we combine them. Similarly, the uncertainty in the Eddington bias is added in quadrature to the uncertainty in the GP calculations (which incorporates both scatter at fixed properties and errors due to the MW property uncertainties; cf. Section 3.2.4). For details on the Eddington bias calculations, refer to supplementary Appendix D.

#### (ii) Calculate flux ratios –

We calculate the flux in each band  $f_{\nu,x}$  relative to the flux in the  $r$ -band  $f_{\nu,r}$  via the relation

$$\log \left( \frac{f_{\nu,x}}{f_{\nu,r}} \right) = \frac{^0(x-r)}{-2.5}, \quad (1)$$

where  $^0(x-r)$  is the Eddington bias-corrected colour in the  $x$  band compared to the  $r$  band, which we have predicted via GPR.

#### (iii) Calculate luminosity –

From the  $r$ -band absolute magnitude combined with the flux ratios, it is straightforward to convert to luminosity. We first calculate the  $r$ -band luminosity as

$$\log(L_{\nu,r}) = \frac{(^0M_r - 34.04)}{-2.5} \left[ \log \left( \frac{W}{Hz} \right) \right], \quad (2)$$

where  $^0M_r$  is the Eddington bias-corrected  $r$ -band absolute magnitude for the MW obtained via GPR. This formula is derived via the relation for converting flux to AB magnitudes in combination with the area of a 10 pc-radius sphere to convert flux to luminosity. The luminosity in any other band can then be calculated via the relation

$$\log(L_{\nu,x}) = \log \left( \frac{f_{\nu,x}}{f_{\nu,r}} \right) - \log(L_{\nu,r}). \quad (3)$$

We can then add log frequency ( $\log \nu$ ) to obtain  $\log \nu L_{\nu,x}$ .

#### (iv) Calculate errors –

To convert uncertainties in magnitudes and colours to uncertainties in  $\log \nu L_{\nu,x}$ , we make use of propagation of errors. First to determine  $\sigma_{\log \nu L_{\nu,r}}$ , we calculate the partial derivative of  $\log \nu L_{\nu,r}$  (which is equation 2 +  $\log \nu$ ) with respect to  $^0M_r$ . This yields

$$\sigma_{\log \nu L_{\nu,r}} = 0.4 \sigma_{^0M_r}. \quad (4)$$

Errors in the other bands are calculated in a similar manner, but they depend on both the error in colour and the error in  $^0M_r$ :

$$\sigma_{\log \nu L_{\nu,x}} = 0.4 \times \sqrt{\sigma_{^0(x-r)}^2 + \sigma_{^0M_r}^2}. \quad (5)$$

**Table 1.** The passbands and corresponding power and uncertainties for the predicted SED of the MW, as plotted in Fig. 9. These values have already had Eddington bias subtracted out.

Passband	$\lambda_{\text{eff}}$ ( $\mu\text{m}$ )	$\log \nu L_\nu$ (log W)	$\sigma_{\log \nu L_\nu}$ (log W)
<i>FUV</i>	0.155	35.53	0.20
<i>NUV</i>	0.2275	35.63	0.20
<i>u</i>	0.354	36.01	0.06
<i>g</i>	0.4750	36.44	0.02
<i>r</i>	0.622	36.62	0.09
<i>i</i>	0.763	36.66	0.01
<i>z</i>	0.905	36.70	0.02
<i>J</i>	1.25	36.67	0.04
<i>H</i>	1.65	36.54	0.05
<i>Ks</i>	2.15	36.39	0.06
<i>W1</i>	3.368	35.59	0.08
<i>W2</i>	4.618	35.58	0.10
<i>W3</i>	12.082	35.61	0.20
<i>W4</i>	22.194	35.49	0.30

In the plots that follow, we do not plot the contribution to errors from  $\sigma_{M_r}$  as it is fully covariant across all bands; as a result, when templates are normalized to match the observed SED, any error in  $^0M_r$  would simply change the normalization. We do provide its value for reference. Thus, the error bars presented in the MW SED plots are equivalent to  $0.4\sigma_{(x-r)}$ .

The colour predictions used to derive the luminosities used for our SED are provided for reference in Table A1 and the value of  $^0M_r$  is provided in the absolute magnitude table (Table A3). We present our estimated luminosities and associated uncertainties, incorporating Eddington bias corrections in each case, in Table 1.

#### 4.2.2 Interpreting the SED of the MW

In Fig. 9, we present our full predicted SED for the MW, along with a variety of empirical template galaxy spectra from the literature. We plot  $\log \nu L_\nu$  (the power emitted per log interval in frequency) on the vertical axis, in units of log Watts. The horizontal axis corresponds to rest-frame wavelength in units of  $\mu\text{m}$ ; each photometric band used is labelled at its effective wavelength along the top of the plot. The black open circles represent the estimates for the MW’s luminosity along with their associated errors, calculated as described in Section 4.2.1. We also show the error bar corresponding to the uncertainty in  $^0M_r$  near the bottom of Fig. 9(a) for reference. The numerical values for the MW SED corresponding to the plotted points are provided in Table 1.

While detailed fitting of the MW’s SED using physical models lies beyond the scope of this paper, we will compare to observed SEDs of individual galaxies and composite galaxy templates from the literature as a sanity check on the realism of our results. Using photometry for extragalactic samples to constrain the SED of the MW, and then comparing the results to observed galaxy photometry (albeit for different objects) are somewhat circular. However, given that our analysis has treated every band completely independently, there were no guarantees that we should get a sensible SED when combining GPR results across the spectrum.

In Fig. 9(a), we compare our predicted MW SED to templates from Benítez et al. (2004), which are refinements to the templates from Coleman et al. (1980) and (for starburst galaxies) Kinney et al. (1996). The Coleman et al. (1980) templates were based on averaging the observed SEDs of relatively blue galaxies of a given morphological type. Given the broad range of observed SEDs for

objects with similar morphological classification, these should not be considered universally applicable for all galaxies of a given type; however, we use the same labelling as Benítez et al. (2004) for consistency with the literature.

In this and successive plots, we have normalized all templates to match the estimated MW SED in the *r* band. We normalize in an optical band as those bands have the smallest errors in the predicted SED; which particular optical band we choose has minimal effect on our comparisons. It is apparent that the MW SED is generally consistent with the (Benítez et al. 2004) ‘Sbc’ galaxy template. The galaxies used to construct this template, M51 (NGC 5194) and NGC 2903, are undergoing moderate amounts of star formation, with log sSFRs of  $-10.1$  and  $-10.4$  (versus  $-10.5$  for the MW) (Muñoz-Mateos et al. 2007). The MW is generally expected to have an SBbc morphological type [see e.g. Bland-Hawthorn & Gerhard (2016), for a recent review on MW structure, as well as Hodge (1983), Kennicutt (2001), and Efremov (2011)], though given how it was constructed, we should not read too much into the agreement of our SED with the Benítez et al. (2004) ‘Sbc’ template in particular. It is clear, however, that the GPR method yields results that resemble composite SED templates from the literature.

Benítez et al. (2004) provide only a sparse set of composite templates that may not match the SEDs of an individual galaxy. Additionally, those templates do not span the full wavelength range of the MW SED we have produced. Thus, we also compare to the set of 129 observed galaxy SEDs from Brown et al. (2014). This SED atlas encompasses a variety of bright galaxies in the very local Universe. Unlike the templates shown in Fig. 9(a), these correspond to SEDs of individual galaxies (not averages) with minimal modelling used to interpolate between photometric and spectroscopic coverage.

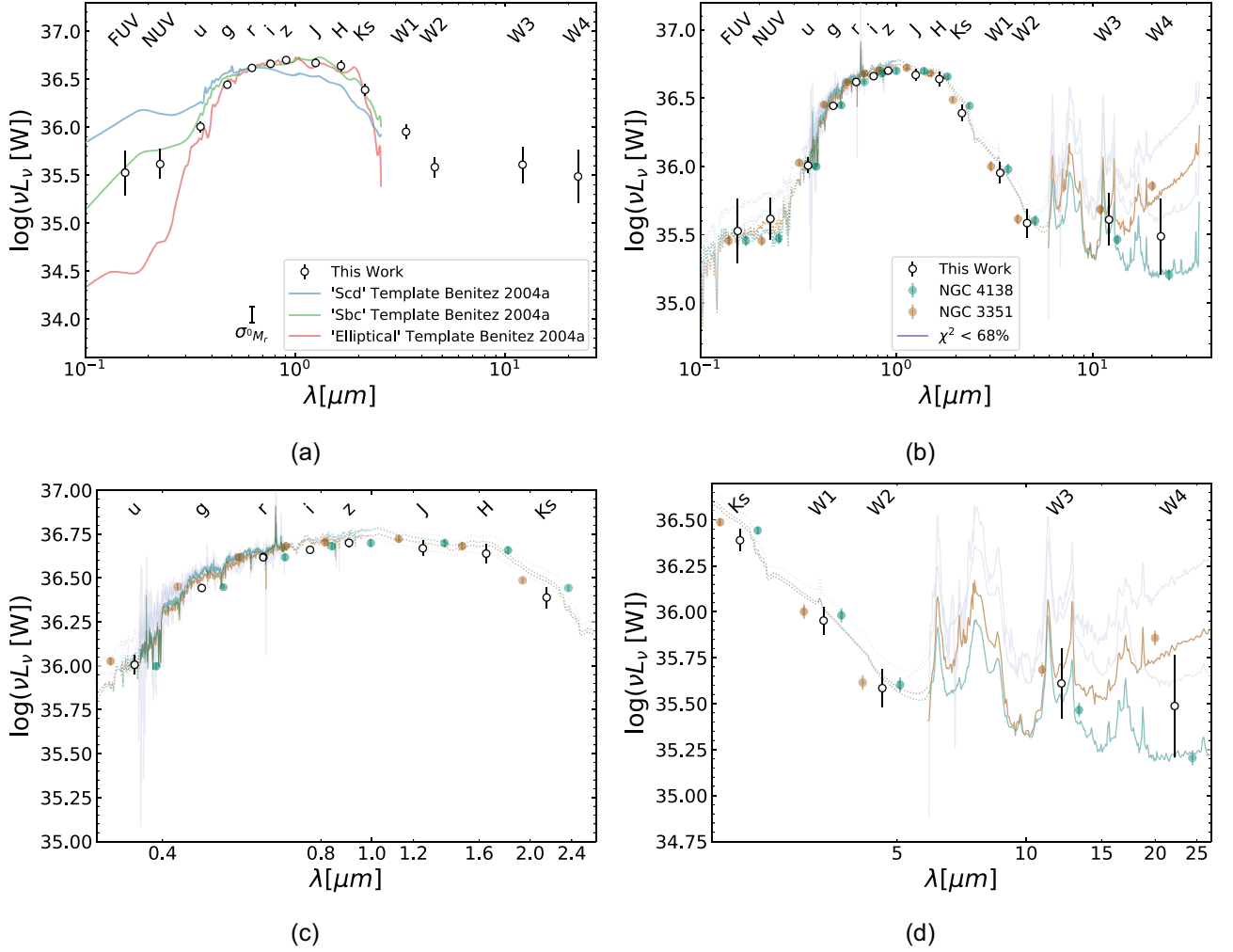
The data tables from Brown et al. (2014) provide extinction-corrected photometry as well as a variety of summary values such as luminosity distance. Because the magnitudes are presented in the AB system, we can use the relation

$$\log f_\nu = \frac{m_{\text{AB}} - 8.9}{-2.5}, \quad (6)$$

to calculate flux, where  $f_\nu$  is the flux in units of Jansky and  $m_{\text{AB}}$  is the observed magnitude in each band. We neglect *k*-corrections as these galaxies are very nearby ( $3.1 < D_L < 249.2$  Mpc at the most extreme), so the corrections are generally negligible. We then can use the flux–luminosity relation  $L_\nu = 4\pi D_L^2 f_\nu$  to calculate the luminosity in each band. Via propagation of errors, the uncertainty in  $\log \nu L_\nu$  for these templates is equivalent to  $\sigma_{\log \nu L_\nu} = 0.4\sigma_{m_{\text{AB}}}$ .

We take a few further steps before comparing the Brown et al. (2014) SEDs to our Galaxy’s. First, we have obtained the axial ratios (as a proxy for inclination) for 89 out of the 129 galaxies in the Brown et al. (2014) sample from the Siena Galaxy Atlas, which have been distributed as part of DESI Legacy imaging surveys DR9. All galaxies with *b/a* below 0.5 or unknown axial ratios were excluded from comparisons. Lower axial ratios should correspond to highly inclined galaxies for which reddening will strongly affect the SED (e.g. Unterborn & Ryden 2008; Maller et al. 2009), making them inappropriate comparisons to our face-on SED for the MW. We then normalize the observed SEDs to match the Galactic SED in the *r* band, as was done for the Benítez et al. (2004) templates before.

Finally, we calculate the  $\chi^2$  difference between our predicted SED for the MW and each of the galaxy SEDs presented in the Brown et al. (2014) atlas. We emphasize that no fitting is performed in this comparison other than matching in the *r* band. We then calculate  $\chi^2$



**Figure 9.** The predicted SED for the MW from a six-parameter GPR, depicted by black open circles in all panels; the method of calculation is described in Section 4.2.1. (a) We compare to empirical galaxy templates from Benítez et al. (2004) (re-calibrated from Coleman, Wu & Weedman 1980) normalized to match the MW in the  $r$  band. The MW SED is consistent with the ‘Sbc’ template from this set, which is labelled as such because it was originally based on the average of the SEDs of two blue galaxies of morphological type Sbc. (b) We compare the SED of the MW to the most closely matched templates from Brown et al. (2014), which are based on spectra and model fits to bright nearby galaxies. We show all templates whose  $\chi^2$  values in comparison to the MW SED are below the 95 per cent upper limit value of  $\chi^2$  for 13 degrees of freedom; again, we normalize in the  $r$  band. The two galaxies with the smallest  $\chi^2$  (which fall below the 68 per cent upper limit) have their photometry plotted as round points, which have been offset in the wavelength direction for clarity. Their accompanying spectra from the SED atlas are also plotted, without an offset. We also show as fainter curves the SEDs for the remaining two galaxies with potentially matching SEDs, which have  $\chi^2$  values below the 68 per cent limit. Images of the four best-fitting galaxies are provided in Fig. 10. Portions of the spectra from the SED atlas that are based on models are depicted using dotted lines, while those that are directly based on observations are shown using solid lines. (c) The optical and near-IR portion of the MW SED, with Brown et al. (2014) templates, as in (b). (d) The mid-IR portion of (b). The GPR method produces an MW SED that is consistent in shape with both composite SED templates and individual observed SEDs.

using log quantities, as that is the space in which we perform our predictions and for which errors are (by construction) symmetric:

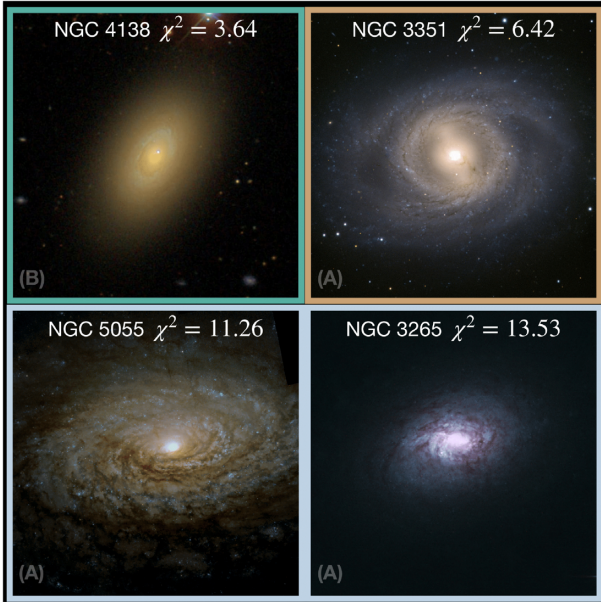
$$\chi^2 = \sum \left( \frac{\log vL_v^{\text{atlas}} - \log vL_v^{\text{MW}}}{\sigma_{\log vL_v}} \right)^2, \quad (7)$$

where  $\sigma_{\log vL_v}$  combines in quadrature the total error in the MW SED for a given band, the uncertainties in the Brown et al. (2014) photometry, and  $\log_{10}$  (1.1), which corresponds to a 10 per cent error in  $vL_v$ . This extra error is added to account for systematic uncertainties in the photometry for a given band relative to others; if this were not included, optical bands would dominate the  $\chi^2$  value due to their small nominal uncertainties. We calculate  $\chi^2$  using the 14 bandpasses in which we have measured the MW’s predicted SED,

which yields 13 total degrees of freedom (one is lost due to the  $r$ -band normalization performed).

Fig. 9(b) overplots the Brown et al. (2014) SEDs for galaxies whose  $\chi^2$  values fall within the 68 per cent upper limits for a  $\chi^2$  distribution with 13 degrees of freedom (corresponding to  $\chi^2 = 14.8$ ). We also examine galaxies that fall below the 95 per cent limit (with  $\chi^2 < 22.4$ ), but exclude them from plotting for brevity. Objects below the 68 per cent upper limit (four in total) are clearly consistent with the SED of the MW, while those between the 68 and 95 per cent limits (comprising three objects) are in some tension with the SED of the MW, but could still be a match. While we calculate  $\chi^2$  using the broad-band photometry for each galaxy, we also plot the full SED from Brown et al. (2014) for each of these galaxies for reference.





**Figure 10.** Postage stamp images corresponding to galaxies within the Brown et al. (2014) SED atlas with  $\chi^2$  values below the 95 per cent upper limit when compared to the estimated SED of the MW. At the top of each image, we indicate the galaxy NGC number as well as the  $\chi^2$  value for the comparison of its SED to the Galactic one; we present them in order from smallest to largest  $\chi^2$ . At the bottom left of each tile is a letter marked (A) or (B) that denotes the source of the given image. (A) images are from ESA/Hubble (2021) and (B) images are from SDSS (2021). The border around each image matches the colour used for its SED in Fig. 9(b). It is evident that galaxies that may have similar SEDs to the MW exhibit a broad range of visual morphologies, further emphasizing the lack of a unique mapping between morphology and galaxy SED.

In Fig. 9(b), the two galaxies with the smallest  $\chi^2$  are labelled and plotted with the highest opacity in teal and gold. For these two galaxies, we also plot the observed photometry as points offset in the wavelength direction so they are easier to compare to the MW values. The higher  $\chi^2$  objects are plotted with low opacity in pale blue. We also provide more detailed plots of two separate ranges of the SEDs. The optical through near-IR regime is depicted in Fig. 9(c), while the near- to mid-IR SED is depicted in Fig. 9(d). Portions of the spectra that are based on observations are plotted with solid lines, while modelled portions are plotted with dotted lines.

We also provide images for the four Brown et al. (2014) atlas galaxies that fall below the 68 per cent limit in Fig. 10. At the top of each postage stamp image, we list the galaxy NGC number, as well as the  $\chi^2$  value for comparing photometry for that galaxy to our MW SED. The tiles are presented in order from smallest to largest  $\chi^2$  value. At the bottom left of each tile is a letter marking (A) or (B) that refers to the source for each given image. (A) images are from ESA/Hubble (2021) and (B) images are from SDSS (2021). The borders that surround each image match the colour coding in Fig. 9(b).

It is clear from this comparison that our GPR method produces results whose spectral shape is comparable to observed galaxy SEDs. The small total number of SEDs within the Brown et al. (2014) atlas that are consistent with the MW is no surprise; after cutting on inclination, we are reduced to only 70 objects, the majority of which are early-type galaxies, leaving a limited number of examples to cover the full range of star-forming galaxies. We emphasize that there remains a need for more careful investigation and modelling of

the MW SED we have obtained; this will be the topic of a follow-up paper.

However, we will briefly comment on the galaxies from the Brown et al. (2014) atlas whose SEDs are most consistent with the MW. NGC 4138 is a Hubble type SA(r)0 galaxy that contains an AGN and a star-forming ring. Using estimates for mass from Jore, Broeils & Haynes (1996) ( $2.92 \times 10^{10} M_{\odot}$ ) and Kassin (2004) ( $6.23 \times 10^{10} M_{\odot}$ ) and the SFR estimates from Wiegert & English (2014) ( $0.14 M_{\odot} \text{ yr}^{-1}$ ) and Brown et al. (2017) ( $0.2 M_{\odot} \text{ yr}^{-1}$ ) yields a log sSFR of  $\sim(-11.3)-(-11.5)$ , significantly smaller than the MW value ( $-10.52$ ). NGC 4138 has a  $^0(g-r)$  colour of  $0.73 \pm 0.05$  (Brown et al. 2014), matching the MW value.

In contrast, NGC 3351/M95 is a galaxy of Hubble type SB(r)b, versus SBb or SBc for the Galaxy; it contains a pseudo-bulge (Sandage & Bedke 1994; Fisher, Drory & Fabricius 2009; Brown et al. 2014), much as the MW is conjectured to possess (see Bland-Hawthorn & Gerhard 2016, and references therein). According to measurements provided by Leroy et al. (2008) and George et al. (2019), NGC 3351 has a log sSFR of  $-10.43$  per year ( $\text{SFR} = 0.940 M_{\odot} \text{ yr}^{-1}$ ), which is comparable to the log sSFR of the MW of  $-10.52$ . According to the measurements compiled by Brown et al. (2014), NGC 3351 has a  $^0(g-r)$  colour of  $0.74 \pm 0.05$ , very similar to that of the MW's ( $0.73 \pm 0.05$ ; see Table A1). It matches the MW SED equally well as NGC 4138 at most wavelengths, with the exception of the longest wavelength W3 and W4 bands. However, we note that the photometry for NGC 3351 from Dale et al. (2017) does not show the strong red slope in the mid-IR seen in the SED atlas spectrum. If the Dale et al. (2017) measurements were used in the mid-IR, the agreement with the MW SED would be significantly better.

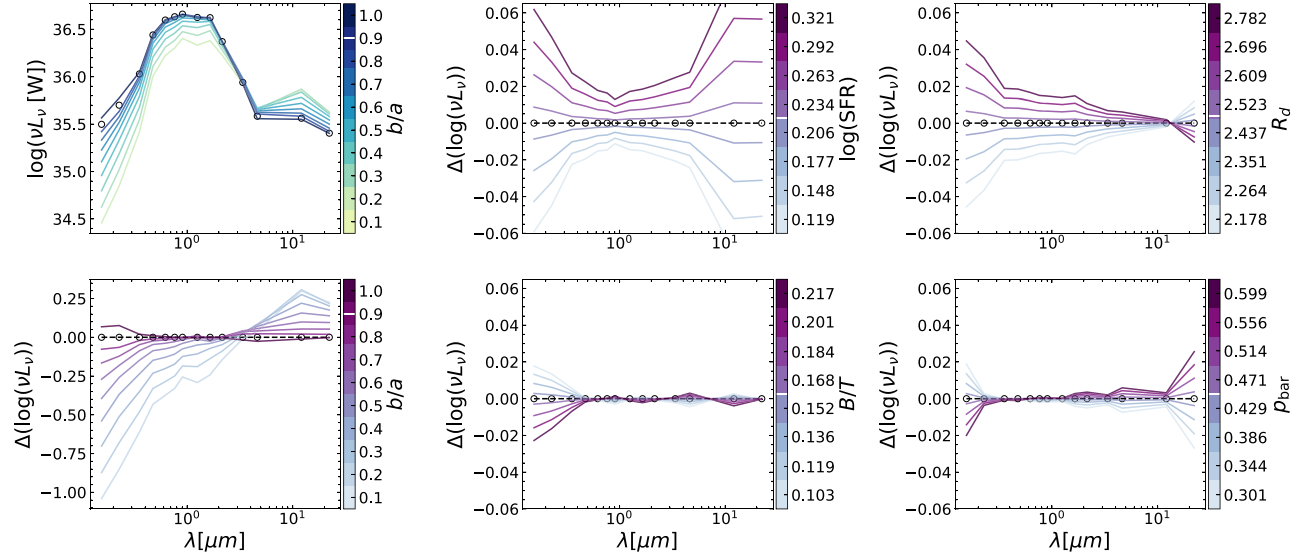
Two other galaxies in the atlas have SEDs for which the  $\chi^2$  value when compared to the MW SED is below the 68 per cent significance level. NGC 5055 is a galaxy of Hubble type SAbc (Brown et al. 2014) with log sSFR of  $-10.53$  as tabulated in Kennicutt et al. (2011). NGC 3265 is of Hubble type SA(rs)0 pec (Ann, Seo & Ha 2015) with log sSFR of  $-9.12$  (Kennicutt et al. 2011). The set of objects whose SEDs are consistent with the MW's spans a diverse range of visual morphologies.

#### 4.2.3 Impact of physical parameters on the estimated SED of the MW

Thus far, we present results based on a set of fiducial values (with uncertainties) for the MW's stellar mass, SFR, axial ratio, disc scale length, bulge-to-total mass ratio, and bar presence (presented in Section 2.3). Because of the generality of the GPR fit, we can vary each of these parameters one at a time and test its impact on the inferred SED.

In our previous analyses, we randomly drew values from the distributions of MW properties and made predictions based on each of those draws, which were then combined, as described in Section 3. This allowed us to incorporate the uncertainties in the MW's physical properties into our results. In this subsection, however, we will neglect these uncertainties in order to isolate the effect of changing the central values for each parameter. We have performed a similar analysis to the one presented below by sampling from MW uncertainties as a cross-check; the results are very similar, so we do not present them here for brevity.

In this analysis, we keep the values for the five MW parameters *not* being studied at their fiducial mean. We then choose a discrete set of values for the sixth parameter at which to evaluate the SED



**Figure 11.** Isolating the contributions of each physical parameter to the SED. Each panel shows the effect of varying one parameter while fixing the other five parameters to the fiducial MW values (see Section 2.3). We vary most parameters over a range of  $\pm 2\sigma$  from the MW’s fiducial values. In the case of axial ratio, we explore a wider range of values to convey better the impact of inclination on the observed SED. The upper left panel depicts the different SEDs that would be inferred for different galaxy axial ratios, with the expected results that an inclined MW would look much fainter than face-on. In the remainder of the panels, we depict the log of the predicted SED divided by the SED evaluated for fiducial MW parameters in order to make differences clearer; the plotted quantity is thus  $\Delta(\log(vL_v)) = \log(\frac{vL_v}{vL_v^{\text{MW}}})$ . The predicted SED for the MW value is plotted with a dashed black line and open circles at the passbands. On the colour bars, the MW’s value is marked by a horizontal white line. In  $b/a$ , the GP captures the effects of dust reddening at higher inclinations. If the MW were to have a higher SFR, the SED would be brighter in both the UV and IR. If the MW’s disc were more extended, we would observe an increased UV brightness and decreased mid-IR brightness. We caution the reader that predictions for disc scale lengths more than  $1\sigma$  below the MW value may not be reliable. Relatively few galaxies of the mass of MW have sizes smaller than the MW, causing the GPR to be poorly trained for small  $R_d$  values (Licquia et al. 2016). Changes in the MW’s  $B/T$  on the SED would be minimal. As we decrease the intensity of the MW’s bar, it seems to have a minimal effect, with a slightly increased UV brightness. In general, these follow our expectations of galaxy evolution that we discuss further in Section 4.2.3.

via GPR. In the case of SFR, bulge-to-total mass ratio, and bar vote fraction, we select eight values that lie evenly spaced between  $\pm 2\sigma$  of the fiducial mean value for the MW (inclusive), in addition to evaluating at the nominal value. For axial ratio, we step through a wide range of possible galaxy axial ratios instead of focusing around 0.9 in order to capture the full effects of inclination on the Galaxy SED. We have excluded mass from this exercise, as changing mass would radically alter the normalization of the predicted SED.

For each of the values that we step through for the given test parameter, the GPR is evaluated as before. We apply a  $12\sigma$  cut-off on the training sample for all parameters *except* the one being varied. By focusing on one parameter at a time, we can explore the impact each has on the predicted SED for the MW and can assess whether the GP is able to capture the expected correlations between galaxy properties.

The results of this analysis are presented in Fig. 11. In the colour bars for each panel, the lighter shades correspond to smaller values for the parameter being varied and darker shades correspond to larger values. In all panels, the fiducial value for the MW is marked by a horizontal white line on the colour bar.

The upper left panel shows the GPR-predicted SED for each axial ratio value considered, with axes similar to Fig. 9. The SED evaluated with  $b/a = 0.9$ , the fiducial axial ratio value used for the MW, is marked by open points. In this panel, one can see the effects of inclination reddening first hand, an effect that has been seen repeatedly in analyses in spiral galaxies (e.g. Shao et al. 2007; Unterborn & Ryden 2008; Maller et al. 2009; Masters et al. 2010b). The cross-section for dust extinction and scattering generally increases with decreasing wavelength, causing reddening effects to

be the strongest at the shortest wavelengths. Thus, we expect the SEDs of galaxies to appear redder the more inclined they are (Xiao et al. 2012). The increased attenuation at higher inclinations for galaxies in the training sample causes the GPR to predict a redder SED as the inclination increases (lower  $b/a$ ). We find a decreased brightness in the UV/optical and an increased brightness in the IR as discs are viewed more edge-on, an effect also observed by SED modellers (see e.g. Noll et al. 2009).

In the remainder of the panels, we plot SEDs *divided* by the SED evaluated at the fiducial MW values, as effects are more subtle and would be difficult to discern in an un-normalized plot. We include a plot of this type based on varying axial ratio in the lower left panel, though we use a larger y-range than the other normalized plots due to the large dynamic range spanned.

An SED that matches the prediction for the fiducial MW values exactly would fall along the horizontal line at  $\log(vL_v) - \log(vL_v^{\text{MW}}) = 0$ . The photometric predictions for the MW nominal parameters hence correspond to the open circles along this line.

SFR has the clearest effect on the SED, as seen in the top middle panel. The amount of UV flux is a sensitive indicator of star formation as it is dominated by hot, massive, short-lived stars. These stars contribute to the flux at optical and near-IR wavelengths, but are subdominant there; however, in the mid-IR the SED responds strongly to star formation due to light from hot stars that is reprocessed by dust. The GPR predictions reflect all of these phenomena.

The upper right panel depicts the effect of disc scale length on the predicted SED. We note that the predictions become unstable at shorter scale lengths due to the small number of MW-mass galaxies

with radius smaller than our Galaxy in the training samples (cf. Licquia et al. 2016), so results for disc scale lengths more than  $1\sigma$  below the MW value may not be robust. Outside of that regime, it is clear that MW-like galaxies with shorter scale lengths exhibit significantly less flux in the UV than those with longer scale lengths when  $M_*$ , SFR, etc. are all held fixed, along with smaller effects at optical–IR wavelengths.

A smaller disc, with other properties held fixed, could imply that the gas within the disc is denser and dust columns are correspondingly greater. This would in turn cause the SED to look fainter in the UV relative to the IR compared to if the disc were more extended. We have tested the effect of varying  $b/a$  and  $R_d$  simultaneously and find that we can compensate for a change in one with a change in the other almost perfectly. Given the long-standing scenario that inclination effects on the SEDs of disc galaxies are driven by dust (e.g. Maller et al. 2009), it is reasonable to hypothesize that the effects of varying  $R_d$  must relate to varying dust impact as well.

The bottom middle panel shows the effect of varying only the bulge-to-total ratio. Overall, the impact on the SED is small. We see that if the MW were to have a more massive bulge, it would appear slightly fainter in the UV. This could reflect the fact that bulges have older stellar populations than spiral discs; the effect may be subtle here as to first order the effect is captured by variation in sSFR. Bulges are also susceptible to dust reddening (Tuffs et al. 2004; Driver et al. 2007), so a larger bulge may also suffer from greater reddening effects at short wavelengths.

The lower right panel shows the variation in SEDs as we change the bar vote fraction. As this quantity increases, we see a decrease in UV brightness and an increase in mid-IR brightness. In general, we speculate that most effects caused by a bar may have been captured in the variation with other measured properties (particularly SFR). The observed trend with  $p_{\text{bar}}$  could cohere with a narrative of bar-induced star formation suppression (see e.g. Tubbs 1982; Haywood et al. 2016), which would affect the UV bands the most and lead to a galaxy looking redder once the bar instability has caused the consumption of cold gas in the disc (Masters et al. 2010b). However, this would not explain why galaxies with stronger bars are brighter in the mid-IR while simultaneously being fainter in the UV. Alternatively, one could speculate that processes associated with bars could modulate the properties of interstellar dust (e.g. by affecting the ability of gas to cool and form molecular clouds); having more dust (at fixed  $M_*$ , SFR, and inclination) should cause a reduction of flux in the UV and an increase in flux in the mid-IR, as observed here.

#### 4.2.4 Exploration of other sources of physical parameter measurements

As described in Sections 2.2.1 and 2.2.2, there exist multiple options for the values used for the stellar mass, SFR, and bulge-to-total ratio for SDSS galaxies. In this subsection, we discuss the impact on our results of using different measurements of galaxy parameters or different methods of defining our training samples.

**4.2.4-1 Stellar mass and SFR** The GSWLC-M2 catalogue (Salim et al. 2016, 2018) includes estimates of stellar masses and SFRs computed based on the photometry within the catalogue. While the  $M_*$  values in the GSWLC-M2 catalogue closely match those from the MPA-JHU catalogue (Brinchmann et al. 2004) used for our main results, the SFRs presented in the GSWLC-M2 catalogue are far less bimodal than those presented in the MPA-JHU catalogue; a significant number of galaxies would be classified as star forming in

GSWLC that would be considered quiescent based on the MPA-JHU catalogue.

We have produced an MW SED using the GSWLC-M2 stellar masses and SFRs as features in place of the MPA-JHU values, but otherwise following the same methods used to produce the results in Sections 4.1 and 4.2. The impact on predictions in the optical and IR is small. The effect is more notable in the UV, where the MW is predicted to be closer to the mean colour of star-forming galaxies with the same sSFR [corresponding to smaller values of  $^0(FUV - r)$  and  $^0(NUV - r)$  in Fig. 7]. Even so, the predicted UV colours are still within  $0.5\sigma$  of those resulting from using MPA-JHU  $M_*$  and SFRs. In addition to being brighter in the NUV and FUV, the predicted SED is also marginally fainter in W3 and W4 compared to the results presented in Fig. 9, with shifts that are again well below  $1\sigma$  in each band.

Overall, we find that using the stellar masses and SFRs derived from the GSWLC-M2 catalogue instead of the MPA-JHU catalogue has little effect on our predictions for the MW, and is subdominant to other sources of uncertainty.

**4.2.4-2 Bulge-to-total ratio** There are also multiple options for which band to measure bulge-to-total ratios; the Simard et al. (2011) catalogue contains bulge and disc decompositions performed both in the  $g$  and  $r$  bands. For our main results, we use the  $r$ -band value,  $B/T_r$ , but have also tested the impact of instead using  $B/T_g$  on our results presented in Sections 4.1 and 4.2.

Overall, we find only very small effects on predicted colours from changing to  $B/T_g$ . All results are well within a few hundredths of a magnitude of the previous values, except in the case of  $^0(u - r)$ . For that value, the predicted value for the MW from the GPR trained on  $B/T_g$  is larger than when we use  $B/T_r$  (corresponding to redder colour; cf. Fig. 6), though still within  $0.5\sigma$  of our primary result. As a consequence, the predicted SED for the MW using  $B/T_g$  instead of  $B/T_r$  is almost identical to before.

**4.2.4-3 Treatment of GZ2 votes** As discussed in Section 2.2.3, using citizen science votes from GZ2 requires consideration of responses to previous questions that influence whether the question of interest is even asked. If we wished to select a pure sample of barred galaxies, it would be necessary to ensure not only that a high fraction of people who were asked whether a bar is present voted in the affirmative, but also that a substantial total number of people voted on each of the preceding questions in order to minimize errors in vote fractions (see e.g. Willett et al. 2013). However, with GPR we may gain more information about trends that influence photometric properties by including a broader set of objects, so we do not necessarily wish to exclude non-barred or non-spiral galaxies from the training sample. We explore here how changing the treatment of GZ2 votes influences our GPR predictions.

We have tested what impact restricting the training set to only barred, face-on spirals would have by testing how predicted colours for the MW vary when using training samples with a variety of different constraints: (1) a control sample without any restrictions based on GZ2 vote results; (2) a sample where if the number of votes on whether or not a galaxy has a bar,  $N_{\text{bar}}$ , is less than 10 we set  $p_{\text{bar}} = 0$ ; (3) a sample using the bar selection cuts from Willett et al. (2013) table 3, column 3, and rows 2 and 3 in addition to the vote count thresholds mentioned in Section 2.2.3; or (4) a sample using the same cuts as Willett et al. (2013) *except* setting  $p_{\text{bar}} = 0$  when  $N_{\text{bar}} < 10$ , rather than rejecting objects with low  $N_{\text{bar}}$  from the set entirely.



Applying all of the Willett et al. (2013) cuts reduces the size of the training sample by roughly an order of magnitude, degrading the ability of GPR to predict colours and increasing net errors. Furthermore, requiring  $N_{\text{bar}} \geq 10$  not only shrinks the size of the sample but also greatly biases the luminosity distribution of the training sample compared to a volume-limited sample, which may result in biases in inferred photometry. We have explored how the GPR-predicted MW colours change for each of these four training sample definitions (but otherwise using the methodologies described in Section 3). The results from (2) compared to our fiducial case, (1), are nearly identical, so the particular values of  $p_{\text{bar}}$  assigned to objects with poorly constrained vote counts cannot have had a large systematic impact on our predicted MW photometry. We also find that restricting to training sample (3) or (4) yields much ( $\gtrsim 2 \times$ ) larger errors on all predictions. Results from (3) and (4) are still within  $1\sigma$  of those from (1) and (2), however. Therefore, we conclude that with GPR we get better predictions when we include more objects (including some with noisier vote fractions) than when we instead restrict training to just the best-constrained objects. Therefore, we perform *no* GZ2-based cuts on the galaxy sample used for training, and instead include objects spanning the full range of bar, face-on, and feature vote fractions in the training set.

## 5 SUMMARY AND CONCLUSIONS

### 5.1 Summary

In this work, we have set out to estimate a full SED for the MW, spanning wavelengths from the UV to the IR. Our central motivation is twofold: (1) to improve our understanding of how the MW compares to the general galaxy population and by doing so (2) guide the tuning of parameters in simulations in order to create more realistic galaxies.

The previous work by Licquia et al. (2015) constrained the optical colours and luminosity of the MW using MW analogue galaxies selected based on their stellar mass and SFR, obtaining the best constraints on the MW’s photometric properties available previous to this work. Here, we have been able to reduce the uncertainties on these constraints further by incorporating information from additional parameters such as disc scale length and bulge-to-total ratio, which also connect to a galaxy’s evolutionary history (Cappellari 2016; Saha & Cortesi 2018), and have for the first time developed predictions for MW photometry at wavelengths beyond the optical.

We have shown that the MW analogue method breaks down when we attempt to match the Galaxy in many physical parameters; the number of MW analogues rapidly approaches zero in higher dimensional spaces (cf. Fig. 1). Expanding to a wider wavelength range requires information from data sets that do not cover the full SDSS footprint, making the problem worse. We instead have predicted the photometric properties of the MW using GPR, which provides an optimal means of interpolating information from a limited training set. We have performed a series of tests throughout this paper that have demonstrated that GPR is able to produce realistic and reliable photometric predictions.

We have compared predictions for the MW to the broader local galaxy population in colour–mass, colour–sSFR, and colour–colour diagrams. As exemplified by Fig. 5, we obtain similar results in the optical to those reported by Mutch et al. (2011) and LNB15, though with reduced errors, further confirming that the MW has optical colours consistent with the green valley population. For the first time, we have also predicted UV (Fig. 7) and IR colours (Fig. 8) for the MW, which provide more sensitive diagnostics of the evolutionary

status of a galaxy. We find that in both these regimes the MW appears to lie on the star-forming side of the green valley.

In this work, we have determined the luminosity and colours of the MW for *GALEX FUV* and *NUV*, SDSS *ugriz*, 2MASS *JHKs*, and *WISE W1 – W4* bands in an entirely self-consistent way, giving us unprecedented constraints on its SED. We have constructed the first multiwavelength SED for the MW. This SED has a shape consistent with both composite galaxy templates (Fig. 9a) and observed SEDs of individual galaxies (Fig. 9b). The GPR method produces a realistic SED with errors and captures previously known galaxy property correlations, such as those between reddening in spiral galaxies and viewing angle or between SFR and UV and IR flux (Fig. 11). High-resolution hydrodynamical simulators (e.g. Guedes et al. 2011; Sawala et al. 2016; Wetzel et al. 2016) no longer have to compare their mocks of the MW blindly to photometric constraints from broad galaxy populations that span a wide range of properties. Rather, it should now be possible to tune the treatment of star formation efficiency, threshold gas density for star formation, and dust properties to produce galaxies that match the photometric properties of the MW directly, while simultaneously exploiting those properties that we can measure well from inside the Galaxy.

### 5.2 Discussion: the MW as a red spiral

As previously suggested in a variety of works (e.g. Salim 2014; Schawinski et al. 2014; Licquia et al. 2015), definitions of the green valley that rely only on optical bands may lead to misleading conclusions. The MW has an sSFR that is higher than the canonical values for green valley galaxies,  $\log \text{sSFR} = -10.52$  as compared to  $-11.8 < \log \text{sSFR} < -10.8$  for transitioning galaxies from Salim (2014), even though it has red optical colours for a star-forming object. However, at UV and IR wavelengths the colours of the MW more clearly place it among the star-forming population. This combination of red optical colours when viewed face-on with significant star formation evident at UV and IR wavelengths is characteristic of the previously identified population of red spiral galaxies.

A population of red spiral galaxies in clusters was first identified by van den Bergh (1976). Since then, these ‘passive spirals’ have been identified at a range of redshifts and in multiple data sets. As noted by Cortese (2012), for galaxies with a stellar mass above  $10^{10} M_{\odot}$  like the MW ( $M_{*} = 5.48^{+1.18}_{-0.94} \times 10^{10} M_{\odot}$ ), the blue cloud and red sequence overlap in their optical colours (this is also consistent with findings by Salim 2014). This makes optical photometry a poor choice for constraining the star formation activity for galaxies like our own. However, these massive objects still exhibit a distinct colour bimodality in the UV, as shown by Wyder et al. (2007) and Salim (2014) and is evident from comparing Figs 6 and 7. In comparison to their lower mass counterparts, massive galaxies produced the great majority of their stars at earlier epochs (e.g. Boselli et al. 2001). This causes the optical colours of massive galaxies to be dominated by relatively old stellar populations as opposed to probing recent star formation activity (Wyder et al. 2007; Chilingarian & Zolotukhin 2012; Cortese 2012). Hao et al. (2019) and Zhou et al. (2020) provide evidence that this is the case for red spirals. Direct or re-radiated light from young stars still dominates the red spiral SEDs at UV and IR wavelengths, however.

Reflecting that, both Cortese (2012) and Smethurst et al. (2015) find that red spiral galaxies tend to be UV bright; this can be driven by a relatively small amount of total star formation. In order to facilitate comparison of the MW to the red spiral galaxy population, we have overplotted the red spirals from the Masters et al. (2010b) catalogue

(based on a GZ2 and optical colour selection) that are also part of our cross-matched galaxy catalogue on all colour–mass, colour–sSFR, and colour–colour diagrams presented here. In each diagram, the MW falls near the middle of the red spiral population.

Cortese (2012) notes that 85–90 per cent of objects in their red sample maintain SFRs of  $\sim 1 M_{\odot} \text{ yr}^{-1}$ ; Masters et al. (2010b) found that red spiral galaxies selected from GZ2 typically had lower rates of ongoing star formation than blue spirals of the same mass, but still non-negligible. For comparison, the MW has an SFR of  $1.65 \pm 0.19 M_{\odot} \text{ yr}^{-1}$  (Licquia & Newman 2015), while the average SFR of galaxies of approximately the same mass as the MW ( $\pm 0.3$  in log stellar mass) with  $B/T < 0.75$  (to exclude ellipticals) within our cross-matched galaxy sample is  $1.69 M_{\odot} \text{ yr}^{-1}$ ; the Galaxy is very close to average in this respect.

Masters et al. (2010b) also find that red spiral galaxies have a significantly higher bar fraction compared to blue spirals of the same mass: 70 per cent versus 27 per cent. This matches with the clear evidence that the MW possesses a bar (e.g. Blitz & Spergel 1991; Bland-Hawthorn & Gerhard 2016; Shen & Zheng 2020). Masters et al. (2010b) note that one possible evolutionary scenario for red spirals is bar-driven gas inflows. This removes gas from the outer disc and funnels it into central star formation, which in turn causes the disc to appear more and more red over time (Masters et al. 2012; Saintonge et al. 2012; Cheung et al. 2013; Fraser-McKelvie et al. 2020).

Masters et al. (2011) and Fraser-McKelvie et al. (2020) find that barred spirals tend to have redder colours than their unbarred counterparts. It thus may be the case that the bar has played a role in the colours that we observe for the MW in this work. For example, bar quenching may play a role in the development of red spirals, as Bamford et al. (2009) find many high-stellar-mass red spirals in the field and Smethurst et al. (2015) note that field red spirals most likely evolve primarily via secular evolution due to the lack of nearby galaxies. In the case of the MW, work by Haywood et al. (2016) and Khoperskov et al. (2018) finds that the bar may have played a substantial role in the star formation history of the MW (leading to a significant decrease in star formation 9–10 Gyr ago, and thereby causing the observed pattern of chemical abundances in the disc). Although the effect of bar vote fraction in Fig. 11 is small, it may be that the effects of a bar are primarily captured by other parameters (e.g. SFR).

We note that Evans, Parker & Roberts (2018) studied a population somewhat similar to red spirals, which they labelled ‘red misfits’. Evans et al. (2018) define this population as corresponding to objects with  $\log(\text{sSFR}) > -10.8$  and rest-frame  $g - r > 0.67$  (i.e. sSFR measured to be above the value for the saddle point in the bimodal distribution and colour redder than the saddle point in the colour bimodality). Based on these divisions, the MW almost certainly meets this definition (which is less stringent than most red spiral classifications).

### 5.3 Outlook

As seen in Fig. 10, there is a significant diversity in the set of galaxies that have SEDs consistent with the MW, given the measurement uncertainties in both our results and the Brown SED atlas. The goal of this paper has been to construct the MW’s UV-to-IR SED to enable comparisons to samples of external galaxies and to improve the tuning of simulations. However, in a follow-up paper we will fit the estimated SED of the MW using population synthesis models to obtain more detailed constraints on how the star formation history, dust reddening properties, and metallicity of the

Galaxy would be interpreted from outside (see e.g. Conroy 2013). This will require proper treatment of covariances between different photometric bands; we will address this by employing multi-output GPR in this future work.

The longest wavelength *WISE* W3 and W4 bands could have substantial discriminating power on what SEDs are consistent with the MW’s, if they only had smaller errors, as is evident in Fig. 9. However, currently these bands are poorly measured compared to the optical or near-IR; for most objects used in training the MW SED, the signal-to-noise ratio in these bands is below 1. Given the low effects of dust extinction in these bands, investigation of the flux ratio (or colour) in these bands across the all-sky *WISE* imaging, potentially combining modelling of smooth components of the MW with mapping of the contributions from dust, may provide an alternative method to constrain the colour of the MW at the longest wavelengths. If luminosities in the W3 and W4 bands can be measured relative to the luminosity in W2, long-wavelength measurements could be effectively anchored well to the SED presented here; measuring such relative quantities should be affected less by modelling uncertainties than absolute measurements would be.

The SED presented in the paper (or future improved versions) can be used to identify multiwavelength MW analogue galaxies by matching in unresolved photometric properties. If we do not need to require detailed morphological measurements or citizen science inspection of images, it would greatly increase the size of the parent catalogues that could be used to identify MWAs, which could be useful for a variety of follow-up studies such as determining gas masses for the MW or studying environments of MW-like galaxies.

The MW appears to be atypical in its satellite population and mass assembly history. For example, Evans et al. (2020) find that the assembly history of the MW is only reproduced in 0.65 per cent of MW-mass EAGLE galaxies. This assembly history should be closely related to the local environment surrounding our Galaxy, and environment has been found to play a key role in the formation and evolution of galaxies (e.g. van den Bosch et al. 2008; Peng et al. 2012; Bluck et al. 2016).

In future work, we plan to explore how incorporating measures of galaxy environment (e.g. measures of the local overdensity of galaxies) within a GPR model affects the predicted SED of the MW. The noisiness of environment measures (Hogg et al. 2004) and the impact of SDSS fibre collisions on Local Group-like systems (as typically only one galaxy out of two close neighbours would be observed, causing analogues of an MW–M31 pair to be missed) may limit the information that may be gained from this, however. While we anticipate the environment to have a small impact on the Galactic SED compared to the dominant effects of stellar mass and SFR on galaxy colours (e.g. Grützbauch et al. 2011), assessing the local environments of the most MW-like galaxies may allow us to explore and to what extent our Galaxy’s environment has shaped its exhibited characteristics.

GPR can be useful for a variety of studies beyond the photometric estimates for the MW considered here. For this reason, the authors have provided their analysis code on our project GitHub for full public access for adaption to any other project, under a CC BY-SA 4.0 license.

### ACKNOWLEDGEMENTS

The authors thank Amelia Fraser-McKelvie for helpful discussion and providing her own *WISE* *k*-correction code that we used as a starting point before developing our own. We thank Michael Brown and Richard Beare for helpful discussions related to the *k*-correction

methods applied in this paper. We also thank Rachel Benzanson for helpful discussions and suggestions throughout the course of this work. Lastly, the authors would like to thank the referee for helpful comments and suggestions.

We gratefully acknowledge support from NASA (National Aeronautics Space Administration) Astrophysics Data Analysis Programme grant number 80NSSC19K0588 that made this research possible.

This work made use of PYTHON, along with many community-developed or maintained software packages, including IPYTHON (Perez & Granger 2007), JUPYTER ([jupyter.org](https://jupyter.org)), MATPLOTLIB (Hunter 2007), NUMPY (van der Walt, Colbert & Varoquaux 2011), PANDAS (Wes McKinney 2010), SCIKIT-LEARN (Pedregosa et al. 2011), and SCIPY (Virtanen et al. 2020). This research made use of NASA’s Astrophysics Data System for bibliographic information.

This paper was based in part on observations made with the NASA *Galaxy Evolution Explorer*. GALEX is operated for NASA by the California Institute of Technology under NASA contract NAS5-98034.

This publication makes use of data products from SDSS-IV. Funding for the SDSS-IV has been provided by the Alfred P. Sloan Foundation, the U.S. Department of Energy Office of Science, and the Participating Institutions.

SDSS-IV acknowledges support and resources from the Center for High Performance Computing at the University of Utah. The SDSS website is [www.sdss.org](http://www.sdss.org).

SDSS-IV is managed by the Astrophysical Research Consortium for the Participating Institutions of the SDSS Collaboration including the Brazilian Participation Group, the Carnegie Institution for Science, Carnegie Mellon University, Center for Astrophysics | Harvard & Smithsonian, the Chilean Participation Group, the French Participation Group, Instituto de Astrofísica de Canarias, The Johns Hopkins University, Kavli Institute for the Physics and Mathematics of the Universe (IPMU)/University of Tokyo, the Korean Participation Group, Lawrence Berkeley National Laboratory, Leibniz Institut für Astrophysik Potsdam (AIP), Max-Planck-Institut für Astronomie (MPIA Heidelberg), Max-Planck-Institut für Astrophysik (MPA Garching), Max-Planck-Institut für Extraterrestrische Physik (MPE), National Astronomical Observatories of China, New Mexico State University, New York University, University of Notre Dame, Observatório Nacional/MCTI, The Ohio State University, Pennsylvania State University, Shanghai Astronomical Observatory, United Kingdom Participation Group, Universidad Nacional Autónoma de México, University of Arizona, University of Colorado Boulder, University of Oxford, University of Portsmouth, University of Utah, University of Virginia, University of Washington, University of Wisconsin, Vanderbilt University, and Yale University.

This publication makes use of data products from the Two Micron All Sky Survey, which is a joint project of the University of Massachusetts and the Infrared Processing and Analysis Center/California Institute of Technology, funded by the National Aeronautics and Space Administration and the National Science Foundation.

This publication makes use of data products from the *Wide-field Infrared Survey Explorer*, which is a joint project of the University of California, Los Angeles, and the Jet Propulsion Laboratory/California Institute of Technology, funded by the National Aeronautics and Space Administration.

The Legacy Surveys consist of three individual and complementary projects: the Dark Energy Camera Legacy Survey (DECaLS; Proposal ID #2014B-0404; PIs: David Schlegel and Arjun Dey), the Beijing–Arizona Sky Survey (BASS; NOAO Prop. ID #2015A-

0801; PIs: Zhou Xu and Xiaohui Fan), and the Mayall z-band Legacy Survey (MzLS; Prop. ID #2016A-0453; PI: Arjun Dey). DECaLS, BASS and MzLS together include data obtained, respectively, at the Blanco telescope, Cerro Tololo Inter-American Observatory, NSF’s NOIRLab; the Bok telescope, Steward Observatory, University of Arizona; and the Mayall telescope, Kitt Peak National Observatory, NOIRLab. The Legacy Surveys project is honored to be permitted to conduct astronomical research on Iolkam Du’ag (Kitt Peak), a mountain with particular significance to the Tohono O’odham Nation. NOIRLab is operated by the Association of Universities for Research in Astronomy (AURA) under a cooperative agreement with the National Science Foundation.

BASS is a key project of the Telescope Access Programme (TAP), which has been funded by the National Astronomical Observatories of China, the Chinese Academy of Sciences (the Strategic Priority Research Programme ‘The Emergence of Cosmological Structures’ Grant # XDB09000000), and the Special Fund for Astronomy from the Ministry of Finance. The BASS is also supported by the External Cooperation Programme of Chinese Academy of Sciences (Grant # 114A11KYSB20160057), and Chinese National Natural Science Foundation (Grant # 11433005).

The Legacy Survey team makes use of data products from the *Near-Earth Object Wide-field Infrared Survey Explorer* (NEOWISE), which is a project of the Jet Propulsion Laboratory/California Institute of Technology. NEOWISE is funded by the National Aeronautics and Space Administration.

The Legacy Surveys imaging of the DESI footprint is supported by the Director, Office of Science, Office of High Energy Physics of the U.S. Department of Energy under Contract No. DE-AC02-05CH1123, by the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility under the same contract, and by the U.S. National Science Foundation, Division of Astronomical Sciences under Contract No. AST-0950945 to NOAO. This project used data obtained with the Dark Energy Camera (DECam), which was constructed by the Dark Energy Survey (DES) collaboration. Funding for the DES Projects has been provided by the U.S. Department of Energy, the U.S. National Science Foundation, the Ministry of Science and Education of Spain, the Science and Technology Facilities Council of the United Kingdom, the Higher Education Funding Council for England, the National Center for Supercomputing Applications at the University of Illinois at Urbana–Champaign, the Kavli Institute of Cosmological Physics at the University of Chicago, Center for Cosmology and Astro-Particle Physics at the Ohio State University, the Mitchell Institute for Fundamental Physics and Astronomy at Texas A&M University, Financiadora de Estudos e Projetos, Fundacao Carlos Chagas Filho de Amparo, Financiadora de Estudos e Projetos, Fundacao Carlos Chagas Filho de Amparo a Pesquisa do Estado do Rio de Janeiro, Conselho Nacional de Desenvolvimento Científico e Tecnológico and the Ministerio da Ciencia, Tecnologia e Inovacao, the Deutsche Forschungsgemeinschaft, and the Collaborating Institutions in the Dark Energy Survey. The Collaborating Institutions are Argonne National Laboratory, the University of California at Santa Cruz, the University of Cambridge, Centro de Investigaciones Energeticas, Medioambientales y Tecnologicas-Madrid, the University of Chicago, University College London, the DES-Brazil Consortium, the University of Edinburgh, the Eidgenössische Technische Hochschule (ETH) Zurich, Fermi National Accelerator Laboratory, the University of Illinois at Urbana–Champaign, the Institut de Ciències de l’Espai (IEEC/CSIC), the Institut de Física d’Altes Energies, Lawrence Berkeley National Laboratory, the Ludwig Maximilians Universität München and the associated Excellence



Cluster Universe, the University of Michigan, NSF's NOIRLab, the University of Nottingham, the Ohio State University, the University of Pennsylvania, the University of Portsmouth, SLAC National Accelerator Laboratory, Stanford University, the University of Sussex, and Texas A&M University.

## DATA AVAILABILITY

The data used in this article are provided publicly on our catalogue GitHub. We include both the original volume-limited sample of LNB15 and the cross-matched final sample used in this work. The ReadMe also provides a detailed description of the columns within these tables, which is also provided by each respective catalogue that our data originate from.

## REFERENCES

- Abramson L. E., Williams R. J., Benson A. J., Kollmeier J. A., Mulchaey J. S., 2014, *ApJ*, 793, 49
- Aihara H. et al., 2011, *ApJS*, 193, 29
- Alatalo K., Cales S. L., Appleton P. N., Kewley L. J., Lacy M., Lisenfeld U., Nyland K., Rich J. A., 2014, *ApJ*, 794, L13
- Ann H. B., Seo M., Ha D. K., 2015, *ApJS*, 217, 27
- Baldry I. K., Balogh M. L., Bower R., Glazebrook K., Nichol R. C., 2004, in Allen R. E., Nanopoulos D. V., Pope C. N., eds, AIP Conf. Proc. Vol. 743, The New Cosmology: Conference on Strings and Cosmology. Am. Inst. Phys., New York, p. 106
- Bamford S. P. et al., 2009, *MNRAS*, 393, 1324
- Beare R., Brown M. J. I., Pimbblet K., 2014, *ApJ*, 797, 104
- Behroozi P. S., Marchesini D., Wechsler R. H., Muzzin A., Papovich C., Stefanon M., 2013, *ApJ*, 777, L10
- Bell E. F., de Jong R. S., 2001, *ApJ*, 550, 212
- Bell E. F. et al., 2004, *ApJ*, 608, 752
- Benítez N. et al., 2004, *ApJS*, 150, 1
- Bland-Hawthorn J., Gerhard O., 2016, *ARA&A*, 54, 529
- Blanton M. R., Roweis S., 2007, *AJ*, 133, 734
- Blitz L., Spergel D. N., 1991, *ApJ*, 379, 631
- Bluck A. F. L. et al., 2016, *MNRAS*, 462, 2559
- Boardman N. et al., 2020a, *MNRAS*, 491, 3672
- Boardman N. F. et al., 2020b, *MNRAS*, 498, 4943
- Bocquet S., Heitmann K., Habib S., Lawrence E., Uram T., Frontiere N., Pope A., Finkel H., 2020, *ApJ*, 901, 5
- Boselli A., Gavazzi G., Donas J., Scodeggio M., 2001, *AJ*, 121, 753
- Bovy J., Rix H.-W., 2013, *ApJ*, 779, 115
- Brinchmann J., Charlot S., White S. D. M., Tremonti C., Kauffmann G., Heckman T., Brinkmann J., 2004, *MNRAS*, 351, 1151
- Brown M. J. I. et al., 2014, *ApJS*, 212, 18
- Brown M. J. I. et al., 2017, *ApJ*, 847, 136
- Bundy K. et al., 2015, *ApJ*, 798, 7
- Cappellari M., 2016, *ARA&A*, 54, 597
- Cardelli J. A., Clayton G. C., Mathis J. S., 1989, *ApJ*, 345, 245
- Carlesi E., Hoffman Y., Gottlöber S., Libeskind N. I., Knebe A., Yepes G., Pilipenko S. V., 2020, *MNRAS*, 491, 1531
- Chapelon S., Contini T., Davoust E., 1999, *A&A*, 345, 81
- Cheung E. et al., 2013, *ApJ*, 779, 162
- Chilingarian I. V., Zolotukhin I. Y., 2012, *MNRAS*, 419, 1727
- Cho J., Park C., 2009, *ApJ*, 693, 1045
- Coleman G. D., Wu C. C., Weedman D. W., 1980, *ApJS*, 43, 393
- Conroy C., 2013, *ARA&A*, 51, 393
- Conroy C., Schiminovich D., Blanton M. R., 2010, *ApJ*, 718, 184
- Cortese L., 2012, *A&A*, 543, A132
- Dale D. A. et al., 2017, *ApJ*, 837, 90
- Devereux N., 1987, *ApJ*, 323, 91
- Dey A. et al., 2019, *AJ*, 157, 168
- Díaz-García S., Moyano F. D., Comerón S., Knapen J. H., Salo H., Bouquin A. Y. K., 2020, *A&A*, 644, A38
- Driver S. P., Popescu C. C., Tuffs R. J., Liske J., Graham A. W., Allen P. D., de Propris R., 2007, *MNRAS*, 379, 1022
- Efremov Y. N., 2011, *Astron. Rep.*, 55, 108
- ESA/Hubble, 2021, ESA/Hubble Image Site, available at <https://esahubble.org/images/>
- Evans F. A., Parker L. C., Roberts I. D., 2018, *MNRAS*, 476, 5284
- Evans T. A., Fattahi A., Deason A. J., Frenk C. S., 2020, *MNRAS*, 497, 4311
- Faber S. M. et al., 2007, *ApJ*, 665, 265
- Fielder C. E., Mao Y.-Y., Newman J. A., Zentner A. R., Licquia T. C., 2019, *MNRAS*, 486, 4545
- Fisher D. B., Drory N., Fabricius M. H., 2009, *ApJ*, 697, 630
- Fraser-McKelvie A., Merrifield M., Aragón-Salamanca A., 2019, *MNRAS*, 489, 5030
- Fraser-McKelvie A. et al., 2020, *MNRAS*, 499, 1116
- Gaia Collaboration, 2018, *A&A*, 616, A1
- Gallazzi A., Charlot S., Brinchmann J., White S. D. M., Tremonti C. A., 2005, *MNRAS*, 362, 41
- Galloway M. A. et al., 2015, *MNRAS*, 448, 3442
- George K., Joseph P., Mondal C., Subramanian S., Subramanian A., Paul K. T., 2019, *A&A*, 621, L4
- Gillessen S., Eisenhauer F., Trippe S., Alexander T., Genzel R., Martins F., Ott T., 2009, *ApJ*, 692, 1075
- Gladders M. D., Oemler A., Dressler A., Poggianti B., Vulcani B., Abramson L., 2013, *ApJ*, 770, 64
- Gordon T. A., Agol E., Foreman-Mackey D., 2020, *AJ*, 160, 240
- Görtler J., Kehlbeck R., Deussen O., 2019, preprint ([arXiv:1905.01127](https://arxiv.org/abs/1905.01127))
- Gravity Collaboration, 2019, *A&A*, 625, L10
- Grützbauch R., Chuter R. W., Conselice C. J., Bauer A. E., Bluck A. F. L., Buitrago F., Mortlock A., 2011, *MNRAS*, 412, 2361
- Guedes J., Callegari S., Madau P., Mayer L., 2011, *ApJ*, 742, 76
- Hao C.-N., Shi Y., Chen Y., Xia X., Gu Q., Guo R., Yu X., Li S., 2019, *ApJ*, 883, L36
- Hart R. E. et al., 2016, *MNRAS*, 461, 3663
- Hawarden T. G., Mountain C. M., Leggett S. K., Puxley P. J., 1986, *MNRAS*, 221, 41P
- Hawarden T. G., Huang J. H., Gu Q. S., 1996, in Buta R., Crocker D. A., Elmegreen B. G., eds, ASP Conf. Ser. Vol. 91, IAU Colloq. 157: Barred Galaxies. Astron. Soc. Pac., San Francisco, p. 54
- Haywood M., Lehnert M. D., Di Matteo P., Snaith O., Schultheis M., Katz D., Gómez A., 2016, *A&A*, 589, A66
- Hodge P. W., 1983, *PASP*, 95, 721
- Hodges J. L., Lehmann E. L., 1963, *Ann. Math. Stat.*, 34, 598
- Hogg D. W. et al., 2004, *ApJ*, 601, L29
- Hummel E., van der Hulst J. M., Kennicutt R. C., Keel W. C., 1990, *A&A*, 236, 333
- Hunter J. D., 2007, *Comput. Sci. Eng.*, 9, 90
- Jarrett T. H. et al., 2011, *ApJ*, 735, 112
- Jarrett T. H. et al., 2017, *ApJ*, 836, 182
- Jore K. P., Broeils A. H., Haynes M. P., 1996, *AJ*, 112, 438
- Kassin S. A. J., 2004, PhD thesis, Ohio State Univ., Columbus, OH
- Kauffmann G. et al., 2003, *MNRAS*, 341, 33
- Kelson D. D. et al., 2014, *ApJ*, 783, 110
- Kennicutt R. C. J., 2001, in Woodward C. E., Bica M. D., Shull J. M., eds, ASP Conf. Ser. Vol. 231, Tetons 4: Galactic Structure, Stars and the Interstellar Medium. Astron. Soc. Pac., San Francisco, p. 2
- Kennicutt R. C. et al., 2011, *PASP*, 123, 1347
- Khoperskov S., Haywood M., Di Matteo P., Lehnert M. D., Combes F., 2018, *A&A*, 609, A60
- Kinney A. L., Calzetti D., Bohlin R. C., McQuade K., Storchi-Bergmann T., Schmitt H. R., 1996, *ApJ*, 467, 38
- Kourkchi E., Tully R. B., Neill J. D., Seibert M., Courtois H. M., Dupuy A., 2019, *ApJ*, 884, 82
- Krishnarao D. et al., 2020, American Astronomical Society Meeting Abstracts, Vol. 235, #182.02
- Kroupa P., Weidner C., 2003, *ApJ*, 598, 1076
- Kruk S. J. et al., 2018, *MNRAS*, 473, 4731
- Lackner C. N., Gunn J. E., 2012, *MNRAS*, 421, 2277

- Leroy A. K., Walter F., Brinks E., Bigiel F., de Blok W. J. G., Madore B., Thornley M. D., 2008, *AJ*, 136, 2782
- Licquia T. C., Newman J. A., 2015, *ApJ*, 806, 96
- Licquia T. C., Newman J. A., 2016, *ApJ*, 831, 71
- Licquia T. C., Newman J. A., Brinchmann J., 2015, *ApJ*, 809, 96 (LNB15)
- Licquia T. C., Newman J. A., Bershadsky M. A., 2016, *ApJ*, 833, 220
- Lintott C. et al., 2011, *MNRAS*, 410, 166
- McKinney W., 2010, in van der Walt S., Millman J., eds, Proc 9th Python Sci. Conf.: SciPy 2010. SciPy, Austin, Texas, p. 56
- Majewski S. R. et al., 2017, *AJ*, 154, 94
- Maller A. H., Berlind A. A., Blanton M. R., Hogg D. W., 2009, *ApJ*, 691, 394
- Martell S. L. et al., 2017, *MNRAS*, 465, 3203
- Martin C., GALEX Team, 2005, in Colless M., Staveley-Smith L., Stathakis R. A., eds, Proc. IAU Symp. 216, Maps of the Cosmos. Astron. Soc. Pac., San Francisco, p. 221
- Martinet L., Friedli D., 1997, *A&A*, 323, 363
- Masters K. L., Giovanelli R., Haynes M. P., 2003, *AJ*, 126, 158
- Masters K. L. et al., 2010a, *MNRAS*, 404, 792
- Masters K. L. et al., 2010b, *MNRAS*, 405, 783
- Masters K. L. et al., 2011, *MNRAS*, 411, 2026
- Masters K. L. et al., 2012, *MNRAS*, 424, 2180
- Mendel J. T., Simard L., Ellison S. L., Patton D. R., 2013, *MNRAS*, 429, 2212
- Mendez A. J., Coil A. L., Lotz J., Salim S., Moustakas J., Simard L., 2011, *ApJ*, 736, 110
- Morselli L., Renzini A., Popesso P., Erfanianfar G., 2016, *MNRAS*, 462, 2355
- Muñoz-Mateos J. C., Gil de Paz A., Boissier S., Zamorano J., Jarrett T., Gallego J., Madore B. F., 2007, *ApJ*, 658, 1006
- Mutch S. J., Croton D. J., Poole G. B., 2011, *ApJ*, 736, 84
- Noll S., Burgarella D., Giovannoli E., Buat V., Marcellac D., Muñoz-Mateos J. C., 2009, *A&A*, 507, 1793
- Pedregosa F. et al., 2011, *J. Mach. Learn. Res.*, 12, 2825
- Peng Y.-j., Lilly S. J., Renzini A., Carollo M., 2012, *ApJ*, 757, 4
- Perez F., Granger B. E., 2007, *Comput. Sci. Eng.*, 9, 21
- Pompea S. M., Rieke G. H., 1990, *ApJ*, 356, 416
- Rasmussen C. E., Williams C. K. I., 2006, *Gaussian Processes for Machine Learning*. MIT Press, Cambridge, MA
- Ruchti G. R. et al., 2015, *MNRAS*, 450, 2874
- Saha K., Cortesi A., 2018, *ApJ*, 862, L12
- Saintonge A. et al., 2012, *ApJ*, 758, 73
- Salim S., 2014, *Serb. Astron. J.*, 189, 1
- Salim S. et al., 2007, *ApJS*, 173, 267
- Salim S. et al., 2016, *ApJS*, 227, 2
- Salim S., Boquien M., Lee J. C., 2018, *ApJ*, 859, 11
- Sandage A., Bedke J., 1994, *The Carnegie Atlas of Galaxies*, Vol. 638. Carnegie Inst. Wash., Washington, DC
- Sawala T. et al., 2016, *MNRAS*, 457, 1931
- Schawinski K. et al., 2014, *MNRAS*, 440, 889
- Schlegel D. J., Finkbeiner D. P., Davis M., 1998, *ApJ*, 500, 525
- SDSS, 2021, SDSS Image Gallery, available at <https://www.sdss.org/index.php>
- Shao Z., Xiao Q., Shen S., Mo H. J., Xia X., Deng Z., 2007, *ApJ*, 659, 1159
- Shen J., Zheng X.-W., 2020, *Res. Astron. Astrophys.*, 20, 159
- Silva L. et al., 2011, *MNRAS*, 410, 2043
- Simard L. et al., 2002, *ApJS*, 142, 1
- Simard L., Mendel J. T., Patton D. R., Ellison S. L., McConnachie A. W., 2011, *ApJS*, 196, 11
- Smethurst R. J. et al., 2015, *MNRAS*, 450, 435
- Strateva I. et al., 2001, *AJ*, 122, 1861
- Strauss M. A. et al., 2002, *AJ*, 124, 1810
- Taylor E. N. et al., 2015, *MNRAS*, 446, 2144
- Tubbs A. D., 1982, *ApJ*, 255, 458
- Tuffs R. J., Popescu C. C., Völk H. J., Kylafis N. D., Dopita M. A., 2004, *A&A*, 419, 821
- Tuttle S. E., Tonnesen S., 2020, *ApJ*, 889, 188
- Unavane M., Wyse R. F. G., Gilmore G., 1996, *MNRAS*, 278, 727
- Unterborn C. T., Ryden B. S., 2008, *ApJ*, 687, 976
- van de Sande J. et al., 2018, *Nat. Astron.*, 2, 483
- van den Bergh S., 1976, *ApJ*, 206, 883
- van den Bosch F. C., Aquino D., Yang X., Mo H. J., Pasquali A., McIntosh D. H., Weinmann S. M., Kang X., 2008, *MNRAS*, 387, 79
- van der Walt S., Colbert S. C., Varoquaux G., 2011, *Comput. Sci. Eng.*, 13, 22
- Vera M., Alonso S., Coldwell G., 2016, *A&A*, 595, A63
- Virtanen P. et al., 2020, *Nat. Methods*, 17, 261
- Weiner B. J. et al., 2005, *ApJ*, 620, 595
- Wetzel A. R., Hopkins P. F., Kim J.-h., Faucher-Giguère C.-A., Kereš D., Quataert E., 2016, *ApJ*, 827, L23
- Wiegert T., English J., 2014, *New Astron.*, 26, 40
- Willett K. W. et al., 2013, *MNRAS*, 435, 2835
- Wright E. L. et al., 2010a, *AJ*, 140, 1868
- Wyder T. K. et al., 2007, *ApJS*, 173, 293
- Xiao T., Wang T., Wang H., Zhou H., Lu H., Dong X., 2012, *MNRAS*, 421, 486
- York D. G. et al., 2000, *AJ*, 120, 1579
- Zhou S., Li C., Hao C.-N., Guo R., Mo H., Xia X., 2020, *ApJ*, 916, 38

## SUPPORTING INFORMATION

Supplementary data are available at *MNRAS* online.

Please note: Oxford University Press is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

This paper has been typeset from a  $\text{\LaTeX}$  file prepared by the author.