

University of Louisville

## ThinkIR: The University of Louisville's Institutional Repository

---

Faculty Scholarship

---

3-2-2023

### WALLABY Pilot Survey: hydra cluster galaxies UV and H I morphometrics

Benne W. Holwerda

University of Louisville, benne.holwerda@louisville.edu

Frank Bigiel

Albert Bosma

Helene M. Courtois

Nathan Deg

*See next page for additional authors*

Follow this and additional works at: <https://ir.library.louisville.edu/faculty>



Part of the [External Galaxies Commons](#)

---

#### Original Publication Information

Benne W Holwerda and others, WALLABY Pilot Survey: hydra cluster galaxies UV and H I morphometrics, *Monthly Notices of the Royal Astronomical Society*, Volume 521, Issue 1, May 2023, Pages 1502–1517, <https://doi.org/10.1093/mnras/stad602>

#### ThinkIR Citation

Holwerda, Benne W.; Bigiel, Frank; Bosma, Albert; Courtois, Helene M.; Deg, Nathan; Dénes, Helga; Elagali, Ahmed; For, Bi-Qing; Koribalski, Baerbel; Leahy, Denis A.; Lee-Waddell, Karen; López-Sánchez, Ángel R.; Oh, Se-Heon; Reynolds, Tristan N.; Rhee, Jonghwan; Spekkens, Kristine; Wang, Jing; Westmeier, Tobias; and Wong, O Ivy, "WALLABY Pilot Survey: hydra cluster galaxies UV and H I morphometrics" (2023). *Faculty Scholarship*. 898.

<https://ir.library.louisville.edu/faculty/898>

This Article is brought to you for free and open access by ThinkIR: The University of Louisville's Institutional Repository. It has been accepted for inclusion in Faculty Scholarship by an authorized administrator of ThinkIR: The University of Louisville's Institutional Repository. For more information, please contact [thinkir@louisville.edu](mailto:thinkir@louisville.edu).

---

## Authors

Benne W. Holwerda, Frank Bigiel, Albert Bosma, Helene M. Courtois, Nathan Deg, Helga Dénes, Ahmed Elagali, Bi-Qing For, Baerbel Koribalski, Denis A. Leahy, Karen Lee-Waddell, Ángel R. López-Sánchez, Se-Heon Oh, Tristan N. Reynolds, Jonghwan Rhee, Kristine Spekkens, Jing Wang, Tobias Westmeier, and O Ivy Wong

# WALLABY Pilot Survey: Hydra Cluster Galaxies UV and H I morphometrics

Benne W. Holwerda<sup>1\*</sup>, Frank Bigiel<sup>2</sup>, Albert Bosma<sup>3</sup>, Helene M. Courtois<sup>4</sup>,  
Nathan Deg<sup>5</sup>, Helga Dénes<sup>6</sup>, Ahmed Elagali<sup>7</sup>, Bi-Qing For<sup>7,8</sup>, Baerbel Koribalski<sup>9,10,8</sup>,  
Denis A. Leahy<sup>11</sup>, Karen Lee-Waddell<sup>7,12</sup>, Ángel R. López-Sánchez<sup>13,15,8</sup>,  
Se-Heon Oh<sup>16</sup>, Tristan N. Reynolds<sup>7,8</sup>, Jonghwan Rhee<sup>7,8</sup>, Kristine Spekkens<sup>17</sup>,  
Jing Wang<sup>18</sup>, Tobias Westmeier<sup>7,8</sup>, and O. Ivy Wong<sup>7,8,12</sup>

<sup>1</sup> University of Louisville, Department of Physics and Astronomy, 102 Natural Science Building, 40292 KY Louisville, USA.

<sup>2</sup> Argelander-Institut für Astronomie, Universität Bonn, Auf dem Hügel 71, 53121 Bonn, Germany

<sup>3</sup> Aix Marseille Univ, CNRS, CNES, LAM, Marseille, France

<sup>4</sup> Univ Lyon, Univ Claude Bernard Lyon 1, IUF, IP2I Lyon, F-69622, Villeurbanne, France

<sup>5</sup> Department of Physics, Engineering Physics, and Astronomy, Queen's University, Kingston, ON, K7L 3N6, Canada

<sup>6</sup> ASTRON, Netherlands Institute for Radio Astronomy, 7991 PD Dwingeloo, The Netherlands

<sup>7</sup> ICRAR, The University of Western Australia, 35 Stirling Highway, Crawley WA 6009, Australia

<sup>8</sup> ARC Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D), Australia

<sup>9</sup> Australia Telescope National Facility, CSIRO Space & Astronomy, PO Box 76, Epping, NSW 1710, Australia

<sup>10</sup> School of Science, Western Sydney University, Locked Bag 1797, Penrith, NSW 2751, Australia

<sup>11</sup> Department of Physics and Astronomy, University of Calgary, 2500 University Dr. NW, Calgary, T2N 1N4, Canada

<sup>12</sup> Australia Telescope National Facility, CSIRO Space & Astronomy, PO Box 1130, Bentley, WA 6102, Australia

<sup>13</sup> Australian Astronomical Optics, Macquarie University, 105 Delhi Rd, North Ryde, NSW 2113, Australia

<sup>14</sup> Department of Physics and Astronomy, Macquarie University, NSW 2109, Australia

<sup>15</sup> Macquarie University Research Centre for Astronomy, Astrophysics & Astrophotonics, Sydney, NSW 2109, Australia

<sup>16</sup> Department of Physics and Astronomy, Sejong University, 209 Neungdong-ro, Gwangjin-gu, Seoul, Republic of Korea

<sup>17</sup> Department of Physics and Space Science, Royal Military College of Canada, P.O. Box 17000, Station Forces, Kingston, Ontario, Canada K7K 7B4

<sup>18</sup> Kavli Institute for Astronomy and Astrophysics, Peking University, Beijing 100871, China

Accepted XXX. Received YYY; in original form June 2022

## ABSTRACT

Galaxy morphology in atomic hydrogen (H I) and in the ultra-violet (UV) are closely linked. This has motivated their combined use to quantify morphology over the full H I disk for both H I and UV imaging. We apply galaxy morphometrics: Concentration, Asymmetry, Gini,  $M_{20}$  and Multimode-Intensity-Deviation statistics to the first moment-0 maps of the WALLABY survey of galaxies in the Hydra cluster center. Taking advantage of this new H I survey, we apply the same morphometrics over the full H I extent on archival GALEX *FUV* and *NUV* data to explore how well H I truncated, extended ultraviolet disk (XUV) and other morphological phenomena can be captured using pipeline WALLABY data products. Extended H I and UV disks can be identified relatively straightforward from their respective concentration. Combined with WALLABY H I, even the shallowest GALEX data is sufficient to identify XUV disks. Our second goal is to isolate galaxies undergoing ram-pressure stripping in the H I morphometric space. We employ four different machine learning techniques, a decision tree, a k-nearest neighbour, a support-vector machine, and a random forest. Up to 80% precision and recall are possible with the Random Forest giving the most robust results.

**Key words:** galaxies: disc – galaxies: ISM – galaxies: kinematics and dynamics – galaxies: spiral – galaxies: statistics – galaxies: structure

\* Contact e-mail: benne.holwerda@louisville.edu

## 1 INTRODUCTION

In principle, the appearance of galaxies is a direct result of the processes that formed and shaped them. Turning their appearance into a qualified value, while accounting for viewing angle is a long-standing undertaking in observational extragalactic astronomy.

While most morphology quantifications focus on the stellar component of galaxies, a very information-rich view of gas-rich galaxies is the 21cm line emission of neutral hydrogen (H I). This gas spans most often a larger disk than the stars and it is considered more sensitive to early interaction (e.g., [Hibbard et al. 2001](#)). There is an active interest in the outskirts of spiral galaxy disks because they are the sites of the most recent, readily observable acquisition of gas for these systems (e.g. [Sancisi et al. 2008](#)), as well as low-level star-formation (e.g., [Dong et al. 2008](#); [Koribalski & López-Sánchez 2009](#); [Bigiel et al. 2010](#); [Alberts et al. 2011](#); [López-Sánchez et al. 2015](#); [Watson et al. 2016](#)).

This low-level star-formation was first discovered in H $\alpha$  emission by [Ferguson et al. \(1998\)](#) and [Lelièvre & Roy \(2000\)](#) and later in GALEX ultraviolet imaging as Extended UltraViolet (XUV) disks ([Thilker et al. 2005a,b, 2007](#); [Gil de Paz et al. 2005, 2007](#); [Zaritsky & Christlein 2007](#); [Lemonias et al. 2011](#); [Meurer 2017](#); [Koribalski 2017](#)). These XUV disk complexes are generally  $\sim 100$  Myr old, which explains why most lack H $\alpha$  ([Alberts et al. 2011](#)). UV knots in interactions may be an exception ([López-Sánchez et al. 2015](#)). A different explanation, i.e. a top-light Initial Mass Function (IMF), as proposed by [Meurer et al. \(2009\)](#); [Koda et al. \(2015\)](#); [Bruzzeze et al. \(2015\)](#); [Watts et al. \(2018\)](#), remains a possibility but the IMF is stochastically sampled. The metallicities of these complexes, as derived from emission lines, are sub-solar, in the range of  $0.1 - 1 Z_{\odot}$  ([Gil de Paz et al. \(2007\)](#); [Bresolin et al. \(2009\)](#); [Werk et al. \(2010\)](#); [López-Sánchez et al. \(2015\)](#)).

It was noted early by several authors that the atomic hydrogen as observed by the 21cm fine structure line and the ultraviolet structure seem closely related. That the H I disk extends well beyond the optical disk of spiral galaxies has been known for a long time (e.g., [Bosma 1978](#); [Begeman 1989](#); [Meurer et al. 1996, 1998](#); [Swaters et al. 2002](#); [Noordermeer et al. 2005](#); [Walter et al. 2008](#); [Boomsma et al. 2008](#); [Elson et al. 2011](#); [Heald et al. 2011b,a](#); [Zschaechner et al. 2011](#); [de Blok et al. 2008](#)) but only recently the close correlation with UV structure drove a direct comparison (cf [de Paz et al. 2008](#); [Lemonias et al. 2011](#)).

A lopsided appearance of the outer H I disk ([Jog & Combes 2009](#); [van Eymeren et al. 2011a,b](#)) or an asymmetry ([Giese et al. 2016](#); [Reynolds et al. 2020](#)) can have a myriad of internal or external processes responsible for it: accretion along cosmic web filaments ([Bournaud et al. 2005](#)), minor satellite accretion ([Zaritsky & Rix 1997](#)), tidal interactions ([Jog & Combes 2009](#); [Koribalski & López-Sánchez 2009](#)), ram-pressure stripping ([Moore & Gottesman 1998](#)) as well as sloshing within the dark matter halo ([Stinson et al. 2009](#)). The anecdotal similarity and the potential common origin of gas accretion is what drove the use of morphometrics on combined data-sets ([Holwerda et al. 2011b, 2012](#)).

Parameterisation of H I disk appearance is different from stellar parametrization because the H I disk is based on line emission and therefore has a lower dynamic range. The area covered by the disk is larger, but the spatial resolution is typically an order of magnitude lower due to the much larger H I beam (or PSF).

A good companion data-set is typically GALEX ultraviolet (FUV and NUV) because the spatial sampling is often of similar to H I scales, and H I and the UV from newly formed young stars are related in distribution ([Thilker et al. 2005a](#); [Gil de Paz et al. 2005](#))

and possibly linked through mutual formation processes ([Heiner et al. 2014](#); [López-Sánchez et al. 2018](#)). The main benefit of using both H I and UV imaging information is that UV images are inherently very clumpy in appearance. It becomes increasingly difficult to distinguish which UV source is a star-forming region belonging to the disk and which is a background source. By using the H I outer contour, the extent of the UV disk can be identified more reliably (e.g. the discussion on outermost H II regions in [Hunter et al. 2018](#)).

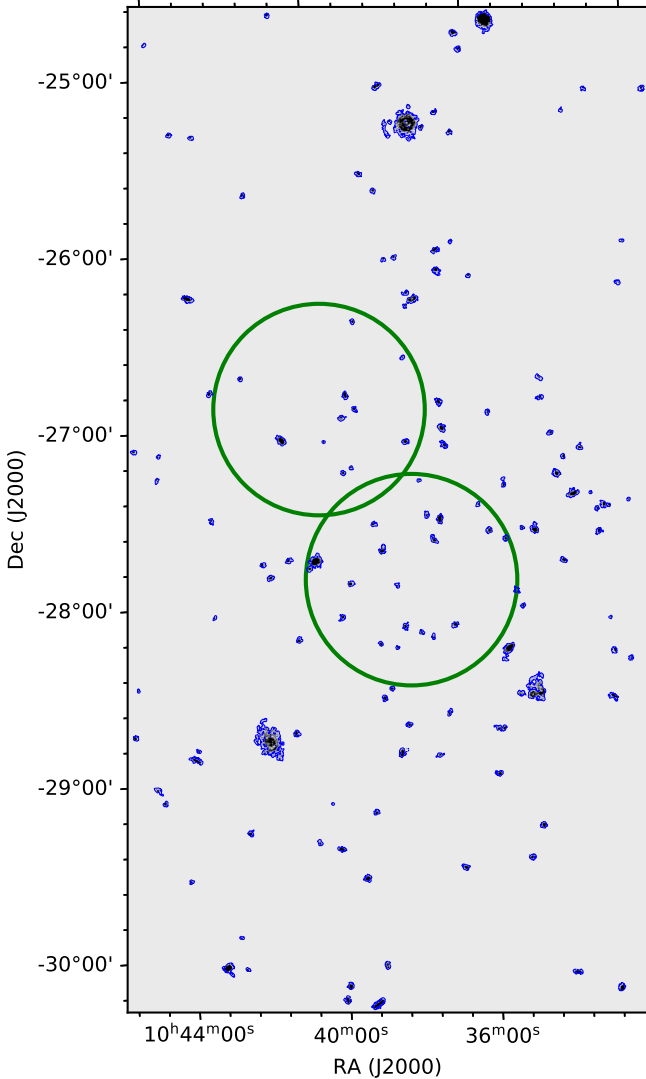
In this paper we apply the galaxy morphometrics originally developed for stellar disks which were applied with some success on H I data in the past ([Holwerda et al. 2011a,b,c,d,e, 2012](#); [Giese et al. 2016](#); [Reynolds et al. 2020](#)) but on often heterogeneous data. For example [Giese et al. \(2016\)](#) pointed out that these depend strongly on the signal-to-noise ratio (s/n) of each object, complicating their use across surveys or with varying s/n. The optimal application is therefore within a single survey although careful comparisons can successfully be made between instrument setups of different surveys ([Reynolds et al. 2020](#)). We use the STATMORPH implementation of these morphometrics ([Rodríguez-Gomez et al. 2019](#)).

Starting with the early WALLABY data ([Serra et al. 2015b](#); [Koribalski et al. 2020](#); [For et al. 2019](#); [Kleiner et al. 2019](#); [Lee-Waddell et al. 2019](#); [Elagali et al. 2019](#); [Reynolds et al. 2019](#)), we apply the image morphometrics on H I and GALEX FUV data to explore their utility in identifying disturbed H I disks, H I deficient or truncated disks, and extended ultraviolet disks. This paper is organized as follows: Section 2 describes the early WALLABY ASKAP and GALEX data products used, Section 3 describes the morphometrics employed on both H I and GALEX imaging, Section 4 describes the results for morphometrics to identify extended and ram-pressure stripped disks, Section 5 describes the H I morphometrics of ram-pressure galaxies, Section 6 describes the machine learning applied to the H I morphometric space to identify RPS galaxies, and Section 7 briefly discusses all of our results and a future outline of their potential use in WALLABY survey proper. Throughout we use the Planck (2015) cosmology ( $H_0 = 67.74$  km/s/Mpc,  $\Omega_0 = 0.3075$ , [Planck Collaboration et al. 2016](#)).

## 2 WALLABY SURVEY

The Widefield ASKAP L-band Legacy All-sky Blind survey (or WALLABY, [Koribalski et al. 2020](#)) is one of several key surveys that will soon begin on the Australian SKA Pathfinder (ASKAP [Johnston et al. 2008, 2007](#); [Johnston 2007](#); [Johnston et al. 2009](#); [McConnell et al. 2016](#); [Hotan et al. 2014, 2021](#); [Koribalski 2022](#)), which is an imaging radio telescope located at the Murchison Radio-astronomy Observatory in Western Australia. The aim of WALLABY is to use the powerful wide-field phased-array technology of ASKAP to observe a large fraction of the whole sky in the 21-cm line of H I at 30-arcsec angular resolution, which will yield information for hundreds of thousands of external galaxies in the local Universe.

Science goals include an understanding of the role of stellar and black hole feedback, gas accretion and galaxy interactions in galaxy evolution. Currently the WALLABY Pilot Survey observations covered three 60 deg<sup>2</sup> fields, of which the Hydra field is one. The Pilot Survey is allowing the team to develop, deploy and commission its pipeline processing, data verification and post-processing algorithms, as well as measure the properties of hundreds of galaxies, including their distance, neutral hydrogen mass, and total mass.



**Figure 1.** The moment-0 map of the central part of the Hydra cluster with H I detections outlined in blue (set to 10 Jy/beam $\times$ Hz in the moment-0). The location of the two GALEX fields are outlined in green. The full TR2 coverage extends to either side.

## 2.1 Hydra Cluster Field

The Hydra field was observed with ASKAP as part of the WALLABY Pilot Survey. The first and second internal data releases (Hydra TR1 and TR2) are available in the sense that the original data cubes produced by the ASKAP pipeline are available to team members on the CSIRO data archive CASDA. In addition, WALLABY has run its own SoFiA software (Serra et al. 2015a; Westmeier et al. 2021) to produce source catalogues and source images/spectra for the team to work on. The WALLABY team internal data release covers a sub-region of the core of the Hydra cluster in a region of the cube that is relatively clean from artefacts (Reynolds et al. 2021; Wang et al. 2021). The region covers the RA range of approximately 10:32:00 to 10:45:00 and the declination range of approximately -30:20:00 to -24:40:00, corresponding to about  $5.5 \times 3.5$  degrees on the sky or roughly one third of the eventual full Hydra field (Figure 1). The redshift range considered here covers

$cz = 500$  to 15,000 km/s ( $z < 0.05$ ), roughly half of the RFI-free band available to WALLABY.

In order to reduce the impact of residual continuum emission on the source finding, the positions of all NRAO VLA Sky Survey (NVSS Condon et al. 1998) sources with a flux density in excess of 150 mJy were manually flagged in the form of a box of size  $11 \times 11$  pixels centred on the pixel containing each NVSS position.

The great benefit for calculating morphometrics in this field is that the galaxies are all at comparable distances and the observations all probe similar spatial scales.

## 2.2 SoFiA Source Finding

H I detections done with SoFiA (Serra et al. 2015a; Westmeier et al. 2021) were linked across a spatial and spectral radius of 2 with a minimum size requirement for a reliable source of 8 spatial pixels and 5 spectral channels. SoFiA’s reliability filter was then applied to remove all detections with a reliability below 0.6, using a Gaussian kernel density estimator of size 0.3 times the covariance. All remaining sources were then parameterised, assuming a restoring beam size of 30 arcsec for all integrated flux measurements.

The resulting detections were inspected by eye to remove obvious artefacts that were unlikely to be genuine H I sources. Many of these artefacts are related to inadequate flagging of the upper-left and lower-left corner beams of the mosaic that have strayed into the region considered here. After manual removal of artefacts, 272 detections remain. This data set was released internally to the WALLABY team as Hydra TR2.

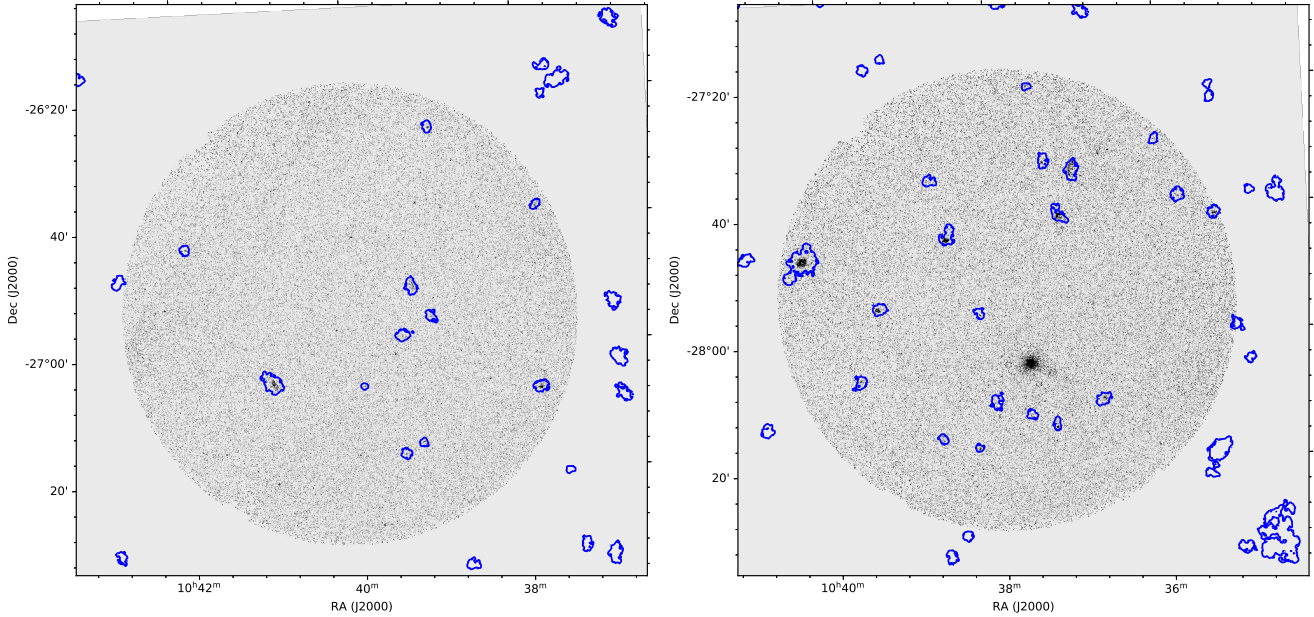
We use the internal moment-0 maps (total line intensity across frequencies associated with the galaxy) produced (Figure 1), the 3D mask cube, and the SoFiA source catalog. These formed the basis for a segmentation image, an image denoting which pixels belong to an object in the catalog. We replaced non-zero pixels in the mask cube, collapsed on the frequency axis, with the source identification in the SoFiA catalog to be used as the segmentation image input in STATMORPH (Rodríguez-Gomez et al. 2019, see §3). To compute the morphometrics, we use the moment-0 and the collapsed mask cutouts from SoFiA, not the full mosaic out of memory management considerations.

## 2.3 GALEX observations of Hydra

Two fields were observed for the original All Sky Imaging Surveys (AIS) with the GALEX telescope in both the FUV and NUV channels (1350-1750 and 1750-2800 Å, respectively) with 210s exposures (Figure 2). Overlap with the WALLABY central FOV is good but not all the H I detections are covered by UV observations (30 out of 272). While 63 galaxies have non-zero morphometrics, only 30 galaxies overlap with GALEX data, the remainder is the result of morphometrics on the empty sky contribution. Statmorph runs on a segmentation map, in this case the H I one and will measure morphometrics where no source flux is present. Some SWIFT observations exist but only cover a single galaxy in the H I field and thus were not used.

## 3 MORPHOMETRIC CATALOGS

An observational approach to characterizing galaxy appearances is “non-parametric” or morphometric parameters. In essence, several authors and groups have tried to come up with parameters that can be applied to images that do not depend too much on resolution



**Figure 2.** The FUV filter GALEX observations of the Hydra cluster. H I detections are marked in blue (contour at 10 Jy/beam×Hz in the moment-0 map). There is an overlap of 30 H I detections in the GALEX fields with associated GALEX morphometrics. The bright FUV source in the second field is a foreground galactic star.

or a preconceived idea about the shape of the profile. These non-parametric or “morphometric” parameters<sup>1</sup> can then be used as quantities to classify galaxies along the Hubble Tuning fork or to identify mergers in a population of galaxies.

The morphometric parameters considered here are Concentration, Asymmetry and Smoothness from [Conselice \(2003\)](#),  $M_{20}$  and Gini from [Lotz et al. \(2004\)](#), and the MID parameters from [Peth et al. \(2016\)](#); [Rodríguez-Gomez et al. \(2019\)](#). We use the `STATMORPH` package described in [Rodríguez-Gomez et al. \(2019\)](#) to compute the morphometrics. We utilize a Gaussian smoothing kernel with a 1 pixel FWHM (6”) for both H I and UV implementations of `STATMORPH`. This choice is not critical for the work here as this input smoothing kernel is only used for the Sérsic profile fit in `STATMORPH`, the index of which we do not use here. Both H I and UV profiles are typically not well described with a Sérsic profile (cf [Leroy et al. 2008](#); [Bigiel et al. 2011](#); [Wang et al. 2014](#); [Swaters et al. 2002](#), and Reynolds et al. in prep. for result of H I radial profiles).

In many cases, these morphometrics still require inputs other than the image. The relevant input parameters are the central position of the galaxy ( $x_c, y_c$ ), and a definition of the area over which these parameters are computed, the segmentation of the image (i.e. which pixels belong to the object of interest). These two measurements, the center of the object and which parts of the image belong to it, are often obtained from source extractor ([Bertin & Arnouts 1996](#); [Holwerda 2005](#)).

Here, the segmentation map was generated from the the SoFiA 3D mask and the central positions were computed by `STATMORPH`. By using the H I map for the segmentation, the granular nature of the FUV/NUV imaging can be accounted for and the oversegmentation of the UV image can be avoided. The use of the H I contour

effectively isolates distant galaxies from Galactic stellar sources. Foreground Galactic stars are less of a consideration (cf Figure 2).

### 3.1 Concentration-Asymmetry-Smoothness (CAS) Morphometrics

CAS refers to the now commonly used Concentration-Asymmetry-Smoothness space ([Conselice 2003](#)) for morphological analysis of distant galaxies. Concentration of the light, symmetry around the centre and smoothness as an indication of substructure.

Concentration is defined by [Bershady et al. \(2000\)](#) as:

$$C = 5 \log(r_{80}/r_{20}) \quad (1)$$

with  $r_f$  as the radius containing percentage  $f$  of the light of the galaxy (see definitions of  $r_f$  in [Bertin & Arnouts \(1996\)](#); [Holwerda \(2005\)](#)). In the optical, typical values for the concentration index are  $C = 2 - 3$  for discs,  $C > 3.5$  for massive ellipticals, while peculiars span the entire range ([Conselice 2003](#)).

The asymmetry is defined as the level of *point-*, (or rotational-) symmetry around the centre of the galaxy ([Abraham et al. 1994](#); [Conselice 2003](#)):

$$A = \frac{\sum_{i,j} |I(i,j) - I_{180}(i,j)|}{\sum_{i,j} |I(i,j)|} - A_{bgr}, \quad (2)$$

where  $I(i,j)$  is the value of the pixel at the position  $[i,j]$  in the image, and  $I_{180}(i,j)$  is the pixel at position  $[i,j]$  in the galaxy’s image, after it was rotated 180° around the centre of the galaxy. In the `STATMORPH` implementation, the asymmetry is calculated in the inner 1.5 Petrosian<sup>2</sup> radii (typical size of the stellar disk), the background asymmetry is subtracted and A is minimized by moving the

<sup>1</sup> A term first coined here [Davenport \(2015\)](#). The term “morphometrics” originates from biology to describe animal appearance in quantitative form, e.g. “length of fins” or how far the eyes are set apart.

<sup>2</sup> The Petrosian radius is one of several definitions to automatically assign a size and aperture to inherently fuzzy galaxies. For a comprehensive treatment on them, see [Graham et al. \(2005\)](#); [Graham & Driver \(2005\)](#).

center of rotation. This is a different implementation than used in [Holwerda et al. \(2012\)](#). Note that asymmetry maximum value is 2 (all pixels off-center) and can be negative if the background asymmetry value is large. We note that we do not subtract a background when using the moment-0 H I maps as these are extracted from the field using the 3D source mask.

The fact that this Smoothness has another input parameter in the form of the size of the smoothing kernel, makes it highly “tunable”, meaning one gets out exactly what was optimized for. It is very difficult to reliably compare between catalogs or even implementations. For this reason, “Smoothness” is not considered further here.

### 3.2 Gini and $M_{20}$

[Abraham et al. \(2003\)](#) and [Lotz et al. \(2004\)](#) introduce the Gini parameter to quantify the distribution of flux over the pixels in an image. They use the following definition:

$$G = \frac{1}{\bar{I}n(n-1)} \sum_i (2i - n - 1) I_i, \quad (3)$$

$I_i$  is the value of pixel  $i$  in an ordered list of the pixels,  $n$  is the number of pixels in the image, and  $\bar{I}$  is the mean pixel value in the image. This is the computationally least expensive implementation where not the entire mosaic has to be loaded.

The Gini parameter is an indication of equality in a distribution (initially an economic indicator [Gini 1912](#); [Yitzhaki 1991](#)), with  $G=0$  the perfect equality (all pixels have the same fraction of the flux) and  $G=1$  perfect inequality (all the intensity is in a single pixel). Its behaviour is therefore in between that of a structural measure and concentration. Gini appears quite sturdy as it does not require the center of the object to be computed. It remains relatively unchanged, even when the object is lensed ([Florian et al. 2016](#)) and it is popular for this reason. However, it depends strongly on the image’s signal-to-noise ([Lisker 2008](#)). In essence, noise can add pixels with no fraction of the flux in them, artificially increasing the Gini value.

[Lotz et al. \(2004\)](#) also introduced a way to parameterize the extent of the light in a galaxy image. They define the spatial second order moment as the product of the intensity with the square of the projected distance to the centre of the galaxy. This gives more weight to emission further out in the disk. It is sensitive to substructures such as spiral arms and star-forming regions but insensitive to whether these are distributed symmetrically or not. The second order moment of a pixel  $i$  is defined as:

$$M_i = I_i \times [(x - x_c)^2 + (y - y_c)^2], \quad (4)$$

where  $[x, y]$  is the position of a pixel with intensity value  $I_i$  in the image and  $[x_c, y_c]$  is the central pixel position of the galaxy in the image.

The total second order moment of the image is given by:

$$M_{tot} = \sum_i M_i = \sum_i I_i [(x_i - x_c)^2 + (y_i - y_c)^2]. \quad (5)$$

[Lotz et al. \(2004\)](#) use the relative contribution of the brightest 20% of the pixels to the second order moment as a measure of disturbance of a galaxy after sorting the list of pixels by intensity ( $I_i$ ):

$$M_{20} = \log \left( \frac{\sum_i M_i}{M_{tot}} \right), \text{ for } \sum_i I_i < 0.2 I_{tot}. \quad (6)$$

The  $M_{20}$  parameter is sensitive to bright regions in the outskirts of disks and higher values can be expected in galaxy images (in the optical and UV) with star-forming outer regions as well as those images of strongly interacting disks.

### 3.3 multimode–intensity– deviation (MID) morphometrics

The MID morphometrics ([Freeman et al. 2013](#); [Peth et al. 2016](#)) were introduced as an alternative to the Gini– $M_{20}$  and CAS morphometrics to be more sensitive to recent mergers. However, these new morphometrics have not been tested as extensively as the Gini– $M_{20}$  and CAS statistics, especially using hydrodynamic simulations ([Lotz et al. 2008, 2010, 2011](#); [Bignone et al. 2017](#)), see also the discussion in the implementation in STATMORPH ([Rodríguez-Gomez et al. 2019](#)). In the case of H I data, none of the morphometrics have been extensively tested, and are to be viewed as purely phenomenological.

We explore them here because the UV morphology tends to be much more “clumpy” than the optical one and, by using the H I outer contour, one can still consider such galaxies as a single object. The MID morphometrics were developed to identify mergers by distinct cores, not segregated by the segmentation, and may prove useful.

## 4 RESULTS

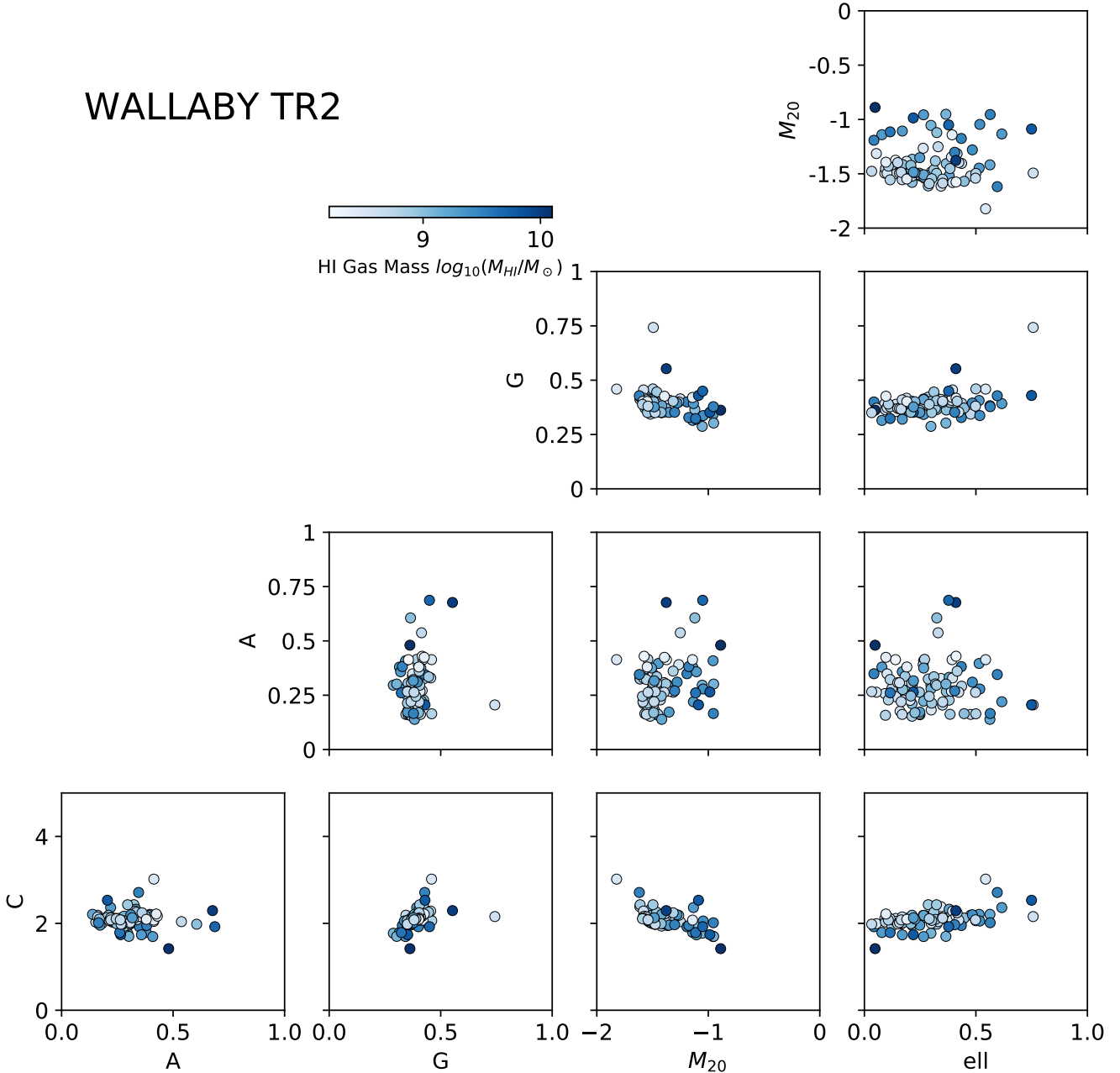
In this section we discuss the two morphometric catalogs computed using the WALLABY H I data and the WALLABY and GALEX data combined.

### 4.1 H I Morphometrics

Figure 3 shows the distribution of the most commonly used morphometrics for the H I moment-0 map. There is a good range of values, despite the much lower dynamical range typically present in H I maps. The concentration values for H I are clustered around a value of 2. There is a full range of asymmetry values. Asymmetry is calculated with the inclusion of a background component by STATMORPH, but none is included here since the values in the moment-0 map not belonging to an object are set to zero. There are also reasonably high values for the Gini and  $M_{20}$  parameters indicating a fair amount of substructure resolved by the WALLABY observations. Concentration and  $M_{20}$  are anti-correlated to some degree as can be expected because  $M_{20}$  up-weighs pixels far from the center. The range of ellipticity values indicates a wide distribution of viewing angles on the H I disks of these galaxies. These ellipticity values are based on second order moments and not ellipse fitting of the H I cube. The H I morphometrics show a wide range in value. The only parameters that show a narrow range are concentration and Gini, which both can be attributed to the lower dynamic range of the moment-0 maps.

### 4.2 MID H I Morphometrics

The multimode-intensity-deviation (MID) morphometrics for the H I moment-0 maps are compared to the more traditional C-A-G- $M_{20}$  in Figure 4. Morphometrics were never meant to be an orthogonal parameter space but these MID morphometrics may probe slightly different phenomena. Deviation and  $M_{20}$  are clearly correlated and Multimode and Intensity may be weakly correlated with  $M_{20}$  as well. Both Multimode and Intensity values are predominantly close to 0 but there are notable outliers. These MID morphometrics, which are sensitive to substructure within an object, may prove useful for H I and UV science in addition to the other, original morphometrics developed for optical morphology.



**Figure 3.** The H I morphometrics corner plot. Concentration, Asymmetry, Gini and M20 together with the ellipticity computed by STATMORPH with the H I mass from WALLABY.

### 4.3 Truncated and extended H I disks

[Holwerda et al. \(2011e\)](#) noted that H I computed concentration may be linked to H I stripping and truncation. Figure 5 shows that relatively few galaxies in the overlap sample exhibit the low values of H I concentration ( $C < 1.5$ ) associated with truncated H I disks. In the optical, the typical exponential disc value is  $C = 2 - 3$ , a spheroidal galaxy much higher.

We view these low concentration H I disks as possibly truncated, similar to those observed by [Chung et al. \(2009\)](#) for the Virgo cluster at slightly higher spatial resolution (15") with the VLA. [Hol-](#)

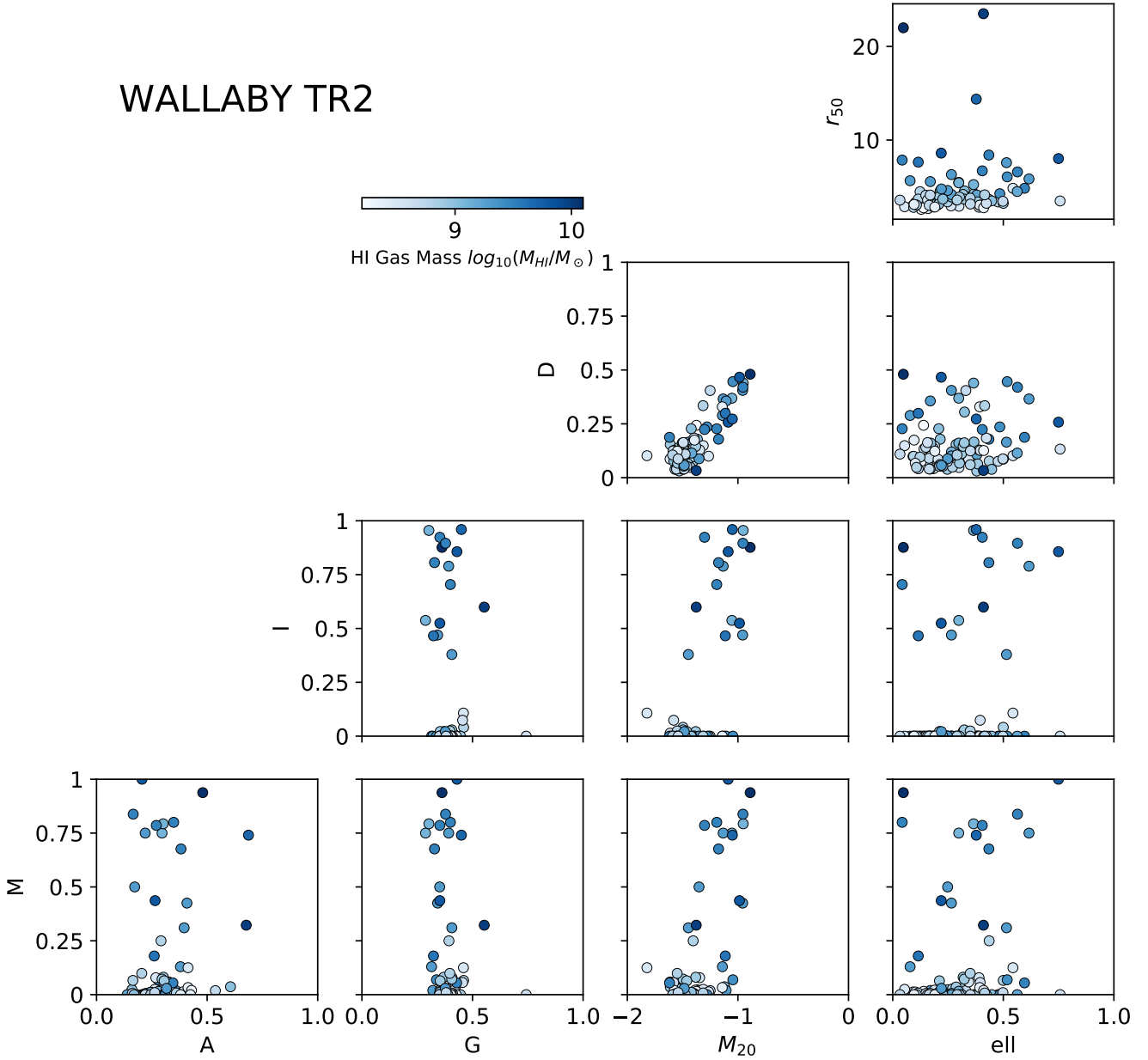
[werda et al. \(2011e\)](#) found in the VIVA data that low-concentration H I disks are more likely truncated.

Conversely, high values of concentration could be an indication of an extended H I or FUV disk. Figure 5 shows high values for mutually exclusive subsamples (listed in tables 2 and 1).

### 4.4 FUV Morphometrics

Figure 6 shows the corner plot for both FUV and NUV galaxies. Some of the objects included have unphysical values (e.g.  $A=0$ ) and are likely included despite no flux in the H I contour. In general, the NUV/FUV catalog is much smaller than the H I catalog (30 objects





**Figure 4.** The multimode-intensity-deviation (MID) morphometrics compared against the Concentration-Asymmetry-Gini- $M_{20}$  of the H I morphometrics computed by `STATMORPH` with the H I mass from WALLABY.

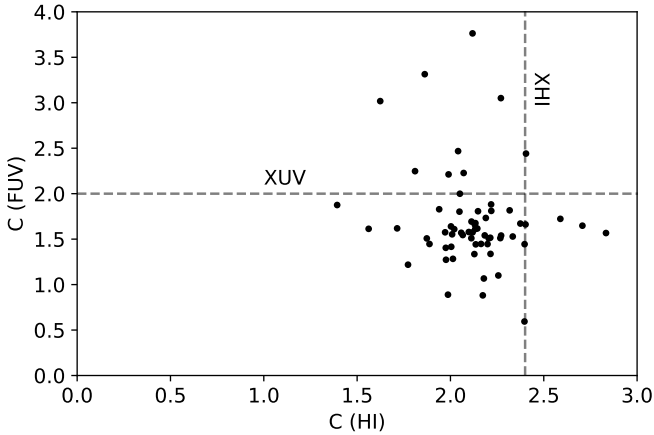
in the GALEX fields with 63 with nonzero GALEX morphometric values and 272 H I morphometric values), calculated for all the objects in GALEX footprint.

#### 4.5 XUV disk candidates

[Holwerda et al. \(2012\)](#) noted that the H I outer contour is a good way to compute the UV morphology as it clearly delineates which UV clumps of emission belong to a disk galaxy and which do not. Figure 7 shows the FUV  $M_{20}$  and Asymmetry values with concentration colour-coded. If we use the  $M_{20} - A$  criterion from [Holwerda et al. \(2012\)](#), very few objects in the Hydra cluster fall under the XUV classification. However, the implementation of Asymmetry differs

from [Holwerda et al. \(2012\)](#). This emphasizes our point that for each survey, wavelength and implementation, new criteria will need to be calibrated to identify objects of interest. Low values of asymmetry and  $M_{20}$  with high concentration values will be the candidates for XUV disks. [Holwerda et al. \(2012\)](#) noted that the fraction identified and specific criteria used to identify XUV disks depends on both H I and UV depth and H I resolution (e.g. WHISPS vs THINGS resolution) Based on this initial identification, 5/32 galaxies (~15%, see Table 1) in the Hydra cluster have an XUV disk, a substantial reduction (factor 5-6) compared to the field ([Lemonias et al. 2011](#); [Moffett et al. 2012](#)).

Figure 8 shows a few examples of FUV images of high FUV concentration with the WALLABY H I contour. It highlights how



**Figure 5.** The concentration values for H I and FUV for the overlap in samples. Extended H I and UV disks potentially stand out by their high concentration value. The dashed lines are the criteria for extended sources from [Holwerda et al. \(2012\)](#).

FUV cat id	$C_{FUV}$	$A_{FUV}$	$M_{20,FUV}$	$C_{HI}$	
J100720-262426	2.23	-3.27	-0.65	2.07	*
J101035-254920	3.02	-0.13	-1.31	1.62	
J101359-253824	2.44	-0.15	-0.87	2.40	*
J101434-274133	3.31	0.11	-1.40	1.86	
J101945-272719	3.05	0.26	-1.04	2.27	*
J102207-282201	2.47	-0.61	-1.04	2.04	*
J102413-284853	2.21	-0.21	-0.84	1.99	
J102600-280334	3.76	0.18	-0.73	2.12	*
J102605-280710	2.25	-1.92	-0.79	1.81	
NUV cat id	$C_{NUV}$	$A_{NUV}$	$M_{20,NUV}$	$C_{HI}$	
J100720-262426	2.21	-2.91	-0.79	2.07	
J101035-254920	2.83	0.03	-1.25	1.62	
J101359-253824	3.11	-0.04	-1.32	2.27	
J101434-274133	2.06	-0.05	-0.86	1.99	
J101945-272719	2.25	-1.28	-0.65	1.81	

**Table 1.** The galaxies with high concentration values in FUV (Figure 5) and NUV with their asymmetry and  $M_{20}$  values. These are potential XUV disks in the Hydra cluster. Asterix indicates a candidate XUV disks based on the asymmetry and  $M_{20}$  values.

extended FUV sources can be found through computed apertures, but outer disk flux contributions will be missed without the H I contour definition. Even with the most shallow GALEX observations typically available (AIP 200 s.) and greater distances than the resolved galaxies of THINGS and WHISP ([Holwerda et al. 2012](#)), one can identify extended UV disks through their morphometrics.

## 5 RAM-PRESSURE STRIPPING

[Wang et al. \(2021\)](#) present estimates of the ram-pressure stripping (RPS) that H I galaxies in the Hydra cluster are undergoing based on the same data. They use a holistic approach to classify galaxies as undergoing RPS using H I appearance, kinematics and stellar mass surface density. They flag galaxies undergoing RPS and candidate

cat id	$C_{HI}$	A	$M_{20}$
J100539-282633	2.47	0.40	-1.29
J100634-295615	2.52	0.36	-1.52
J102341-291347	2.45	0.39	-1.42
J102416-284343	2.41	0.46	-1.56
J103015-270743	2.55	0.39	-1.63
J103420-264728	3.02	0.41	-1.82
J103543-255954	2.49	0.63	-1.28
J103603-245430	2.43	0.32	-1.61
J103818-285023	2.71	0.35	-1.62
J103853-274100	2.80	0.63	-1.48
J103902-291255	2.43	0.30	-1.50
J103915-301757	2.53	0.21	-1.09
J104339-285157	2.76	0.24	-1.59
J104442-290119	2.89	0.39	-1.61
J104447-270553	2.42	0.43	-1.48

**Table 2.** The galaxies with high concentration values in H I (Figure 5) but with no FUV or NUV counterpart. These are extended H I disks that may hold XUV disks in the Hydra cluster.

galaxies for undergoing RPS (RSP-r1). Figures 9 and 10 show the H I morphometric feature space with the RPS flag and candidate RPS flag (RPS-r1) indicated.

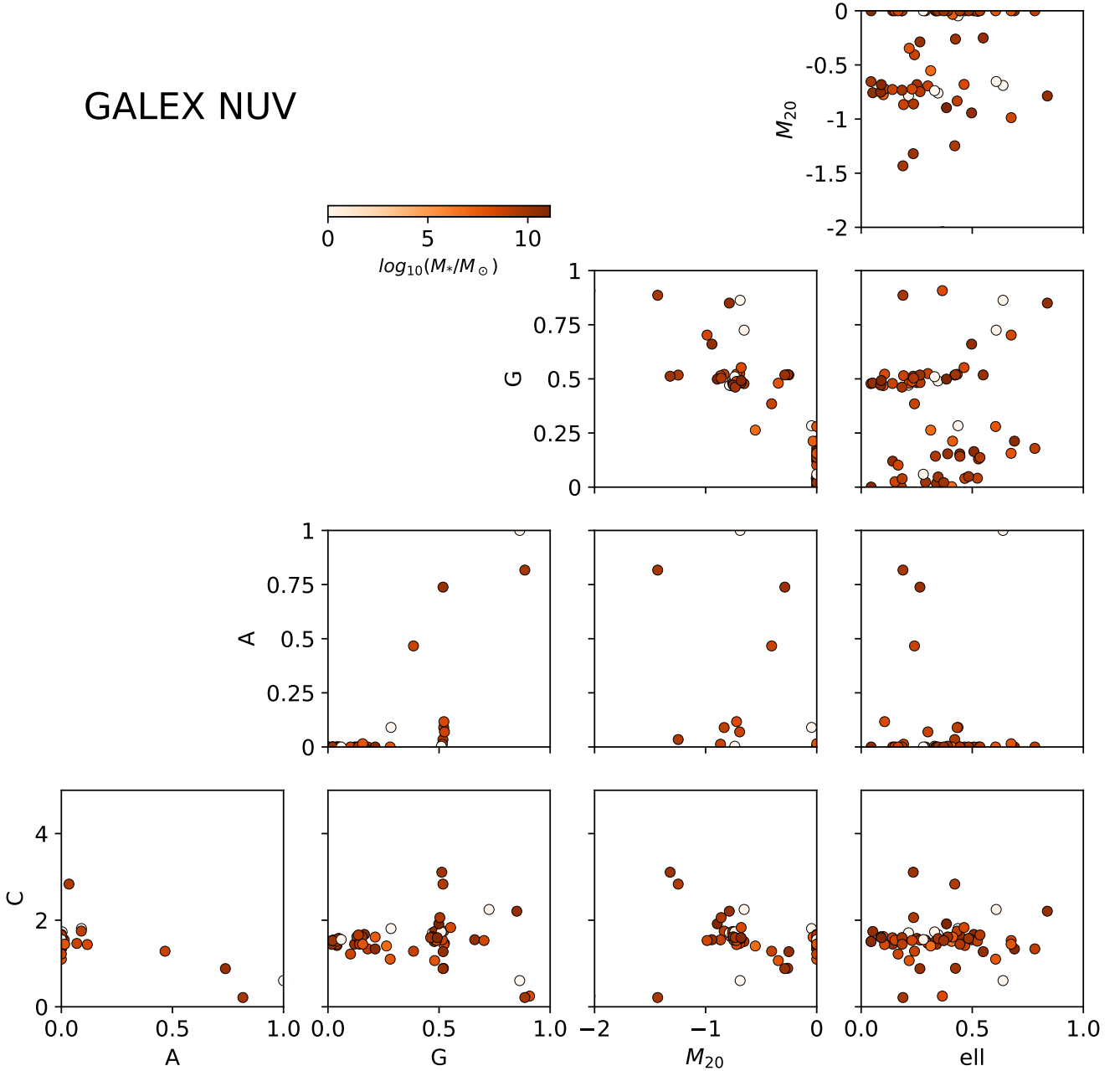
Both flags would be of interest for Machine Learning (hereafter ML) as one could use morphology to identify RPS without inspection or one could flag candidate RPS galaxies. In the H I morphometrics (Figures 9 and 10), the two populations are well-mixed, spanning the full ranges of values. It may well be that in the full feature space or a subset thereof, the two may well be separable. With no immediate clear separation in this parameter space, ML techniques developed to identify separation with a single hyperplane or a limited set of choices potentially could.

## 6 MACHINE LEARNING CLASSIFICATION OF RPS FROM H I MORPHOMETRICS

With a substantial feature space in the H I morphology and no immediate intuition on where RPS galaxies should lie within this space, we attempted to use the H I morphometric features to classify the RPS flags of [Wang et al. \(2021\)](#) using different ML classifier from SKLEARN ([Pedregosa et al. 2011](#)). The goal is to identify which H I morphometric parameters are best in pre-selecting RPS, what cuts in that space could lead to RPS candidates being identified in the remaining WALLABY survey, and how well ML can identify RPS in the H I feature space.

Because the non-RPS galaxies outnumber the RPS galaxies in the Hydra catalogs, we balance the data using SMOTE in the IMBLEARN package which supplements the SKLEARN one for machine learning. The sample is normalized (whitened) using SKLEARN's standard scaler. The sample is then split 80-20% randomly for training and testing respectively (400/100 instances after SMOTE). This is a smaller sample than typical in machine learning applications. Our aim however is to evaluate the future applicability of the H I morphometrics feature space to identify RPS.

We apply a decision tree, k-nearest neighbour, a support-vector machine, and a random forest classifier to estimate how well one can distinguish RPS galaxies in the H I morphometric space. We calculate the accuracy (fraction of correct classifications), the precision (the fraction of true positives over true negatives), the recall (the fraction of true positives of total positives) and the F1 = (precision



**Figure 6.** The distribution of NUV and FUV morphometric parameters for the objects with ultraviolet observations. There is NUV/FUV information on 30 objects (63 with catalog entries). Those with e.g.  $A=0$  are the result of morphometrics on fixed pixel values of the area outside the FOV, they encode the H I shape. The colour bar is the stellar mass of these galaxies.

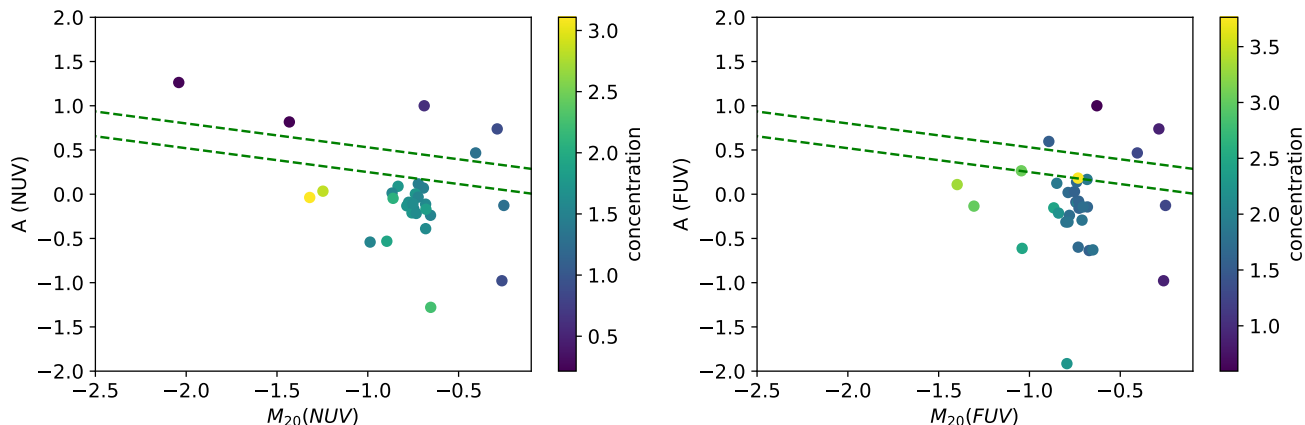
$\times \text{recall} / (\text{precision} + \text{recall})$ . Before SMOTE to rebalance the data, the F1 scores were poor for all ML algorithms that follow.

### 6.1 Decision Tree

A decision tree classifier is a series of criteria in feature space that ultimately split the training sample according to the labels provided. We split our Hydra sample into 80% training and 20% test after SMOTE rebalancing. We trained on classifying either the RPS flag or the candidate RPS flag (RPS-r1) from Wang et al. (2021).

### 6.2 H I Morphometrics Decision Tree

We tried two decision trees, one for the RPS flag and one for the candidate RPS flag (RPS-r1). Both went to over a dozen forks in depth. Every permutation of the H I morphometric feature space yielded a result (accuracy, precision, recall and F1) with the test sample. These are summarized in Table 3, showing the accuracy (combined precision for positive and negative), precision (the fraction of selected positives correctly identified), recall (the fraction of positives correctly identified), and F1 ( $= \text{precision} \times \text{recall} / (\text{precision} + \text{recall})$ ) metrics for each decision tree. Precision and recall are low for all



**Figure 7.** The asymmetry and  $M_{20}$  parameters from STATMORPH in the GALEX FUV and NUV channels with concentration colour-coded. The criterion identified by Holwerda et al. (2012) as potentially useful to identify XUV disks is marked with green dashed lines. In between these lines were the known XUV disks identified in their WHISPS/GALEX catalog. Galaxies with high values of concentration, low values of asymmetry and  $M_{20}$  are good candidates for XUV disks.

	Accuracy	Precision	Recall	F1
<i>f<sub>RPS</sub></i>				
HI C-A-G-M20-M-I-D-R50	0.80	0.71	0.89	0.39
HI C-A-G-M20	0.80	0.84	0.86	0.43
HI MID-R50	0.80	0.73	0.82	0.39
<i>f<sub>r1,RPS</sub></i>				
HI C-A-G-M20-M-I-D-R50	0.83	0.85	0.82	0.42
HI C-A-G-M20	0.78	0.80	0.80	0.40
HI MID-R50	0.80	0.80	0.82	0.41

**Table 3.** The Accuracy, Precision, Recall and F1 metrics of the decision tree based on different feature choices to identify the unresolved and resolved ram-pressure stripping flags from Wang et al. (2021).

	Accuracy	Precision	Recall	F1
<i>f<sub>RPS</sub></i>				
HI C-A-G-M20-M-I-D-R50	0.76	0.65	1.00	0.39
HI C-A-G-M20	0.81	0.72	0.93	0.41
HI M-I-D-R50	0.74	0.64	0.93	0.38
<i>f<sub>r1,RPS</sub></i>				
HI C-A-G-M20-M-I-D-R50	0.80	0.75	0.92	0.41
HI C-A-G-M20	0.69	0.67	0.84	0.37
HI M-I-D-R50	0.74	0.74	0.80	0.38

**Table 4.** The Accuracy, Precision, Recall and F1 metrics of the kNN based on different feature choices to identify the unresolved and resolved ram-pressure stripping flags from Wang et al. (2021).

trees, with Gini-R50 performing best for the RPS flag ( $f_{RPS}$ ) and the C-A-G-M20 space for the candidate RPS flag ( $f_{r1,RPS}$ ).

At present best, the H I morphometric feature space can identify about 80% of the galaxies undergoing RPS (as identified by Wang et al. 2021). Similar to what we found in Holwerda et al. (2011e), there are multiple processes influencing the H I appearance and perhaps this feature space cannot do much better. However, the sample is still small compared to typical ML applications.

### 6.3 K-Nearest Neighbour on H I Morphometrics

K-Nearest Neighbour classifies using the N nearest neighbours in the feature space. Table 4 shows the metrics for the RPS flag in different H I morphometrics with N=3. Performance is similar or poorer than the decision trees; the precision is typically worse, the recall better, resulting in comparable F1 metrics. KNN classifiers works best in coherent, homogeneous and isometric areas within the feature space. If the labels are sharply split in the feature space, this approach has less applicability.

### 6.4 Support Vector Machine on H I Morphometrics

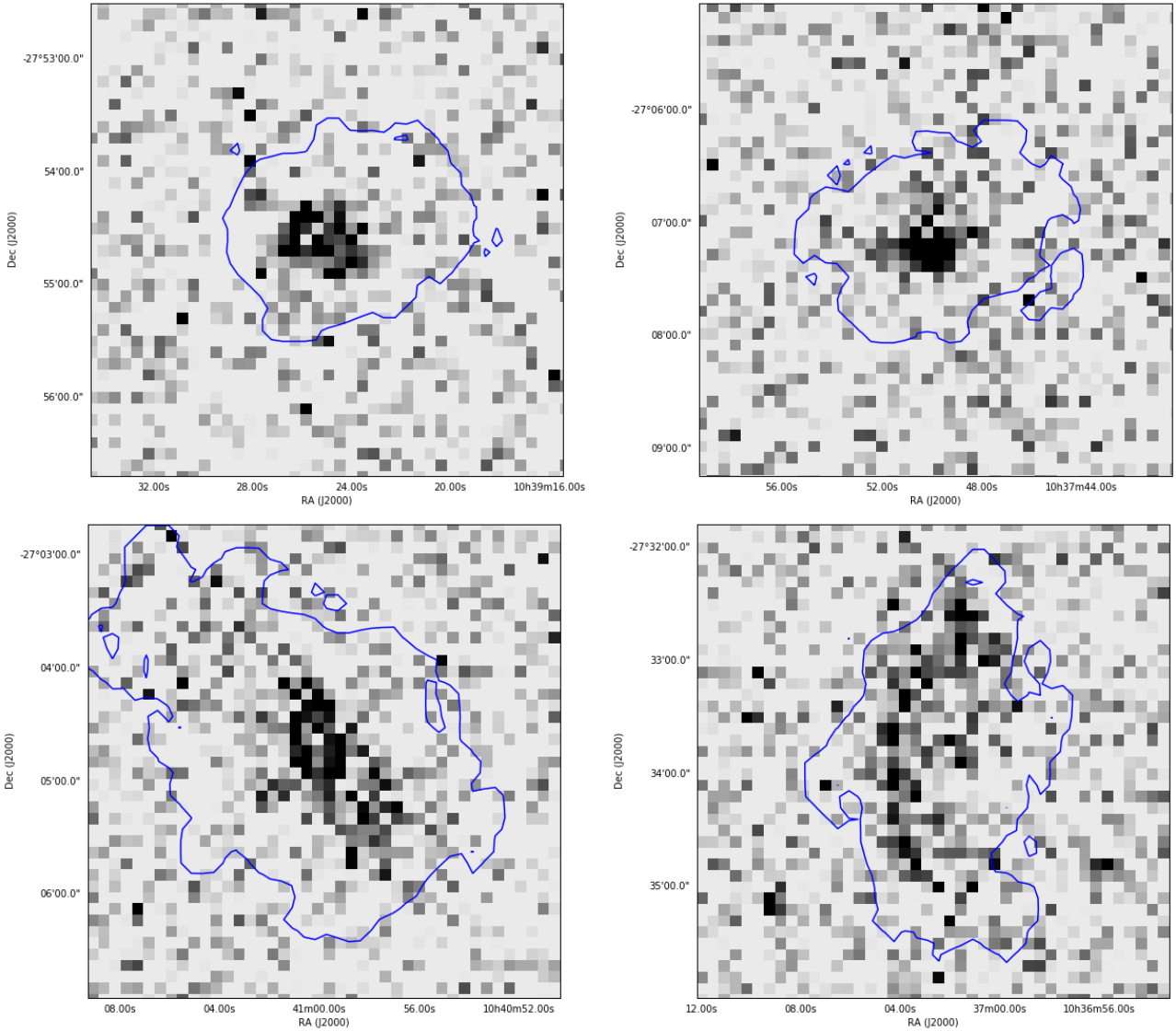
Support Vector Machine (SVM) is a machine learning algorithm optimized to disentangle labeled populations within a feature space. The SVM algorithm is to find a hyperplane in an N-dimensional space, where N is the number of features, that distinctly classifies the data points according to their labels (in this case RPS or not). We use the SKLEARN implementation of SVM to try to classify using the H I morphometrics.

The main hyper-parameters regulating SVMs are the regularization parameter  $C_{reg}$  and the choice and tuning of a convolution kernel. We opt for “sigmoid” kernel, the default, and a “polynomial” one, and plot the metrics as a function of the regularization parameter in figure 11. Most combinations of feature space perform optimally around  $C \sim 5$ . The metrics for  $C_{reg} = 5$  and the “sigmoid” kernel for each different feature set are listed in Table 5. While the accuracy initially looks reasonably promising, both precision and recall leave much to be desired (neither realistically top 75%). We note that the H I morphometrics catalog is still small for machine learning applications with 272 total entries and 148 with RPS classifications and there is room for future improvements with larger samples.

Changing to a different kernel (from sigmoid to polynomial) does not improve performance.

### 6.5 Random Forest

A random forests is an ensemble machine learning method for classification that works by constructing a multiple decision trees at training time into an ensemble (trees combined into a forest). For



**Figure 8.** Examples of GALEX FUV cutouts with the H I contour overlaid. These are reliable XUV detections with high UV values of concentration. The top examples could have been detected using an optical isophote but lower examples show the benefit of an H I contour ( $10 \text{ Jy}/\text{beam}\times\text{Hz}$ ). J103924-275442 (top-left), J103749-270715 (top-right), J104059-270456 (bottom-left), and J103702-273359 (bottom-right).

classification, the output of the random forest is the class selected by most trees. It has the benefit of more robustness and less sensitivity to outliers in the training set. For our catalog classification, this is the final option before one resorts to classifications with the images as inputs with e.g. convolutional neural networks etc.

Table 7 shows the performance of the random forest for different choices of the parameter space. It performs quite well with rebalancing the training data. Similar results on classification based on these morphometrics for simulated JWST images came to a similar conclusion that morphometrics and a random forest are well paired (Rose et. al. *in preparation*).

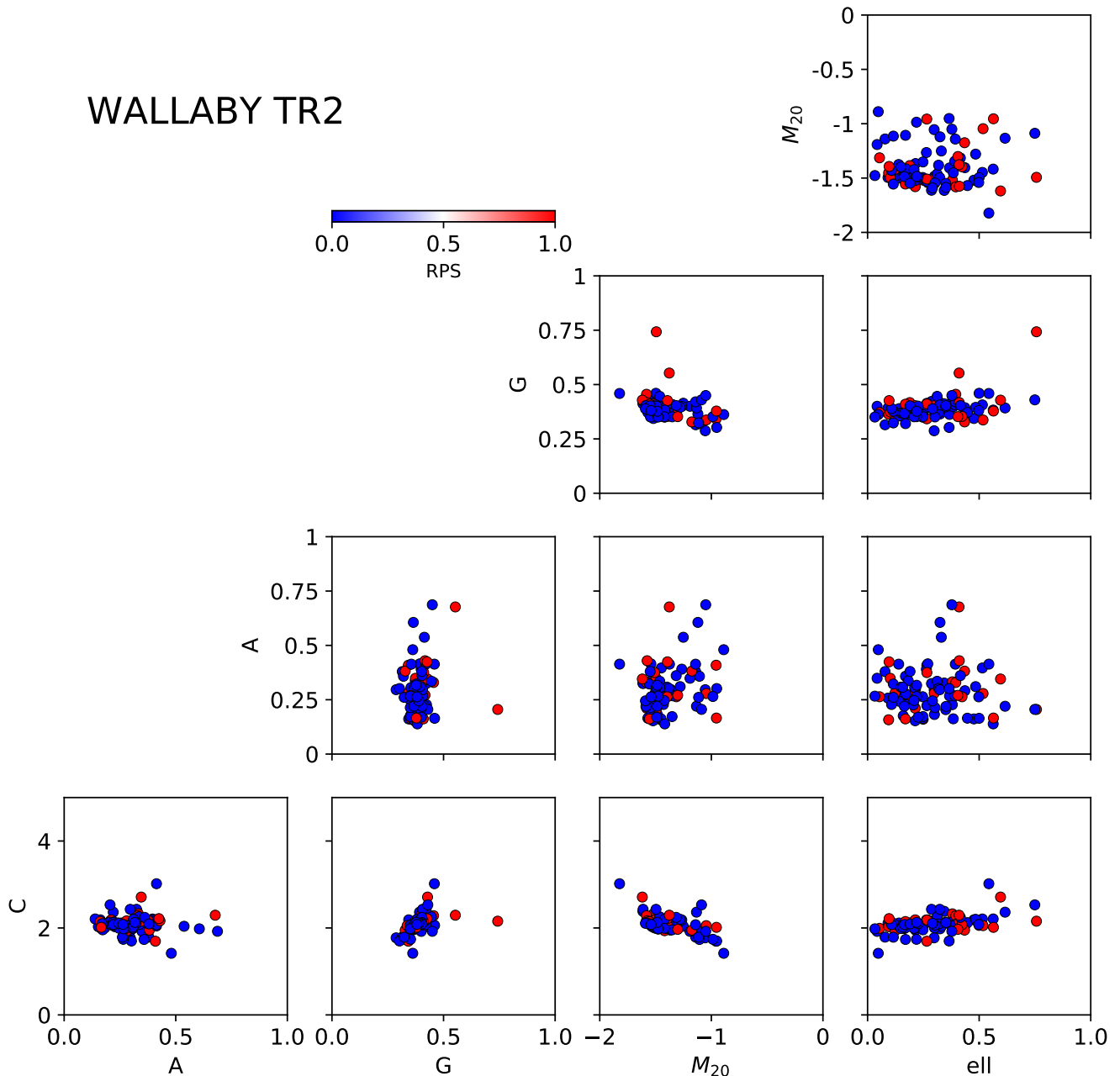
## 7 CONCLUDING REMARKS

The initial quality of the Hydra cluster WALLABY field is already enough to produce reliable morphology maps from the moment-0 (intensity) maps. We combined the H I information with GALEX

imaging for 30 out of the 272 galaxies in the field. In this way, the H I information allows one to define the complete disk of galaxies, useful to identify which UV sources should be considered part of the galaxy and which are background sources (cf the discussion of outermost H II regions in Hunter et al. 2018).

The fraction of XUV disks in the Universe is a potential constraint on how many disk galaxies are still accreting gas to their outskirts and how many are not. The outermost disk remains an excellent edge-case for star-formation in low-density ISM (Watson et al. 2016), a potential way to rejuvenate spheroidal galaxies (Moffett et al. 2012), and for tracing the chemical enrichment of disk galaxies (López-Sánchez et al. 2015).

Structural searches for XUV galaxy disks by Thilker et al. (2007) and Lemonias et al. (2011) find that some 20–30% of spirals have an XUV disk and 40% of S0s (Moffett et al. 2012), making this type of disk common but not typical for spiral and S0 galaxies. This is based mostly on local Universe data and on visual inspection (see also Thilker 2016) or a comparison with Spitzer radial profiles



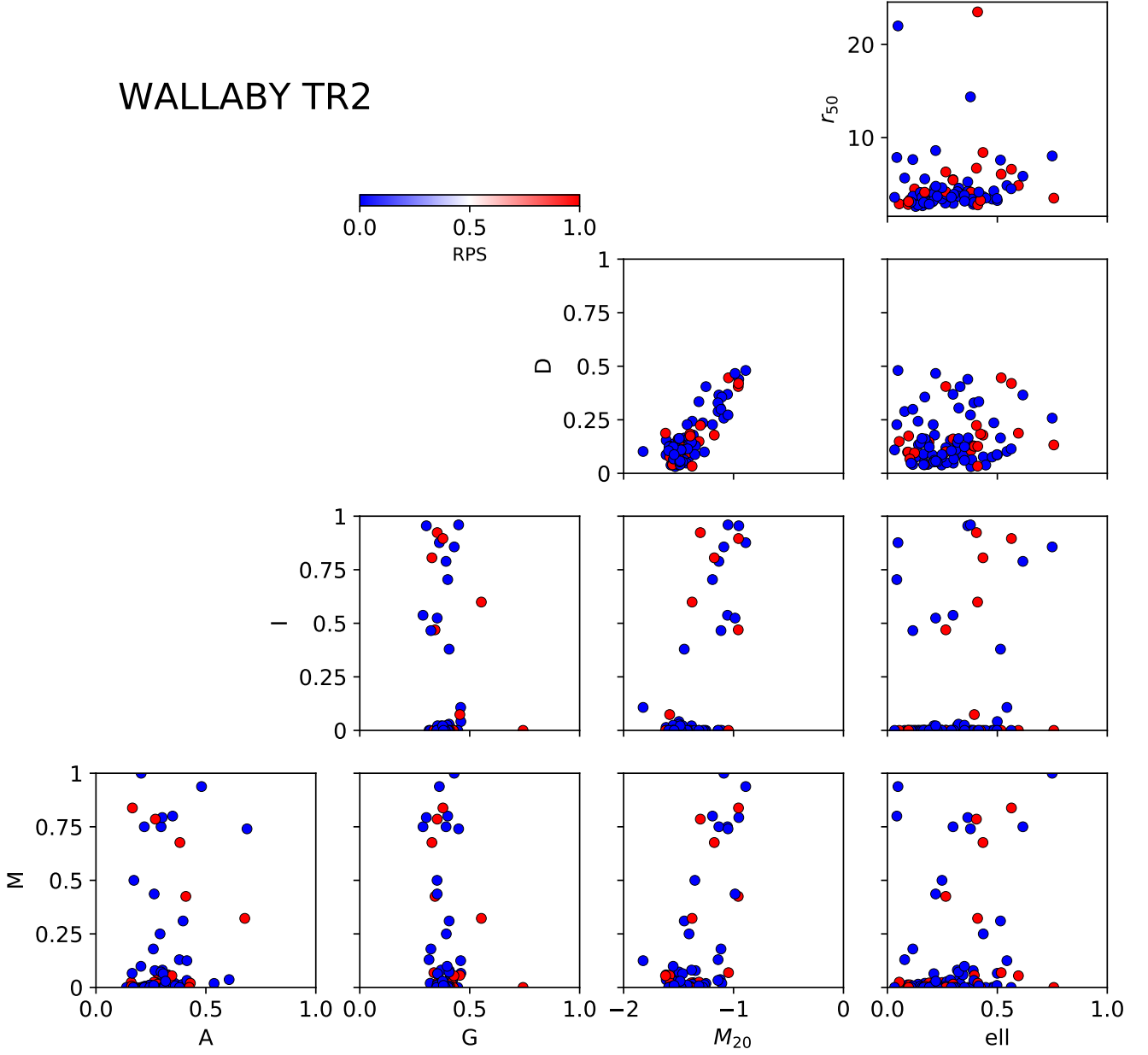
**Figure 9.** The C-A-G-M20-ellipticity H I morphometric feature space with the RPS flag from Wang et al. (2021) marked with the color: red galaxies are undergoing RPS.

in deep imaging (Bouquin et al. 2015; Bouquin & de Paz 2016; Bouquin et al. 2018). However, with WALLABY and GALEX information combined, FUV morphometrics within the H I contour poses a viable way to identify these XUV disks (5/32, 15%). The fraction of XUV disks is lower than in the field, typically  $\sim 40\%$ . The prerequisites are ASKAP H I and GALEX FUV imaging, where even the shallowest GALEX exposures are sufficient thanks to the accurate delineation using the outer H I contour using ASKAP.

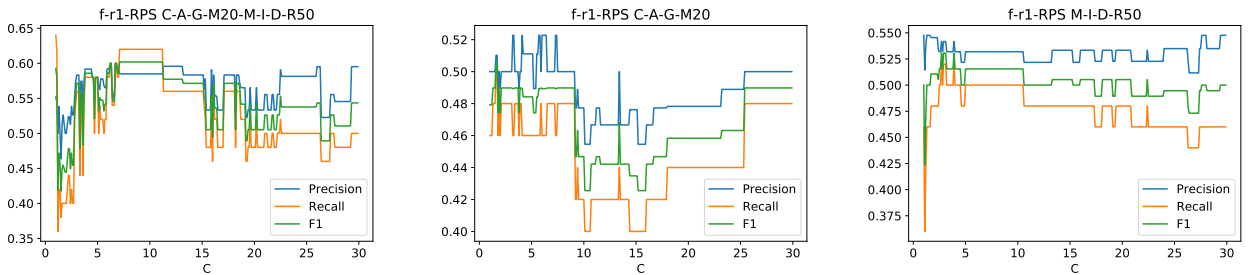
We computed the morphometrics over the moment-0 maps for all 272 galaxies in the Hydra TR2 catalog to act as the feature space to explore. A cluster population of galaxies at comparable distances

with similar signal-to-noise limits on the observations is an ideal sample

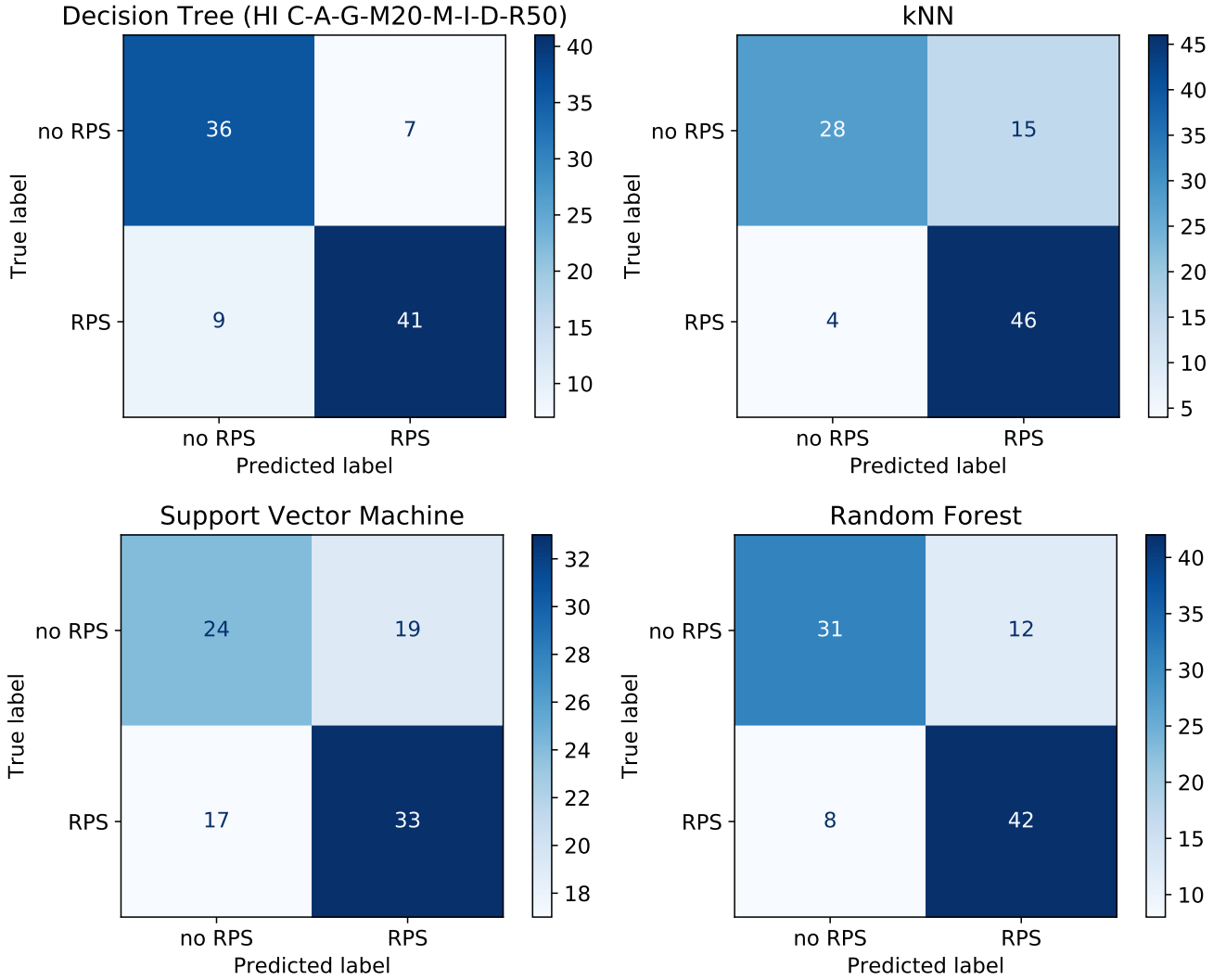
Ram-pressure stripping is considered a principal avenue for galaxy transformation, in addition to tidal interactions, especially in dense environments. Identifying galaxies that are potentially undergoing ram-pressure stripping is therefore a motivator for H I (morphological) studies. We compare our morphometric feature space to the RPS labels from Wang et al. (2021) and find that Machine Learning classifiers –decision trees, Support Vector Machines, K-Nearest Neighbour, Random Forest– do not fully separate RPS galaxies in H I morphometric space yet. Decision trees are impractically deep, the SVM struggle with precision but this problem is slightly better



**Figure 10.** The M-I-D-ell H I morphometric feature space with the RPS flag from Wang et al. (2021) marked with the color: red galaxies are undergoing RPS.



**Figure 11.** The metrics (Precision, recall and F1) as a function of the regularization parameter  $C_{reg}$  of the SVM for three instances of the feature space: CAGM20MIDR50, CAGM20, and MIDR50 for the r1-RPS flag from Wang et al. (2021). SVM performs poorly in separating RPS from non-RPS galaxies with every configuration of the feature space and regularization parameter  $C_{reg}$ .



**Figure 12.** The classification matrices for the full H I feature space (C-A-G-M20-M-I-D-R50) and the r1-RPS flag from Wang et al. (2021) for the decision tree, kNN, SVM and RF.

features	Accuracy	Precision	Recall	F1
<i>f<sub>RPS</sub></i>				
C-A-G-M20-M-I-D-R50	0.74	0.66	0.86	0.37
C-A-G-M20	0.74	0.67	0.80	0.36
M-I-D-R50	0.64	0.56	0.89	0.34
<i>f<sub>r1,RPS</sub></i>				
C-A-G-M20-M-I-D-R50	0.61	0.63	0.66	0.32
C-A-G-M20	0.60	0.63	0.64	0.32
M-I-D-R50	0.49	0.62	0.16	0.13

**Table 5.** The SVM performance for different sub-sets of the feature space with fixed  $C_{reg} = 5$ .  $F1 = 2 \times (\text{Recall} \times \text{Precision}) / (\text{Recall} + \text{Precision})$  is the weighted average of precision and recall.

features	Accuracy	Precision	Recall	F1
<i>f<sub>RPS</sub></i>				
C-A-G-M20-M-I-D-R50	0.52	0.48	1.00	0.32
C-A-G-M20	0.51	0.47	1.00	0.32
M-I-D-R50	0.51	0.47	1.00	0.32
<i>f<sub>r1,RPS</sub></i>				
C-A-G-M20-M-I-D-R50	0.52	0.73	0.16	0.13
C-A-G-M20	0.56	0.55	0.94	0.35
M-I-D-R50	0.52	0.73	0.16	0.13

**Table 6.** The SVM performance for different sub-sets of the feature space with fixed  $C_{reg} = 5$  and a polynomial kernel convolution.

with kNN and RF. On the whole performance is middling (e.g.  $\sim 0.6$  accuracy, precision and recall for SVM) to fairly decent ( $\sim 0.8$  for the RF and kNN). On the whole, it points to reasonably good performance (80% precision and recall) is possible for populations and

acceptable performance for individual galaxies i.e. a ML algorithm can give a probability it is undergoing ram pressure stripping) with a larger training set.

The feature space of H I morphometrics used is either the full, (C-A-G-M20-M-I-D-R50) or the first (C-A-G-M20) or second half



features	Accuracy	Precision	Recall	F1
<hr/>				
$\hat{f}_{RPS}$				
C-A-G-M20-M-I-D-R50	0.84	0.74	0.98	0.42
C-A-G-M20	0.86	0.79	0.93	0.43
M-I-D-R50	0.71	0.62	0.91	0.37
<hr/>				
$\hat{f}_{r1,RPS}$				
C-A-G-M20-M-I-D-R50	0.78	0.78	0.84	0.40
C-A-G-M20	0.74	0.72	0.84	0.39
M-I-D-R50	0.77	0.78	0.80	0.40

**Table 7.** The random forest performance for different sub-sets of the feature space.

(M-I-D-R50) of that space. No clear benefit of one grouping over the other in ML performance can be identified. With the H I morphometric space and rather simple machine learning tools, one can identify populations of galaxies that are undergoing ram-pressure stripping reasonably well (precision and recall of  $\sim 80\%$ ). This may be useful to construct samples for further inspection but the performance is not yet good enough to reliably infer fractions of galaxies undergoing ram-pressure stripping throughout a survey without further checks. One of our original goals was to ascertain whether this H I morphometric parameter space is good enough to aid in the identification of RPS galaxies. This appears to be feasible. Both larger test samples and possible direct use of the H I maps as input may improve results in the future.

- The morphometrics can be derived, even with the marginally resolved sources typical for WALLABY (Figure 3).
- They are a reasonably parameter space to identify galaxies that may be undergoing ram pressure (Figure 3).
- Using an H I contour for GALEX nearby galaxy data is a good segmentation choice (Figure 8).
- If the whole morphometric space is needed or a subset is viable is not yet clear (Table 3 – 7).
- A random forest is the best algorithm to classify based on the morphometric space. This has been found on JWST morphology papers as well independently. This saves substantial time/effort compared to Convolutional Neural Networks etc. (Figure 12 and Table 7)

For the full WALLABY survey (Koribalski et al. 2020), one could apply these to all the solidly detected and partially resolved (e.g. 3-4 beams across) objects to create a parameter space for machine learning to identify populations such as galaxies undergoing ram-pressure stripping or gravitational interactions.

## 8 ACKNOWLEDGEMENTS

The Australian SKA Pathfinder is part of the Australia Telescope National Facility which is managed by CSIRO. Operation of ASKAP is funded by the Australian Government with support from the National Collaborative Research Infrastructure Strategy. ASKAP uses the resources of the Pawsey Supercomputing Centre. Establishment of ASKAP, the Murchison Radio-astronomy Observatory and the Pawsey Supercomputing Centre are initiatives of the Australian Government, with support from the Government of Western Australia and the Science and Industry Endowment Fund. We acknowledge the Wajarri Yamatji as the traditional owners of the Observatory site.

This research made use of Astropy, a community-developed core Python package for Astronomy (Astropy Collaboration et al. 2013, 2018).

AB acknowledges support from the Centre National d’Etudes Spatiales (CNES), France.

SHOH acknowledges a support from the National Research Foundation of Korea (NRF) grant funded by the Korea government (Ministry of Science and ICT: MSIT) (No. NRF-2020R1A2C1008706).

FB acknowledges funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement No.726384/Empire)

BWH is supported by an Enhanced Mini-Grant (EMG). The material is based upon work supported by NASA Kentucky under NASA award No: 80NSSC20M0047.

BWH would like to thank Robin Allen and the teachers at Engelhart elementary for their hard work on distance learning during the COVID-19 lockdown and the game Minecraft for allowing for enough additional time to write this manuscript.

## 9 DATA AVAILABILITY

Electronic versions of the H I and UV morphometric catalogs and the segmentation map are included as supplemental information for this manuscript.

The full 30 square degree spectral line cubes can be accessed on the CSIRO ASKAP Science Data Archive<sup>3</sup> (CASDA, Chapman 2015; Huynh et al. 2020) with the DOI: <https://doi.org/10.25919/5f7bde37c20b5> for this data-set.

## REFERENCES

- Abraham R. G., Valdes F., Yee H. K. C., van den Bergh S., 1994, *ApJ*, 432, 75
- Abraham R. G., van den Bergh S., Nair P., 2003, *ApJ*, 588, 218
- Alberts S., et al., 2011, *ApJ*, 731, 28
- Astropy Collaboration et al., 2013, *A&A*, 558, A33
- Astropy Collaboration et al., 2018, *AJ*, 156, 123
- Begeman K. G., 1989, *A&A*, 223, 47
- Bershady M. A., Jangren A., Conselice C. J., 2000, *AJ*, 119, 2645
- Bertin E., Arnouts S., 1996, *A&AS*, 117, 393
- Bigiel F., Leroy A., Seibert M., Walter F., Blitz L., Thilker D., Madore B., 2010, *ApJ*, 720, L31
- Bigiel F., Leroy A., Walter F., 2011, in Alves J., Elmegreen B. G., Girart J. M., Trimble V., eds, Vol. 270, *Computational Star Formation*. pp 327–334
- Bignone L. A., Tissera P. B., Sillero E., Pedrosa S. E., Pellizza L. J., Lambas D. G., 2017, *MNRAS*, 465, 1106
- Boomsma R., Oosterloo T. A., Fraternali F., van der Hulst J. M., Sancisi R., 2008, *A&A*, 490, 555
- Bosma A., 1978, PhD thesis, -
- Bouquin A. Y. K., de Paz A. G., 2016, *Proceedings of the International Astronomical Union*, 11, 257
- Bouquin A. Y. K., et al., 2015, *ApJ*, 800, L19
- Bouquin A. Y. K., et al., 2018, *ApJS*, 234, 18
- Bournaud F., Jog C. J., Combes F., 2005, *A&A*, 437, 69
- Bresolin F., Ryan-Weber E., Kennicutt R. C., Goddard Q., 2009, *ApJ*, 695, 580
- Bruzzese S. M., Meurer G. R., Lagos C. D. P., Elson E. C., Werk J. K., Blakeslee J. P., Ford H., 2015, *MNRAS*, 447, 618

<sup>3</sup> <https://data.csiro.au/domain/casdaObservation>

- Chapman J. M., 2015, in IAU General Assembly. p. 2232458
- Chung A., van Gorkom J. H., Kenney J. D. P., Crowl H., Vollmer B., 2009, *AJ*, 138, 1741
- Condon J. J., Cotton W. D., Greisen E. W., Yin Q. F., Perley R. A., Taylor G. B., Broderick J. J., 1998, *AJ*, 115, 1693
- Conselice C. J., 2003, *ApJS*, 147, 1
- Davenport J. R. A., 2015, I really want to find an astronomical application for morphometrics, <https://twitter.com/jradavenport/status/571064841344917504>
- Dong H., Calzetti D., Regan M., Thilker D., Bianchi L., Meurer G. R., Walter F., 2008, *AJ*, 136, 479
- Elagali A., et al., 2019, *MNRAS*, 487, 2797
- Elson E. C., de Blok W. J. G., Kraan-Korteweg R. C., 2011, *MNRAS*, 415, 323
- Ferguson A. M. N., Wyse R. F. G., Gallagher J. S., Hunter D. A., 1998, *ApJ*, 506, L19
- Florian M. K., Li N., Gladders M. D., 2016, *ApJ*, 832, 168
- For B. Q., et al., 2019, *MNRAS*, 489, 5723
- Freeman P. E., Izbicki R., Lee A. B., Newman J. A., Conselice C. J., Koeckmoer A. M., Lotz J. M., Mozena M., 2013, *MNRAS*, 434, 282
- Giese N., van der Hulst T., Serra P., Oosterloo T., 2016, *MNRAS*, 461, 1656
- Gil de Paz A., et al., 2005, *ApJ*, 627, L29
- Gil de Paz A., et al., 2007, *ApJ*, 661, 115
- Gini C., 1912
- Graham A. W., Driver S. P., 2005, *Publications of the Astronomical Society of Australia*, 22, 118
- Graham A. W., Driver S. P., Petrosian V., Conselice C. J., Bershady M. A., Crawford S. M., Goto T., 2005, *AJ*, 130, 1535
- Heald G., Allan J., Zschaechner L., Kamphuis P., Rand R., Józsa G., Gentile G., 2011a, in C. Carignan, F. Combes, & K. C. Freeman ed., *IAU Symposium Vol. 277, IAU Symposium*. pp 59–62
- Heald G., et al., 2011b, *A&A*, 526, A118
- Heiner J. S., Vázquez-Semadeni E., Ballesteros-Paredes J., 2014, preprint
- Hibbard J. E., van Gorkom J. H., Rupen M. P., Schiminovich D., 2001, in Hibbard J. E., Rupen M., van Gorkom J. H., eds, *Astronomical Society of the Pacific Conference Series Vol. 240, Gas and Galaxy Evolution*. pp 657–
- Holwerda B. W., 2005, *astro-ph/0512139*
- Holwerda B. W., Pirzkal N., de Blok W. J. G., Bouchard A., Blyth S.-L., van der Heyden K. J., Elson E. C., 2011a, *MNRAS*, 416, 2401
- Holwerda B. W., Pirzkal N., de Blok W. J. G., Bouchard A., Blyth S.-L., van der Heyden K. J., Elson E. C., 2011b, *MNRAS*, 416, 2415
- Holwerda B. W., Pirzkal N., Cox T. J., de Blok W. J. G., Weniger J., Bouchard A., Blyth S.-L., van der Heyden K. J., 2011c, *MNRAS*, 416, 2426
- Holwerda B. W., Pirzkal N., de Blok W. J. G., Bouchard A., Blyth S.-L., van der Heyden K. J., 2011d, *MNRAS*, 416, 2437
- Holwerda B. W., Pirzkal N., de Blok W. J. G., van Driel W., 2011e, *MNRAS*, 416, 2447
- Holwerda B. W., Pirzkal N., Heiner J. S., 2012, *MNRAS*, 427, 3159
- Hotan A. W., et al., 2014, *PASA*, 31, e041
- Hotan A. W., et al., 2021, *Publ. Astron. Soc. Australia*, 38, e009
- Hunter D. A., et al., 2018, *AJ*, 156, 21
- Huynh M., Dempsey J., Whiting M. T., Ophel M., 2020, in Ballester P., Ibsen J., Solar M., Shortridge K., eds, *Astronomical Society of the Pacific Conference Series Vol. 522, Astronomical Data Analysis Software and Systems XXVII*. p. 263
- Jog C. J., Combes F., 2009, *Phys. Rep.*, 471, 75
- Johnston S., 2007, in *From Planets to Dark Energy: the Modern Radio Universe*. October 1-5 2007, The University of Manchester, UK. Published online at SISSA, *Proceedings of Science*, p.6.
- Johnston S., et al., 2007, *Publications of the Astronomical Society of Australia*, 24, 174
- Johnston S., et al., 2008, *Experimental Astronomy*, 22, 151
- Johnston S., Feain I. J., Gupta N., 2009, in D. J. Saikia, D. A. Green, Y. Gupta, & T. Venturi ed., *Astronomical Society of the Pacific Conference Series Vol. 407, Astronomical Society of the Pacific Conference Series*. pp 446–
- Kleiner D., et al., 2019, *MNRAS*, 488, 5352
- Koda J., et al., 2015, *ApJ*, 802, L24
- Koribalski B. S., 2017, in Gil de Paz A., Knapen J. H., Lee J. C., eds, Vol. 321, *Formation and Evolution of Galaxy Outskirts*. pp 232–234
- Koribalski B. S., 2022, arXiv e-prints, p. arXiv:2208.08245
- Koribalski B. S., López-Sánchez Á. R., 2009, *MNRAS*, 400, 1749
- Koribalski B. S., et al., 2020, *Ap&SS*, 365, 118
- Lee-Waddell K., et al., 2019, *MNRAS*, 487, 5248
- Lelièvre M., Roy J.-R., 2000, *AJ*, 120, 1306
- Lemonias J. J., et al., 2011, *ApJ*, 733, 74
- Leroy A. K., Walter F., Brinks E., Bigiel F., de Blok W. J. G., Madore B., Thornley M. D., 2008, *AJ*, 136, 2782
- Lisker T., 2008, *ApJS*, 179, 319
- López-Sánchez Á. R., Westmeier T., Esteban C., Koribalski B. S., 2015, *MNRAS*, 450, 3381
- López-Sánchez Á. R., Lagos C. D. P., Young T., Jerjen H., 2018, *MNRAS*, 480, 210
- Lotz J. M., Primack J., Madau P., 2004, *AJ*, 128, 163
- Lotz J. M., Jonsson P., Cox T. J., Primack J. R., 2008, *MNRAS*, 391, 1137
- Lotz J. M., Jonsson P., Cox T. J., Primack J. R., 2010, *MNRAS*, 404, 590
- Lotz J. M., Jonsson P., Cox T. J., Croton D., Primack J. R., Somerville R. S., Stewart K., 2011, *ApJ*, 742, 103
- McConnell D., et al., 2016, *PASA*, 33, e042
- Meurer G. R., 2017, in Gil de Paz A., Knapen J. H., Lee J. C., eds, Vol. 321, *Formation and Evolution of Galaxy Outskirts*. pp 172–179
- Meurer G. R., Carignan C., Beaulieu S. F., Freeman K. C., 1996, *AJ*, 111, 1551
- Meurer G. R., Staveley-Smith L., Killeen N. E. B., 1998, *MNRAS*, 300, 705
- Meurer G. R., et al., 2009, *ApJ*, 695, 765
- Moffett A. J., Kannappan S. J., Baker A. J., Laine S., 2012, *ApJ*, 745, 34
- Moore E. M., Gottesman S. T., 1998, *MNRAS*, 294, 353
- Noordermeer E., van der Hulst J. M., Sancisi R., Swaters R. A., van Albada T. S., 2005, *A&A*, 442, 137
- Pedregosa F., et al., 2011, *Journal of Machine Learning Research*, 12, 2825
- Peth M. A., et al., 2016, *MNRAS*, 458, 963
- Planck Collaboration et al., 2016, *A&A*, 596, A100
- Reynolds T. N., et al., 2019, *MNRAS*, 482, 3591
- Reynolds T. N., Westmeier T., Staveley-Smith L., Chauhan G., Lagos C. D. P., 2020, *MNRAS*, 493, 5089
- Reynolds T. N., et al., 2021, *MNRAS*, 505, 1891
- Rodríguez-Gómez V., et al., 2019, *MNRAS*, 483, 4140
- Sancisi R., Fraternali F., Oosterloo T., van der Hulst T., 2008, *A&ARv*, 15, 189
- Serra P., et al., 2015a, *MNRAS*, 448, 1922
- Serra P., et al., 2015b, *MNRAS*, 452, 2680
- Stinson G. S., Dalcanton J. J., Quinn T., Gogarten S. M., Kaufmann T., Wadsley J., 2009, *MNRAS*, 395, 1455
- Swaters R. A., van Albada T. S., van der Hulst J. M., Sancisi R., 2002, *A&A*, 390, 829
- Thilker D. A., 2016, *Proceedings of the International Astronomical Union*, 11, 298
- Thilker D. A., et al., 2005a, *ApJ*, 619, L67
- Thilker D. A., et al., 2005b, *ApJ*, 619, L79
- Thilker D. A., et al., 2007, *ApJS*, 173, 538
- Walter F., Brinks E., de Blok W. J. G., Bigiel F., Kennicutt R. C., Thornley M. D., Leroy A., 2008, *AJ*, 136, 2563
- Wang Y., Ferland G. J., Lykins M. L., Porter R. L., van Hoof P. A. M., Williams R. J. R., 2014, *MNRAS*, 440, 3100
- Wang J., et al., 2021, *ApJ*, 915, 70
- Watson L. C., Martini P., Lisenfeld U., Böker T., Schinnerer E., 2016, *MNRAS*, 455, 1807
- Watts A. B., Meurer G. R., Lagos C. D. P., Bruzese S. M., Kroupa P., Jerabkova T., 2018, *MNRAS*, 477, 5554
- Werk J. K., Putman M. E., Meurer G. R., Thilker D. A., Allen R. J., Bland-Hawthorn J., Kravtsov A., Freeman K., 2010, *ApJ*, 715, 656
- Westmeier T., et al., 2021
- Yitzhaki S., 1991, *American Statistical Association*, 9, 235
- Zaritsky D., Christlein D., 2007, *AJ*, 134, 135
- Zaritsky D., Rix H.-W., 1997, *ApJ*, 477, 118

- Zschaechner L. K., Rand R. J., Heald G. H., Gentile G., Kamphuis P., 2011, *ApJ*, 740, 35
- de Blok W. J. G., Walter F., Brinks E., Trachternach C., Oh S.-H., Kennicutt R. C., 2008, *AJ*, 136, 2648
- de Paz A. G., et al., 2008, in Funes J. G., Corsini E. M., eds, *Astronomical Society of the Pacific Conference Series Vol. 396*, Astronomical Society of the Pacific Conference Series. pp 197–+
- van Eymeren J., Jütte E., Jog C. J., Stein Y., Dettmar R. J., 2011a, *A&A*, 530, A29
- van Eymeren J., Jütte E., Jog C. J., Stein Y., Dettmar R. J., 2011b, *A&A*, 530, A30

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