The automated detection and multiwavelength properties of LADUMA HI galaxies

Leyya Stockenstroom

Supervised by Dr. Jacinta Delhaize (UCT)

Co-Supervised by Prof. Matt Jarvis (Oxford, UWC)

A dissertation submitted in partial fulfilment of the requirements for the degree of Master of Science in Astronomy

University of Cape Town

February 2025

Abstract

In this study, we investigate the neutral atomic hydrogen (HI) and star formation history (SFH) properties of 104 low redshift ($0 < z_{HI} < 0.088$) galaxies from the Looking At the Distant Universe with the MeerKAT Array (LADUMA) survey. LADUMA is a deep and narrow HI emission survey underway with the MeerKAT telescope. We perform automated source finding over the low redshift Early Science LADUMA data cube using the SoFiA-2 algorithm. The source finding mask created in this work will be used to clean the first public data release of the LADUMA data cubes, which will be outlined in an upcoming paper. We present the confirmed SoFiA-2 detections of HI and compare the results to those produced by manual and semi-automated methods of source finding. We find generally good agreement between the methods, suggesting that SoFiA-2 will be an efficient and effective method on subsequent LADUMA data cubes. Additionally, we perform spectral energy distribution (SED) fitting to far-ultraviolet (FUV) to far-infrared (FIR) multiband photometry of our HI sample, thereby determining their stellar masses (M_*), star formation rates (SFRs), and star formation histories (SFHs). We explore the relationship between galaxies' HI and star formation history properties (HI-SFH). To do this, we examine the lookback time of peak star formation (τ_{peak}) versus the HI mass to stellar mass fraction (M_{HI}/M_*) and the peak star formation rate versus M_{HI}/M_*. The latter is investigated here for the first time, to the best of our knowledge. Our findings reveal a strong anti-correlation between M_{HI}/M_* and $\mathrm{SFR_{peak}}$ of our HI sample, with τ_{peak} decreasing as M_{HI}/M_* increases. This suggests that galaxies with lower M_{HI}/M_* have previously undergone more aggressive periods of star formation than galaxies with higher M_{HI}/M_*. Additionally, galaxies with higher M_{HI}/M_* exhibit rising star formation rates at the time of observation and are likely to reach their peak star formation rate in future stages of their evolution. These findings serve as a low redshift pilot study, which can be applied to and compared with similar source finding and HI-SFH studies in upcoming higher redshift data from the LADUMA survey.

Acknowledgments

Finishing a masters dissertation is no easy task and would not have been possible without support from others. I would first like to thank Dr. Jacinta Delhaize and Prof. Matt Jarvis for sharing their knowledge, guidance, and all the opportunities they provided me that has helped me to grow into an astronomer. Thank you to the LADUMA science team for processing and providing the HI data cube needed for this work. Furthermore, I would like to express my gratitude to Dr. Luke Davies, for providing me with the DEVILS images and catalogues, and to Prof Aaron Robotham for sharing his knowledge on ProFound and ProSpect. All of you provided invaluable advice throughout this project.

To my friends and colleagues at the UCT Department of Astronomy, thank you for your advice, support, and encouragement. Finally, thank you to my mom for her unconditional love and support while I pursue my dream of becoming an astrophysicist.

We acknowledge the use of the ilifu cloud computing facility – www.ilifu.ac.za, a partnership between the University of Cape Town, the University of the Western Cape, Stellenbosch University, Sol Plaatje University, and the Cape Peninsula University of Technology. The ilifu facility is supported by contributions from the Inter-University Institute for Data Intensive Astronomy (IDIA – a partnership between the University of Cape Town, the University of Pretoria, and the University of the Western Cape), the Computational Biology division at UCT, and the Data Intensive Research Initiative of South Africa (DIRISA).

This work made use of the CARTA (Cube Analysis and Rendering Tool for Astronomy) software (DOI 10.5281/zenodo.3377984 – https://cartavis.github.io).

The MeerKAT telescope is operated by the South African Radio Astronomy Observatory, which is a facility of the National Research Foundation, an agency of the Department of Science and Innovation.

The financial assistance of the South African Radio Astronomy Observatory (SARAO) towards this research is hereby acknowledged (www.sarao.ac.za).

We also acknowledge the support of an Africa-Oxford (Afox-290) travel award.

We acknowledge partial research support by the National Research Foundation of South Africa (Ref Number CSUR240426216203).

Table of Contents

List of Figures

List of Tables

1 | Introduction

1.1 Galaxy evolution

Galaxies are incredibly beautiful and extremely diverse structures of gravitationally bound stars, stellar remnants, dust, and dark matter. The nature, origin, and evolution of galaxies are of great complexity and have intrigued astronomers since galaxies were first observed. The complexity of galaxy evolution stems from the many physical processes which vary in physical scales and time scales. This spans from smaller-scale, internal processes of galaxies, such as the formation and evolution of individual stars, to much larger-scale interactions between galaxies. The combination of all these processes are thought to drive the overall evolution of galaxies [e.g. Tinsley 1980, Davé et al. 2012, Conselice 2014, Lizarraga et al. 2024].

The majority of the processes that occur within a galaxy far exceed that of a human life. As a result, it is impossible to observe the complete evolution of an individual galaxy. Therefore, to piece together a picture of galaxy evolution, we can study multiple galaxies at varying cosmological distances from us. More distant galaxies have greater observed redshifts (z), meaning their emission is shifted to lower frequencies due to their recessional velocity (ν_{rec}). Since ν_{rec} is due to the expansion of the Universe, which is known as Hubble flow, the z of a galaxy is greater the further away it is from us, i.e, when $z << 1$:

$$z = \frac{\nu_{rec}}{c} = \frac{H_0 D}{c}, \tag{1.1}$$

where H_0 is Hubble's constant, D is the proper (physical) distance, and c is the speed of light.

More distant galaxies have a longer lookback time due to the finite speed of light. Therefore, by observing high-redshift galaxies, we are observing them as they were when the Universe was younger. Consequently, by observing galaxies at varying redshifts, we are able to observe them at different stages of their cosmic evolution.

Various properties of galaxies have been observed to change with redshift, indicating that galaxies have evolved over time. These properties include the cosmic star formation rate density (Ψ_{stars}), the stellar mass density (ρ_{stars}), the cosmic density of atomic gas (ρ_{HI}) and the cosmic molecular hydrogen density (ρ_{H_2}) [reviewed by Madau and Dickinson 2014 and Walter et al. 2020a and illustrated in Figure 1.1]. A decline in Ψ_{stars} has been observed at $z \lesssim 2$, which roughly corresponds to the decline observed in ρ_{H2} (shown in top left and bottom right panels of Figure 1.1). However, this trend is not observed in the evolution of ρ_{HI} and ρ_{stars}, with ρ_{stars} rapidly increasing and ρ_{HI} remaining roughly constant across redshifts (shown in top right and bottom left panels of Figure 1.1).

These observations are perplexing, as it is thought that neutral atomic hydrogen (HI)

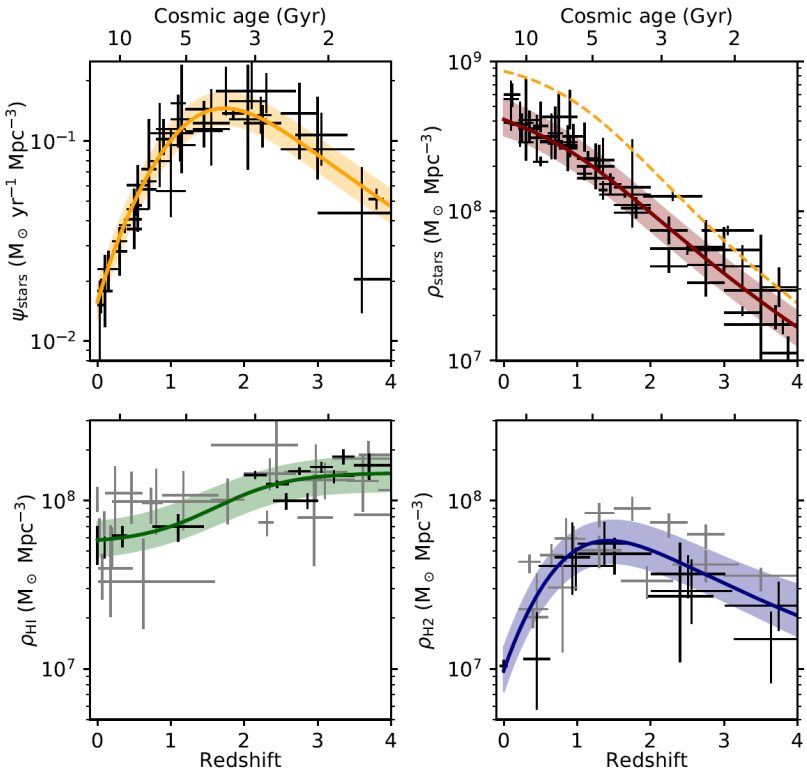


Figure 1.1: Redshift evolution of different baryonic components in galaxies, reproduced from Walter et al. (2020b). (Top left) the cosmic star formation rate density and (top right) stellar mass density from Madau and Dickinson (2014). (Bottom left) the cosmic density of HI. (Bottom right) the cosmic density of H_2. The solid lines are the best-fit function and the shaded regions mark the 1σ region. The yellow dashed line in the stellar mass density plot represents the integration of the best fit function form to the star formation rate density.

condenses into giant molecular clouds (H_2) as it accretes from the galactic halo. Star formation then occurs within these giant molecular clouds [Leroy et al. 2008]. This has been seen in resolved imaging of nearby galaxies, which show star-forming regions (traced via far-ultraviolet observations) correspond to regions with a high surface density of H_2 (traced via CO data) [e.g. Bigiel et al. 2008].

We therefore expect HI to be the reservoir ultimately fueling star formation [Kauffmann 2015, Zhou et al. 2018]. However, if this is the case, then why do we observe a decline in Ψ_{stars} while ρ_{HI} remains constant? Adding to the confusion, if it is assumed that the amount of H_2 gas seen at $z \sim 1.5$ (where ρ_{H_2} peaks) will be fully transformed into stars, this would account for, at most, a quarter of the ρ_{stars} observed at z=0 [Walter et al. 2020a]. So, what is the cause of the high cosmic stellar mass density? Has the flow of HI from the extended reservoirs to the centres of the dark matter halos decreased, resulting in a decline ρ_{H_2} at $z \lesssim 1.5$? Is HI being replenished, resulting in a roughly constant ρ_{HI}?

One way to begin addressing these mysteries is by carefully examining how the HI content (the primary fuel) and star forming properties (the end point) of galaxies are related. Despite the wide range of morphologies, masses, and sizes of the galactic population, they have been observed to follow certain scaling relations. For example, there is a correlation between the HI mass (M_{HI}) and the stellar mass (M_*) of a galaxy, up to a certain stellar mass, where a turnover is observed [e.g. Maddox et al. 2015, Sinigaglia et al. 2022, Rhee et al. 2023]. The M_{HI}-M_* scaling relation implies that more massive galaxies have larger HI reservoirs, up to $\log(M_*) \sim 10 M_\odot$, suggesting a saturation point in HI content.

Additionally, a strong correlation is also observed between the star formation rate (SFR) of galaxies and their M_*, which has been dubbed the star forming main sequence (SFMS; Noeske et al. 2007). Similar to the M_*-M_{HI} scaling relation, the SFMS is linear up to stellar masses of $\log(M_*) \sim 10.5\ M_\odot$, where it then begins to drop off [e.g. Speagle et al. 2014, Fraser-McKelvie et al. 2021, Popesso et al. 2023]. The reason for the decrease of SFR in high stellar mass galaxies remains unknown, however many studies have examined the link between the SFMS drop-off and secular processes such as environmental effects [e.g Gavazzi et al. 2015], AGN feedback [e.g. Mancuso et al. 2016] and halo quenching [e.g. Popesso et al. 2019]. Overall, the drop off in the SFMS is thought to be due to a combination of the evolution of their morphology and environmental effects [Erfanianfar et al. 2016]. By examining and understanding more scaling relations, we may be able to gain insight into the inner workings of galaxy evolution. However, to calculate the properties needed to build these scaling relations, we must conduct multiwavelength studies of galaxies. This is discussed further in the following section.

1.2 Multiwavelength observations of galaxies

Different components of a galaxy emit at different wavelengths of the electromagnetic spectrum (as shown in Figure 1.2). For example, stellar emission dominates the ultraviolet (UV) to near-infrared (NIR) wavelengths (shown by the blue line in Figure 1.2), with short lived OB stars dominating the near-ultraviolet (NUV). Because of this, NUV emission has been used to trace the star formation rate (SFR) of galaxies [Salim et al. 2007, Kennicutt and Evans 2012]. Additionally, NIR emission has been used to track the evolution of the stellar mass density [Dickinson et al. 2003]. However, these measurements are biased by the presence of dust, which absorbs the energy produced by stars and reradiates it into the mid- (MIR) and far-infrared (FIR).

To measure the energy emitted by the different components of a galaxy, we extract the photometry of said galaxy in multiple wavebands. Extracting multiwavelength photometry consists of observing a galaxy with multiple telescopes (which operate in different wavebands), in multiple photometric bands and measuring the flux density of said galaxy in each photometric band. By plotting the extracted photometry of a galaxy, we can examine the amount of energy emitted across all wavebands. This is known as the observed spectral energy distribution (SED) of a galaxy (shown by the black points in Figure 1.2). All of the different processes within a galaxy dominate at different wavelengths and these

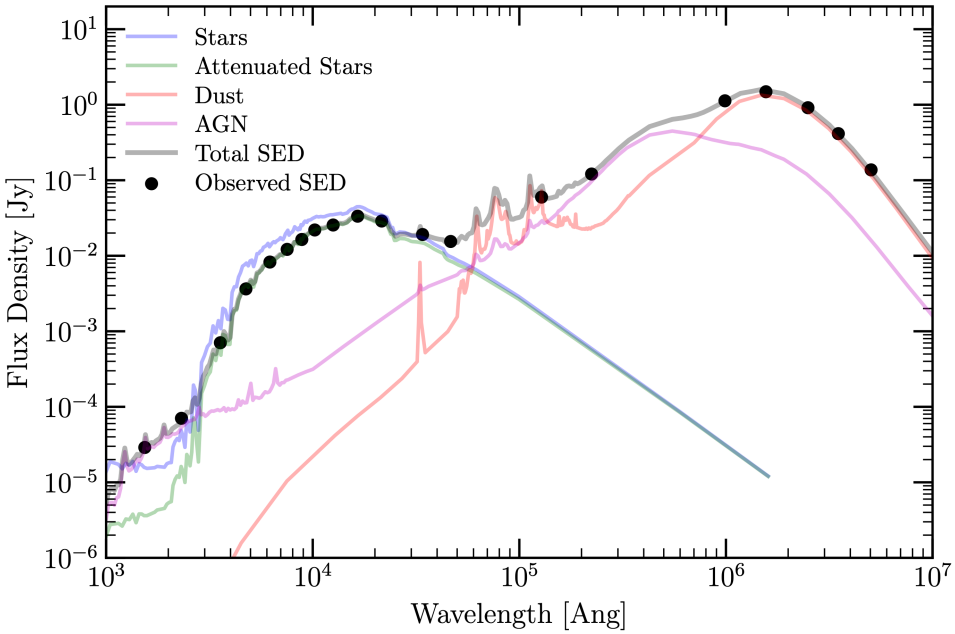


Figure 1.2: Observed and total spectral energy distribution (SED) of a galaxy and the contributing components from ultraviolet (UV) to far-infrared (FIR). The black points represent the photometry extracted in multiple wavebands (which make up the observed SED). The grey line represents the total SED. The blue line is energy contributed by stars while the green line is the energy contributed by attenuated stars. The red line is the emission affected by dust and the purple line is the energy contributed by an active galactic nucleus (AGN). Image created by Dr. Sabine Bellstedt.

processes, along with the energy transfer between them, leave an imprint on the overall shape of the observed SED [Dey et al. 2024]. By fitting the total SED (shown by the grey line in Figure 1.2), we can simultaneously model the individual components of a galaxy.

By breaking down the far-ultraviolet (FUV) to FIR SED of a galaxy into contributions from stars, AGN and dust, we can extract information about the star formation history (SFH), star formation rate (SFR), stellar mass and dust properties of said galaxy [e.g. Da Cunha et al. 2008, Noll et al. 2009, Bellstedt et al. 2020b, Thorne et al. 2021]. This is done by simultaneously modelling the individual components of a galaxy to obtain a total SED that fits the observed SED. Multiple different codes have been developed to perform SED fitting. Examples include ProSpect [Robotham et al. 2020], magphys [Da Cunha et al. 2008], cigale [Noll et al. 2009] and bagpipes [Carnall et al. 2018]. A majority of these codes perform SED fitting by 1) modelling the stellar component of galaxies by combining various stellar templates, which express the age, metallicity and initial mass function (IMF) of the stellar population [e.g. Bruzual and Charlot 2003, Chabrier 2003], 2) parameterising the SFH of the galaxy and 3) modelling the dust attenuation [e.g. Charlot and Fall 2000] and dust emission [e.g. Dale et al. 2014] while usually assuming an energy balance between the attenuated stellar light (shown by the green line in Figure 1.2) and the re-emission at longer wavelengths (shown by the red line

in Figure 1.2) [as per MAGPHYS; Da Cunha et al. 2008 and PROSPECT; Robotham et al. 2020]. At the end of this process, the observed SED will be fit to a total SED template, which has a certain stellar mass and SFR attributed to it.

The parametrisation of the SFH is necessary in the SED fitting process to decipher the distribution of stellar ages within a given galaxy [Schreiber et al. 2018]. The SFH reflects the timeline of a galaxy's star formation activity, plotting the history of the star formation rate of a galaxy. The best approach to model the SFH is still an ongoing debate which is largely split between using a parametric or non-parametric approach (a misleading naming convention, as both are modelled via parameters) [Carnall 2019, Leja et al. 2019]. Non-parametric SFHs require more free parameters and are described as a step function, providing more flexibility in the shape. Whereas parametric modelling assumes the SFH will take a functional form (e.g. exponentially declining, exponentially increasing, etc.). This method requires fewer parameters than the non-parametric method, but is highly dependent on which function is assumed [Thorne et al. 2021]. That being said, a well selected parametric SFH function can produce results comparable to those obtained via a non-parametric model [Robotham et al. 2020].

By modelling the SFH of a galaxy, we can infer when star formation began and how much star formation has occurred during its history. This allows us to examine detailed information of a galaxy's stellar history. For example, we can estimate: 1) the lookback time at which stars began to form in a galaxy and 2) the lookback time at which the star formation was most active in a galaxy. Studying the SFHs of galaxies could give us information on the stellar masses at active star forming epochs and provide potential insight into the physical mechanisms that terminate star formation [Leitner 2012, Schreiber et al. 2018]. Since it is theorised that star formation is intrinsically linked to the availability of HI gas, examining the relationship between a galaxy's SFH and its HI properties could provide further insight into what sustains and terminates star formation within a galaxy. Investigating such a relation is one of the primary goals of this work.

1.3 The 21cm emission of HI

As discussed above, HI plays a fundamental role in the evolution of galaxies. The most direct way to observe HI is via the spin-flip transition that occurs within the ground state of the atom. The quantum spin of the electron and proton can either be parallel or anti-parallel, where the parallel state's energy is 1×10^{-6} eV greater than the energy of the anti-parallel state. When a transition of an electron changes from the parallel state to the anti-parallel state, a photon (with an energy of $E \sim 5.87\ \mu eV$) is emitted at a wavelength of 21.05 cm (corresponding to a frequency of 1420.4058 MHz), which falls within the radio wavelength of the electromagnetic spectrum. The spin-flip transition is a forbidden one, with an average transition time of ∼11 million years in one neutral hydrogen atom (corresponding to an emission coefficient of 2.88×10^{-15} s, Gould 1994). However, in gas clouds, the transition time is significantly shorter due to the collisions between atoms. These collisions make the spin-flip transition occur more frequently, therefore making it

possible to detect with radio telescopes, allowing us to study the HI content of galaxies.

Due to the expanding Universe, galaxies that are beyond the local volume (~ 4.4 Mpc) are moving away from us. This means that the emission of the spin-flip transition from these galaxies is redshifted to lower frequencies due to the expansion of space-time. By measuring the observed frequency of the HI in a galaxy, we are able to calculate its redshift via:

$$z_{obs} = \frac{\nu_{HI}}{\nu_{obs}} - 1, \tag{1.2}$$

where ν_{obs} is the observed frequency and $\nu_{HI} = 1420.4058$ MHz is the emitted frequency of the HI line. From the observed redshift, we can then calculate the luminosity distance of the galaxy (D_L):

$$D_L = \frac{c(1+z)}{H_0} \int_0^z E^{-1}(z')dz', \tag{1.3}$$

where $E(z)$ is the Hubble parameter, which gives the relationship between physical distance and the recessional velocity of a galaxy due to the expansion of the Universe. It is defined as follows:

$$E(z) = \sqrt{\Omega_R(1+z)^4 + \Omega_M(1+z)^3 + \Omega_K(1+z)^2 + \Omega_\Lambda}. \tag{1.4}$$

Using the above equations and the measured HI flux density of a given source, we can calculate its HI mass:

$$\frac{M_{HI}}{M_\odot} = 49.7 \left(\frac{D_L}{\text{Mpc}}\right)^2 \int \frac{S(\nu)}{\text{Jy}} \frac{d\nu}{\text{Hz}}, \tag{1.5}$$

where $S(\nu)$ is the flux density and ν is frequency. All equations given above are reproduced from Meyer et al. (2017).

When observing HI, you can allocate your time to either obtaining deep detections in a small sky region or covering a large field of view at shallower flux limits. Both of these observations are useful for different reasons. Shallow but wide HI surveys allow us to sample a larger variety of galaxies, making it more likely to observe rarer sources like high stellar mass galaxies. Examples of previous shallow but wide HI surveys include the HI Parkes All-Sky Survey (HIPASS; Meyer et al. 2004) which covered the entire southern sky up to $z_{HI} \sim 0.04$ and the Arecibo Legacy Fast ALFA (ALFALFA; Giovanelli et al. 2005) which surveyed ~ 7000 deg^2 of the sky up to $z_{HI} \sim 0.06$.

On the other hand, narrow but deep HI surveys allow us to observe the evolution of the HI content in galaxies and faint objects at low redshifts, which could provide insight into their evolutionary process. That being said, due to the faintness of the HI line, extremely powerful telescopes are required to observe HI beyond the local Universe, limiting previous high redshift surveys. Examples of previous narrow but deep HI surveys include the ALFA Ultra Deep Survey (AUDS; Freudling et al. 2008) which observed HI up to $z = 0.16$ in two fields with a combined area of 1.35 deg^2 and the Blind Ultra Deep HI Environmental Survey (BUDHIES; Jaffé et al. 2013) which detected HI in 150 galaxies

in and around two clusters at $0.16 < z < 0.2$.

However, modern day radio telescopes, such as the Australian SKA Pathfinder (ASKAP) and the Karl G. Jansky Very Large Array, have pushed the newer generation of high redshift HI surveys to previously unobservable depth. Examples of deep new age surveys include the COSMOS HI Large Extragalactic Survey (CHILES; Fernández et al. 2016) which probed up to $z \sim 0.5$, the ongoing FAST Ultra-DEEP Survey (FUDS; Xi, Peng, et al. 2024) which will reach $z \sim 0.4$ and the ongoing Deep Investigation of Neutral Gas Origins (DINGO; Meyer 2009) which will reach $z = 0.43$. Among the modern telescopes is the powerful and extremely sensitive Square Kilometre Array (SKA) precursor, MeerKAT, which will probe HI at even higher redshifts.

1.4 MeerKAT and the LADUMA Survey

In the Karoo region of the Northern Cape province of South Africa is the MeerKAT telescope, with a maximum baseline of 8 km consisting of 64 dishes, each with a main reflector (with a diameter of 13.5 m) and a sub-reflector (with a diameter of 3.8 m). The MeerKAT telescope is currently the most sensitive decimeter-wavelength radio interferometer in the world. Thanks to the use of Grifford-McMahon (G-M) cryogenic refrigerators, which cools parts of the receiver, and its Offset-Gregorian dish design, MeerKAT achieves excellent sensitivity and imaging quality [Jonas 2018]. MeerKAT is equipped with three receivers: the UHF-band, covering a frequency range of 580-1015 MHz, the L-band, covering 900-1670 MHz, and the S-band, covering 1750-3500 MHz. In 2009, a call for proposals for large-scale survey projects (LSPs), that would be carried out by MeerKAT, was made to the international science community. Among the 11 LSPs that were approved is the MeerKAT International GigaHertz Tiered Extragalactic Exploration (MIGHTEE; Jarvis et al. 2016) survey and the Looking At the Distant Universe with the MeerKAT Array (LADUMA; Blyth et al. 2016) survey.

The LADUMA survey is an ongoing ultra-deep survey that aims to investigate the cosmic evolution of HI in galaxies by probing HI emission up to $z_{HI} \sim 1.4$. LADUMA is a single pointing centred at the coordinates 03:32:30, −28:07:57 (J2000) in the Extended Chandra Deep Field South (ECDFS). This field was chosen for multiple reasons, including its southern location, lack of contaminating foreground objects, and its extensive multiwavelength coverage (see Blyth et al. 2016 for details). Examples of multiwavelength imaging data include the VST Optical Imaging of ECDFS and ES1 (VOICE, Vaccari et al. 2016) survey in *ugri* filters, the VISTA Deep Extragalactic Observations (VIDEO, Jarvis et al. 2012) survey in *ZYJHKs* filters and the Spitzer Extragalactic Representative Volume Survey (SERVS, Mauduit et al. 2012). LADUMA will utilise up to ~ 3424 hrs of observation time with the MeerKAT telescope, allocating ~ 300 hrs to L-band observations (which will cover $0 \leq z_{HI} \leq 0.58$) and ~ 3000 hrs to UHF-band observations (which will cover $0.42 \leq z_{HI} \leq 1.45$). Note that both the L-band and UHF-band will cover the redshift range $0.42 \leq z_{HI} \leq 0.58$, as this corresponds to the frequency range where the two bands overlap. This deep survey will have an area from ~ 0.8 deg^2 at $z = 0$ to ~ 5

Key parameters	LADUMA	MIGHTEE
Area covered in ECDFS [deg^2]	0.8-5	8
Frequency range [MHz]	580-1420	900-1420
Redshift range	$0 < z_{HI} < 1.4$	$0 < z_{HI} < 0.58$
Per channel flux sensitivity at 1420MHz [μJybeam^{-1}]	33	100

Table 1.1: Key parameters of LADUMA and MIGHTEE.

deg^2 at $z = 1.4$ with a 5σ detection limit integrated over an assumed HI linewidth of 100km/s of $7.8 \times 10^9 M_\odot$ at $z_{HI} = 0.58$ in the L-band and $3.0 \times 10^{10} M_\odot$ at $z_{HI} = 1.44$ in the UHF-band assuming the observing time of $\sim$ 3000 hrs as reported in Blyth et al. (2016).

MIGHTEE is a wide and semi-deep continuum, polarisation and spectral line survey of four well-studied extragalactic fields. The HI data compiled by MIGHTEE (dubbed as MIGHTEE-HI; Maddox et al. 2021) has the sky area covered by the MIGHTEE continuum survey (20 deg^2), which includes observations in four fields (including the ECDFS field surrounding LADUMA). This gives MIGHTEE-HI a total sky area coverage of $\sim$ 20 deg^2 that covers the redshift range $0 < z_{HI} < 0.58$.

MIGHTEE and LADUMA are complementary in area and depth, with LADUMA probing high redshifts with incredible sensitivity and MIGHTEE covering a large sky area (see Table 1.1). The complementary nature of these two surveys is illustrated in Figure 1.3, which shows the projected HI mass parameter space probed by MIGHTEE (in grey) and LADUMA (in red) [reproduced from Maddox et al. 2021]. From the figure, we see that LADUMA is projected to cover the lower HI mass parameter space at $z_{HI} > 0.03$, while MIGHTEE probes the higher HI mass ranges. For example, LADUMA is expected to detect galaxies with HI masses between $8.9\ M_\odot \lesssim \log(M_{HI}) \lesssim 9.5\ M_\odot$ at $z_{HI} = 0.1$, while MIGHTEE is projected to detect $9.5\ M_\odot \lesssim \log(M_{HI}) \lesssim 10.5\ M_\odot$ at the same redshift.

For our investigation, we utilise data from the LADUMA survey and take advantage of its complementary nature with MIGHTEE for comparison purposes. We outline the aim of this research and the structure of this dissertation in the following section.

1.5 This work

In this work, we identify low redshift ($0 < z < 0.088$) HI galaxies in the LADUMA survey and investigate the relationship between their HI content and star formation history properties. Specifically, we examine the HI properties of these low redshift galaxies against the peak of their star formation and the lookback time of said peak and use results from the MIGHTEE survey with similar research goals for comparison purposes.

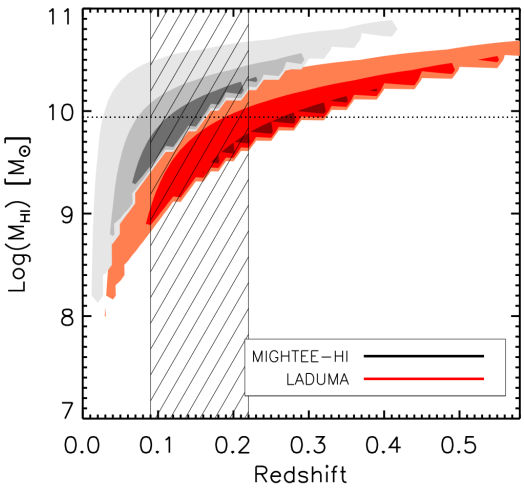


Figure 1.3: The projected HI mass-redshift parameter space, showing where LADUMA (in red) and MIGHTEE (in grey) are sensitive to galaxies, reproduced from Maddox et al. (2021). Light, medium and grey shading correspond to where at least one, 10 and 25 galaxies are expected, respectively. Light, medium and dark red shading shows where at least one, four and eight galaxies are expected, respectively. The horizontal dotted line indicates the $z = 0$ characteristic HI mass from the HI mass function. The hatched region corresponds to the redshift range $0.09 < z < 0.22$, which is heavily affected by radio frequency interference (RFI).

Since HI is thought to be the raw fuel of star formation, examining these relations may address some of the mysteries surrounding galaxy evolution, providing insight into what sustains and terminates star formation within a galaxy. To examine these relations, we begin by performing automated source finding on the high frequency LADUMA data cubes to compile our HI sample. We then utilise the photometric catalogue compiled by the Deep Extragalactic VIsible Legacy Survey (DEVILS; Davies et al. 2018) to perform SED fitting and derive the SFH of our HI sample.

This dissertation is structured as follows: Chapter 2 gives a description of the high-frequency LADUMA data cube (Section 2.1) and the photometric catalogue used in this work (Section 2.2). Chapter 3 outlines the automated source finding method used and presents the HI source finding results. Chapter 4 discuss the multiwavelength follow-up of our HI sample. This includes obtaining multiband photometry (discussed in Section 4.1) and performing spectral energy distribution (SED) fitting (discussed in Section 4.2). Chapter 5 presents the HI scaling relations produced in this work, including the M_{HI}-M_* relation (Section 5.1.1), the SFR-M_* relation (Section 5.1.2), the lookback time of peak star formation rate (τ_{peak}) versus the HI gas mass to stellar mass fraction (M_{HI}/M_*) (Section 5.2.1) and the peak star formation rate ($\mathrm{SFR_{peak}}$) versus M_{HI}/M_* (Section 5.2.2). In Chapter 6 we summarise the main results of this work and discuss a number of avenues that can be explored in future work. Throughout this work, we assume a flat ΛCDM

model with $H_0 = 67.74\ \frac{\mathrm{km/s}}{\mathrm{Mpc}}$, $\Omega_M = 0.3075$ and $\Omega_\lambda = 0.6925$. [Planck Collaboration et al. 2016].

2 | Data

2.1 LADUMA data cube

The HI data used in this work comes from the internal team release of the Early Science data cubes from the LADUMA survey, which was introduced in Section 1.4. The Early Science data includes a portion of the now completed L-band observations, which were observed over 19 night-time tracks with an average duration of 9 hours each. The data spanning the frequency range 880-1670 MHz was obtained at the native 32k channel resolution with a channel width of 26.123 kHz. In other words, the data received from the MeerKAT telescope contained 32000 frequency channels with a spacing of 26.123 kHz between each spectral channel.

To manage the data cube size and also increase sensitivity per channel, the data was downsampled during data processing to 8k resolution with a channel width of 104.52 kHz, which was then split into three independent spectral windows (SPWs) to avoid regions that are heavily affected by radio frequency interference (RFI): 1) 880-933 MHz, 2) 960-1161 MHz and 3) 1304-1420 MHz [Kazemi-Moridani 2024]. The data was processed by members of the LADUMA team using a combination of the `processMeerKAT` pipeline and custom scripts, which will be detailed in a forthcoming paper [Kazemi-Moridani et al., in prep.].

This processing included cleaning and continuum subtraction. Cleaning is the process of deconvolving the telescope's synthesised beam from the 'dirty' cube. The dirty cube contains sidelobes around sources, which are caused by incomplete sampling of the uv plane. This results in 'artificial' patterns and ripples in the data, making source finding and accurate HI mass calculations difficult. Additionally, in the frequencies where HI is observed, continuum emission can be detected, either from the HI sources themselves or other objects along the line-of-sight. This continuum emission contaminates the HI data and must therefore be removed. This is a process known as continuum subtraction. To ensure that source finding is successful, it is important to make use of cleaned and continuum-subtracted data to avoid classifying noise or continuum peaks as real line detections.

In this work, we use only the high frequency SPW, which corresponds to HI redshifts of $0 < z_{HI} < 0.088$, since, at the time, it was the only SPW with sufficient continuum subtraction and cleaning to perform successful automated source finding (see Chapter 3). Cleaning and continuum subtraction of this SPW was performed by other members of the LADUMA team and will be detailed in an upcoming paper [Kazemi-Moridani et al., in prep.]. This high frequency, Early Science data cube has the spatial dimension of 4038×4038 pixels (with a pixel scale of $2'' \times 2''$) and 1102 spectral channels with a spectral resolution of 104.52 kHz. It has a synthesised beam with dimensions of $8'' \times 7.5''$ with a position angle (PA) of 34° and an RMS noise level of 33 μJy beam^{-1} per channel [Kazemi-Moridani 2024]. These properties are summarised in Table 2.1.

Key Property	Value
Centre pointing (J2000)	03h32m30s, −28d07m57s
Frequency range	1304-1420 MHz
Redshift range	0-0.088
Pixel size	$2'' \times 2''$
Image size	4038 × 4038 pixels
Number of channels	1102
Channel width	104.52 kHz
Synthesised beam	$8'' \times 7.5''$
Position Angle of beam	34°
Per channel flux sensitivity	33 μJy beam^{-1}

Table 2.1: Key properties of the high frequency SPW early science LADUMA data cube used in this work.

2.2 DEVILS photometric catalogue

The Deep Extragalactic Visible Legacy Survey (DEVILS, Davies et al. 2018) is an ongoing spectroscopic and multiwavelength survey being performed with the Anglo-Australian Telescope (AAT). DEVILS aims to create an extensive, high completeness database of $\sim$ 60000 galaxies at $z < 1$ in the COSMOS, ECDFS and XMMLSS fields. Along with the spectroscopic survey, DEVILS aims to create an extensive database of multiwavelength photometry and derived properties, such as spectral energy distributions (SEDs), star formation histories (SFHs), stellar masses (M_*), star formation rates (SFRs) and metallicities.

As part of their survey, DEVILS has compiled a photometric catalogue [Davies et al. 2021], which consists of photometric measurements from the most up-to-date imaging available, which was measured by the two-dimensional source finder and image analysis tool, ProFound [Robotham et al. 2018]. This technique ensures a consistent aperture is used from the far-ultraviolet (FUV) to the far-infrared (FIR). This is discussed further in Section 2.2.2.

In this work, we make use of the DEVILS ECDFS photometric catalogue. The DEVILS target in the ECDFS field is centred at the coordinates 03h34m36s, −28d06m00s (J2000) with $\sim$ 6 deg^2 of coverage. The DEVILS observing field in the ECDFS does not cover the entirety of the LADUMA field (depicted in Figure 2.1). In the following section, we will briefly discuss the imaging data used to compile the DEVILS photometric catalogue. It is important to note that many of the surveys discussed below do cover more of the LADUMA field, however, many of the images were created by Davies et al. (2021) to ensure deep coverage of their observed field. This provides an opportunity for future work to compile a comprehensive photometric catalogue of the full LADUMA field, utilising

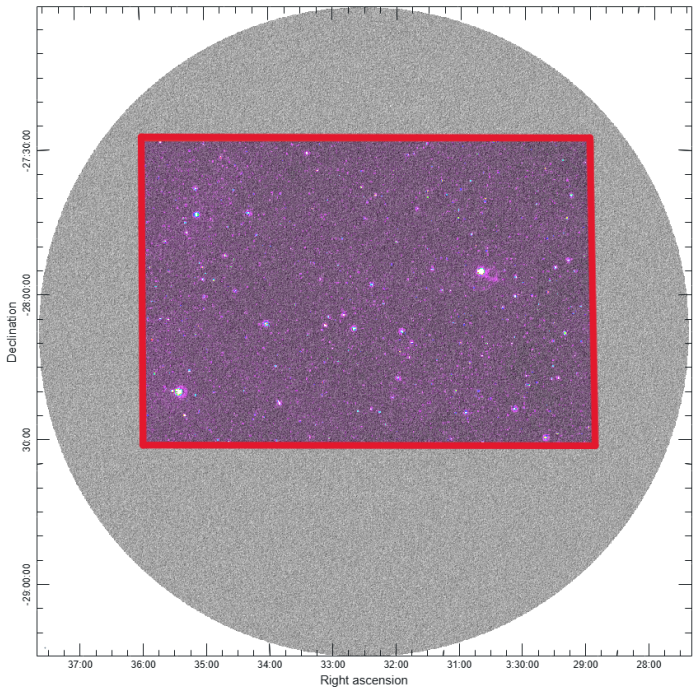


Figure 2.1: Comparison of the LADUMA observing field (in grey) and DEVILS observing field (shown by the *GALEX*-FUV image as an example in purple) of the ECDFS. The red box outlines DEVILS observing field for clarity.

the advanced techniques developed by Davies et al. (2021).

2.2.1 Imaging data sets

The majority of the information discussed in this section is summarised from Davies et al. (2021). The imaging data sets used to compile the DEVILS photometric catalogue are summarised in Table 2.2.

2.2.1.1 *GALEX* Ultraviolet

ECDFS has been observed using the Galaxy Evolution Explorer (*GALEX*) space telescope by various different survey teams as part of multiple different imaging programs. To create the imaging data for DEVILS photometry, all pointings of *GALEX* in the Barbara A. Mikulski Archive for Space Telescopes (MAST) database were overlapped to maximise the depth and coverage of the image data in the DEVILS target field. This was done using SWARP [Bertin et al. 2002], a program designed to co-add FITS images that use any arbitrary astronomical world coordinate system (WSC). SWARP was run with background subtractions turned off, which results in slightly varying depth across the image. However, this is taken into account when measuring photometry by using a local sky background estimate to determine photometric errors. The *GALEX* images used have a measured 5σ point-source depth that varies from $\approx 23.5 - 25.5$ mag across the image.

Facility	Survey	Band	Central Wavelength [Å]	Nominal Depth [5σ AB]
Ultraviolet				
GALEX	multiple	FUV	1540	$\sim$ 25.0
		NUV	2310	$\sim$ 25.0
Optical				
VST	VOICE	u	3790	25.3
		g	4740	26.4
		r	6220	26.1
		i	7760	25.2
Near-Infrared				
VISTA	VIDEO	Y	10200	24.6
		J	12600	24.5
		H	16500	24.0
		Ks	21600	23.5
Mid-Infrared				
Spitzer	SERVS	S36	35300	22.48
		S45	44700	22.07
	SWIRE	S58	56800	19.70
		S80	77500	19.96
		MIPS24	235000	18.00
		MIPS70	704000	13.26
Far-Infrared				
Herschel	HerMES	P100	989000	12.4
		P160	1.56×10^6	11.6
		S250	2.5×10^6	14.1
		S350	3.5×10^6	14.4
		S500	5.04×10^6	13.9

Table 2.2: Overview of multiwavelength data used to compile the DEVILS photometric catalogue adapted from Davies et al. 2021.

2.2.1.2 VOICE Optical

The u,g,r and i band images were observed with the 2.6m VLT Survey Telescope (VST) as part of the VST Optical Imaging of the CDFS and ES1 (VOICE, Vaccari et al. 2016) survey. The VOICE survey contains four pointings which cover 4 deg^2 of the ECDFS field. VOICE reaches a 5σ point source limiting magnitude of 25.3, 26.4, 26.1 and 25.2 mag in the u, g, r, and i bands, respectively.

2.2.1.3 VIDEO Near-Infrared

Images from the VISTA Deep Extragalactic Observations (VIDEO, Jarvis et al. 2012) survey taken by the VISTA-VIRCAM were used for the Y,J,H and Ks band images. VIDEO covers 12 deg^2 of three deep extragalactic fields, including ECDFS. The final images used to compile the DEVILS photometric catalogue have a 5σ point source limiting magnitudes of 24.6, 24.5, and 23.5 mag in the Y, J, H and Ks band, respectively.

2.2.1.4 SERVS and SWIRE Mid-Infrared

To cover the mid-infrared at shorter wavelengths, images from the *Spitzer* Extragalactic Representative Volume Survey (SERVS, Mauduit et al. 2012) were used at 3.6 and 4.5 μm (IRAC1/S36 and IRAC2/S35). SERVS is a medium-deep survey (reaching $\approx$ 2 mJy), which covers 18 deg^2 of the ECDFS field and is captured by the post-cryogenic *Spitzer* Space Telescope. Images from the first data release of SERVS were used to compile the DEVILS photometric catalogue. They reach a 5σ point source limiting magnitude of 22.48 and 22.07 in the S36 and S45 bands, respectively.

For the longer wavelengths, images from the *Spitzer* Wide-area InfraRed Extragalactic (SWIRE, Lonsdale et al. 2003) survey, which cover 5.8, 8, 24 and 70 μm (IRACS3/S58, IRACS4/S80, MIPS24 and MIPS70), were used. Here, images were compiled using data from the fifth data release of SWIRE, which has a 5σ point source limiting magnitude of 19.7, 19.96, 18 and 13.26 in the S58, S80, MIPS24, MIPS70 bands, respectively.

2.2.1.5 HerMES Far-Infrared

DEVILS FIR imaging data was compiled from the *Herschel* Multi-Tiered Extragalactic Survey (HerMES, Oliver et al. 2012a) conducted by *Herschel*'s Spectral and Photometric Imaging Receiver (SPIRE). HerMES covers PACS 100 and 160 μm (P100 and P160) and SPIRE 250, 350 and 500 μm (S100, S350 and S500). To compile the DEVILS photometric catalogue, data from the second HerMES data release, which reach 5σ point source limiting magnitudes of 12.4, 11.6, 14.1, 14.4 and 13,9 mag in the P100, P160, S250, S350 and S500 bands, respectively.

2.2.2 Photometric catalogue compilation with ProFound

In this section, we will give a brief overview of how the photometric catalogue was compiled for the DEVILS survey. This is discussed in much greater detail by Davies et al.

(2021).

Davies et al. (2021) made use of ProFound [Robotham et al. 2018], a two-dimensional source finder and photometric tool, to compile their photometric catalogue. ProFound introduces a new key feature where all photometry is executed on a dilated segmentation map, rather than a traditional elliptical aperture. The use of a segmentation map provides more accurate measurements of photometry, as the entire flux of a galaxy, with minimum inclusion of background sky emission, can be encapsulated by a segment.

ProFound uses the following iterative process to create its segmentation map (summarised from Robotham et al. 2018):

1. The brightest pixel in the given image that has not been assigned to a segment is identified and a new segment is created to encapsulate it.
2. The surrounding pixels that have not yet been assigned to their own segment are then examined by ProFound, which assigns the search pixel as follows:
   (a) If pixels surrounding it have already been assigned to the current growing segment and the search pixel has a lower flux than those pixels, it is also added to the growing segment.
   (b) If a search pixel has a higher flux than its surrounding pixels that are all assigned to the growing segment, the pixel is not assigned to the segment (as this could be flux from a separate, brighter source).
   (c) If no more pixels can be assigned to the current segment, then the process is terminated.
3. The next brightest pixel is then selected and the above process is repeated until all pixels above the specified brightness level have been assigned to a segment, after which the process is terminated.

ProFound produces two segmentation maps using the above method. The first is a colour-optimised segmentation map which covers the central, high signal-to-noise ratio (SNR) region of the source. This serves as a tight detection segmentation map and does not encompass all of a galaxy's flux. The second segmentation map that ProFound produces is known as the total-optimised segmentation map. This map dilates the colour-optimised segmentation map in an attempt to extract the total flux of the sources. Measurements from both of these segmentation maps are included in the final DEVILS photometric catalogue. However, for this work, we only make use of the measurements from the total-optimised segmentation map, as it is best for work that requires complete flux measurements, which is required for this work. An example of a total-optimised segmentation map is shown in Figure 2.2. To produce this, we executed ProFound in a similar manner to that of Davies et al. (2021) on a cutout of a galaxy in the LADUMA field from the VIDEO Y-band image.

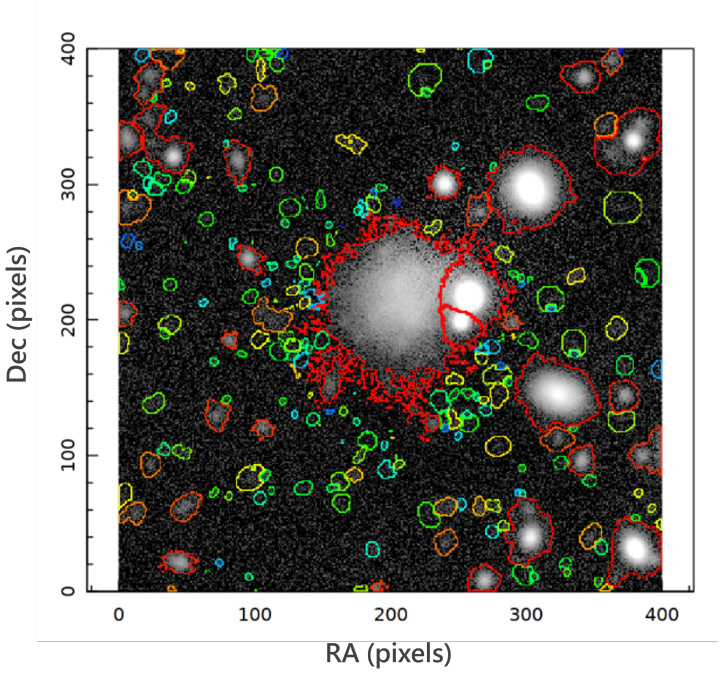


Figure 2.2: Example of a total-optimised segmentation map produced by ProFound. The background image is a cutout of a galaxy in the LADUMA field from VIDEO Y-band image. The colours of the segmentations indicate the relative brightness of the sources identified in the image, with red being the brightest, through to blue being the faintest.

The UV to MIR (UV-MIR from *GALEX*-FUV to *Spitzer*-S80) images have comparable pixel scales and, due to their higher resolution, have minimum confusion between bands. This is not the case for the MIR to FIR bands (MIR-FIR from *Spitzer*-MIPS24 to *Herschel*-S500), which have lower resolutions and varying point spread functions (PSFs). For this reason, different techniques were implemented to measure photometry in these two wavelength regimes.

To extract the UV-MIR photometry, ProFound was first implemented as a source finder to identify all galaxies in the DEVILS field. Source finding was performed on an inverse variance weighted stacked image of the VIDEO Y, J and H band images (to ensure source finding was performed on an image of maximum depth) which has a 0.1 deg^2 buffer around the DEVILS targeted field (to account for sources that may exist on the edge of the field). This process is performed internally by ProFound with the `profoundMultiBand` command. To mitigate the fragmentation of bright, extended sources into multiple different segments, the `reltol` and `cliptol` parameters were implemented, along with the `profoundSegimFix` function. The `reltol` and `cliptol` parameters merge segmentations with a given flux criteria, while the `profoundSegimFix` function allows manual merging and grouping of segments (see Davies et al. 2021 for details). The UV-MIR photometry is then extracted within the bounds of the total-optimised segmentation map (discussed above), which was produced during source finding. The segmentation is allowed to dilate in each band to account for the slight variation of the PSFs from band to band.

In the case of MIR-FIR photometry, it is not appropriate to use this same segmentation map due to the lower resolution and shallower depth of the band images in this wavelength regime. Instead, the `profoundFitMagPSF` function is used, which is designed specifically to measure photometry in low-resolution data. This function fits the PSF of each MIR-FIR band at each source position, which is provided by the UV-MIR catalogue. This ensures that all flux is accounted for in each source in each band. Along with this, `profoundFitMagPSF` identifies any additional sources that do not align with the UV-MIR catalogue, to ensure the inclusion of bright FIR sources that do not emit at other wavelengths. That being said, due to the poor sensitivity and resolution of the MIR-FIR band images, it is inappropriate to include faint NIR sources in the longer wavelength photometric catalogue, as these faint sources are heavily affected by confusion. To avoid this, Davies et al. (2021) do not extract the MIR-FIR photometry of galaxies with Y band magnitude > 21.2 mag.

The `profoundFitMagPSF` function requires an empirical PSF to be defined for each input band. In the PACS bands (P100 and P160), the PSF input is the observed PSF kernel sourced from the PACS Evolutionary Probe (PEP[1]) first data release. For the SPIRE bands, a Gaussian PSF is used with a full-width-half-maximum (FWHM) of 18.15, 25.15 and 36.3 arcsec for the S250, S350 and S500 bands, respectively (as given in Oliver et al. 2012b). For the MIPS bands, the 100K PSFs from Gordon et al. (2008) are used after they are rescaled to the image pixel size of their respective bands.

The ProFound output catalogue includes segmentation statistics (such as image coordinates, radii, number of pixels etc.), colour flux, total flux and their associated error. ProFound takes into account the error in the absolute local sky and noise from the background root mean square (RMS) that exist within the segment, which are added in quadrature to calculate the final flux uncertainty of each source.

2.3 Chapter Summary

The HI data used in this work comes from the internal team release of the Early science data cubes from the LADUMA survey. This work only utilises the high frequency SPW (which corresponds to $0 < z_{HI} < 0.088$), since this is currently the only SPW with sufficient continuum subtraction and cleaning to perform successful source finding. This cube has a synthesised beam with dimensions of $8'' \times 7.5''$ with a PA of $34°$ and a spectral resolution of 104.25kHz. These, along with other key properties of this data cube, are summarised in 2.1.

For SED fitting purposes, we require multiwavelength photometry. This was obtained via the DEVILS photometric catalogue, which consists of photometric measurements from

[1] https://www.mpe.mpg.de/resources/PEP/DR1_tarballs/readme_PEP_global.pdf

the most up-to-date imaging available from FUV to FIR (summarised in Table 2.2). To compile the catalogue, the DEVILS team utilises ProFound, a 2D source finder and image analysis tool, to extract photometry. ProFound executes all photometric measurements on a dilated segmentation map. This technique provides more accurate measurements of photometry and ensures a consistent aperture is used through all wavebands. It is important to note that the DEVILS observation field of ECDFS does not cover the entirety of the LADUMA field. This provides an opportunity for future work to compile a comprehensive photometric catalogue of the field, utilising the advanced techniques developed by the DEVILS team.

3 | HI Source Finding

Finding genuine HI emission in the data cube is the essential first step in our investigation. This process is referred to as source finding. Multiple different methods of source finding exist, including manual, semi-automated and automated methods, each of which has benefits and drawbacks.

Manual source finding involves visually inspecting the data cube and manually constructing the source mask. It has hitherto served as the main method of source finding in survey data. Manual source finding can be a powerful technique, as the human eye is adept at pattern recognition and would likely spot most of the bright emission from large galaxies in the data cube. However, this may not be the case with diffuse emission from smaller galaxies, as their emission would exist close to the noise level of the data cube. On the other hand, automated source finders make use of iterative spatial and spectral smoothing, making it easier to find diffuse emission. Along with this, they are time-efficient, which is not the case with manual source finding. For these reasons, the implementation of source finding algorithms is imperative.

Source finding algorithms can either be semi-automated, where human intervention is required to verify the robustness of the final catalogue, or automated, where minimal human intervention is required. Both manual and semi-automated methods of source finding have already been implemented on the high-frequency (1304-1420 MHz) Early Science LADUMA data cube (hereafter referred to as the LADUMA ES cube for brevity). In this chapter, we present our implementation of automated source finding on the LADUMA ES cube and assess its performance relative to the manual and semi-automated methods. To the best of our knowledge, no HI sources have been previously catalogued in the LADUMA field. Therefore, any sources identified in the LADUMA data will be presented here for the first time[1].

To perform source finding, we make use of the second iteration of the automated Source Finding Algorithm (SoFiA-2; Westmeier et al. 2021 in this work. SoFiA-2 is written in the C-programming language, resulting in faster processing times compared to its predecessor, SoFiA [Serra et al. 2015], which was mainly based in Python. SoFiA-2 is a flexible source finder. However, this also means that to produce results comparable to those produced via other source finding methods, it is vital to optimise the input parameters to the data cube source finding is being performed on. This is one of the primary goals of this work. In this chapter, Section 3.1 details our implementation of SoFiA-2. Section 3.2 assesses the performance of the automated source finding compared to manual and semi-automated methods. Section 3.3 discusses the HI properties of the final SoFiA-2 detected sample.

[1] and in Kazemi-Moridani 2024

3.1 Implementation of SoFiA-2

We implement SoFiA-2 on the non-primary beam (non-pb) corrected, cleaned, and continuum subtracted LADUMA ES cube[2]. Due to the natural sensitivity fall-off, non-pb corrected data cubes have decreasing signal strengths as you move from the beam centre. However, the noise per spectral channel is constant across the primary beam in the non-pb corrected cube. Applying a primary beam (pb) correction compensates for the sensitivity fall-off, recovering the true flux density of sources throughout the beam area. As a result, pb corrected data cubes have decreasing signal strengths away from the beam centre. However, this correction increases the noise at the edges of the cube. Therefore, initial source finding is performed on the non-pb corrected cube, since the noise across the cube is more uniform. Subsequently, targeted source finding using the source mask produced from the non-pb corrected cube is conducted on the pb-corrected cube to obtain accurate flux measurements.

As discussed above, we utilise SoFiA-2 to perform blind HI source finding. The key steps in the SoFiA-2 algorithm are illustrated in Figure 3.1. A data cube and parameter file are two of the inputs accepted by SoFiA-2 (shown in blue in Figure 3.1). The parameter file lists all the user-input parameters (shown in yellow in Figure 3.1). The full SoFiA-2 parameter file used in this work is given in Appendix A.1.

To optimise the parameters of SoFiA-2 to the LADUMA ES cube, we ran SoFiA-2 iteratively. This consisted of performing an initial run of SoFiA-2, comparing the output catalogue to those produced via alternative methods of source finding (manual and semi-automated), determining how many of the SoFiA-2 detections were found by these alternative methods and how many were likely false detections (see Section 3.2), then adjusting the SoFiA-2 parameters to improve the results of the catalogue produced by the following run. This was repeated until a robust final catalogue was produced and adjusting input parameters further yielded no overall improvement. These parameters are summarised in Table 3.1.

For further analysis of the optimal SoFiA-2 parameters found in this work, we compare them to those implemented in Westmeier et al. (2022) and Rajohnson et al. (2024). The goal of Westmeier et al. (2022) was to create an HI catalogue of the Hydra cluster at $z_{HI} \lesssim 0.08$. The Hydra cluster was observed by the Wide-field ASKAP L-band Legacy All-sky Blind Survey (WALLABY; Duffy et al. 2012), a blind HI LSP run on the Australian Square Kilometre Array Pathfinder (ASKAP) telescope. Rajohnson et al. (2024) aimed to search for HI emission in the Vela supercluster at $z_{HI} \lesssim 0.08$, which was observed by the SARAO MeerKAT Galactic Plane Survey (SMGPS; Bordiu et al. 2025), a continuum survey run on the MeerKAT telescope. Both Westmeier et al. (2022) and Rajohnson et al. (2024) utilised SoFiA-2 to perform blind HI source finding and catalogue

[2]produced by other LADUMA team members. This will be discussed in detail in a forthcoming paper [Baker et al. in prep]

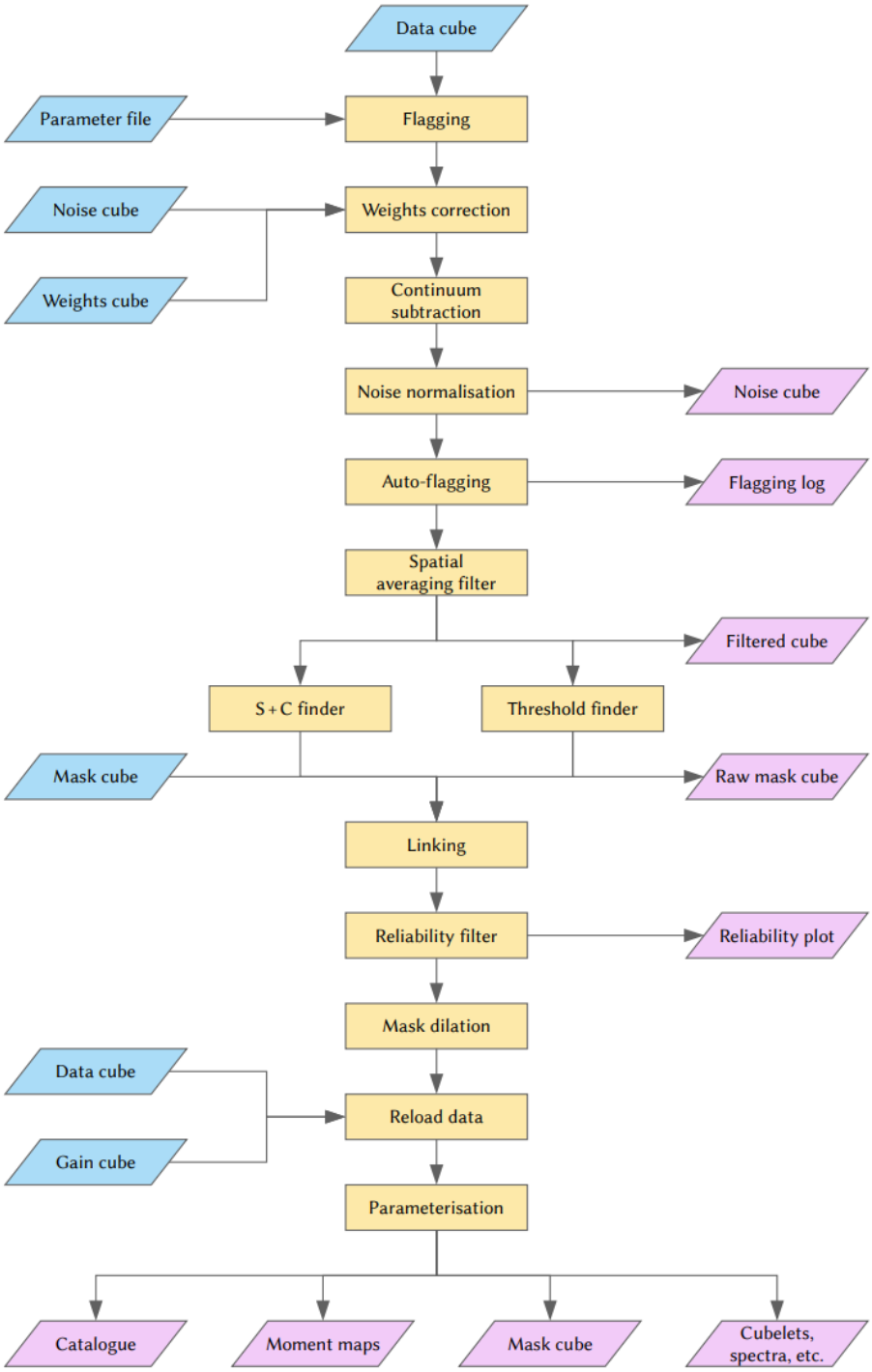


Figure 3.1: Flow diagram illustrating the work flow in which SoFiA-2 performs source finding, reproduced from the SoFiA-2 User Manual [T. Westmeier]. The set of parameters which control the algorithm is shown in yellow, user-input data products are shown in the blue and SoFiA-2 output data products are shown in the purple.

SoFiA Parameter Name	Set value in this work	Corresponding physical value
The Flagging Parameters		
`flag.region`	3533, 3578, 974, 1016, 0, 1102	52.17°, 52.14°, −28.71°, −28.69°, 1304 MHz, 1420 MHz
The S+C Parameters		
`scfind.kernelsXY`	0, 3, 5	0″, 6″, 10″
`scfind.kernelsZ`	0, 3, 5, 7	0, 313.56 kHz, 522.6 kHz, 731.64 kHz
`scfind.threshold`	4.0	132 μJybeam^{-1} per channel
The Linker Parameters		
`linker.radiusXY`	3	6″
`linker.radiusZ`	2	209 kHz
`linker.minSizeXY`	5	10″
`linker.minSizeZ`	5	522.6 kHz
The Reliability Parameters		
`reliability.threshold`	0.5	-
`reliability.scaleKernel`	0.25	-
`reliability.minSNR`	12.0	3

Table 3.1: Optimal input values of various SoFiA-2 parameters implemented to perform source finding over the high frequency (1302-1420 MHz) LADUMA ES cube.

the HI emission of their respective clusters.

Westmeier et al. (2022) and Rajohnson et al. (2024) were chosen for comparison purposes because both utilised the same version of SoFiA-2 used in this work to perform blind HI source finding. Additionally, both utilised data that observed similar redshift ranges as that examined in this work ($0 < z_{HI} < 0.08$). This comparison also allows us to evaluate how SoFiA-2 is implemented on a different telescope (in the case of Westmeier et al. (2022)) and on a different survey from the same telescope (in the case of Rajohnson et al. (2024)). This comparison is detailed in Section 3.1.5.

In Sections 3.1.1, 3.1.2, 3.1.3, and 3.1.4, we discuss the flagging parameters, S+C finder parameters, linker parameters, and reliability parameters, respectively. These parameters were found to have the most influence on the source finding performance. All parameters not discussed in this section were left to the default SoFiA-2 setting. Full details of SoFiA-2 are presented in Westmeier et al. (2021) and the user manual [T. Westmeier[3]] with relevant aspects summarised below.

3.1.1 Flagging Parameters

Before beginning source finding, we utilise one of the flagging parameters built into SoFiA-2 to flag the bright continuum source (PKS 0326-288, with a total continuum flux of 1.04 Jy at 408 MHz; Kapahi et al. 1998), shown in Figure 3.2, which exists in the LADUMA

[3] https://gitlab.com/SoFiA-Admin/SoFiA-2/-/wikis/documents/SoFiA-2_User_Manual.pdf

ES cube. Not flagging PKS 0326-288 would cause SoFiA-2 to label it (and the artefacts caused by it) as a detection, compromising the reliability of the SoFiA-2 output catalogue. To do this, we input the region around PKS 0326-288 with the `flag.region` parameter. This region is then ignored during the source finding process. The spatial and spectral coordinates of PKS 0326-288 in the LADUMA ES cube are $x_{min} = 3533$, $x_{max} = 3578$, $y_{min} = 974$, $y_{max} = 1016$, $z_{min} = 0$, $z_{max} = 1102$, where $x_{min}, x_{max}, y_{min}, y_{max}$ are given in pixel coordinates (represented by the blue box in Figure 3.2) and z_{min}, z_{max} are given in channel number. These values correspond to the following RA and Dec coordinates: $x_{min} = 52.17°$, $x_{max} = 52.14°$, $y_{min} = -28.71°$, $y_{max} = -28.69°$, $z_{min} = 1304$ MHz, $z_{max} = 1420$ MHz.

3.1.2 S+C Finder Parameters

SoFiA-2 offers a choice of two source finding algorithms, the smooth and clip (S+C) finder and the threshold finder. The threshold finder algorithm is the less advanced of the two and simply adds pixels to the source mask if they meet a set absolute flux density threshold. In this work, we opted to rather use the S+C finder, as it is the most sophisticated source finding algorithm of the two [Westmeier et al. 2021]. The S+C Finder works by smoothing the data in several iterations spatially and spectrally. Between each iteration, pixels with an absolute value above a given root mean squared (RMS) threshold are identified and added to a source mask. This source mask is a spectral cube with the same dimensions as the data cube, with a given pixel value at the coordinates where pixels in said data cube meet the RMS threshold and 0 otherwise. Pixels above the RMS threshold are then excluded in subsequent iterations so as not to double-count their flux contributions.

Smoothing is performed by a Gaussian filter in the spatial domain and a Boxcar filter in the spectral domain. The Full Width at Half Maximum (FWHM) of the Gaussian filter used in each smoothing iteration is governed by the `scfind.kernelsXY` parameter while the width of the Boxcar filter is governed by the `scfind.kernelsZ` parameter. For this work, the `scfind.kernelsXY` parameter was set to 0,3,5 pixels, and the `scfind.kernelsZ` was set to 0,3,5,7 channels. This corresponds to the physical scales 0, 6″, 10″ in the spatial domain and 0, 313.56, 522.6, 731.64 kHz in the spectral domain. `scfind.kernelsZ` and `scfind.kernelsZ` were set to these values as these are the expected scales of the sources in the LADUMA ES cube. Setting these parameters in this manner ensures the signal-to-noise ratio (SNR) of said sources is maximised during the smoothing phase, allowing fainter sources to be detected.

The RMS threshold is governed by the `scfind.threshold` parameter, which specifies the flux threshold relative to the global RMS noise level (σ_{RMS}) of the data cube. In this work, `scfind.threshold` was set to $4\sigma_{RMS}$, which corresponds to the physical value of 132 μJybeam^{-1} per channel. All smoothed detections with an absolute flux density greater than the set `scfind.threshold` value is marked as detections and added to the reliability plot. The reliability plot (shown in Figure 3.3) is a diagnostic plot outputted by SoFiA-2. It shows the 2D projection of the distribution of positive (in blue) and negative

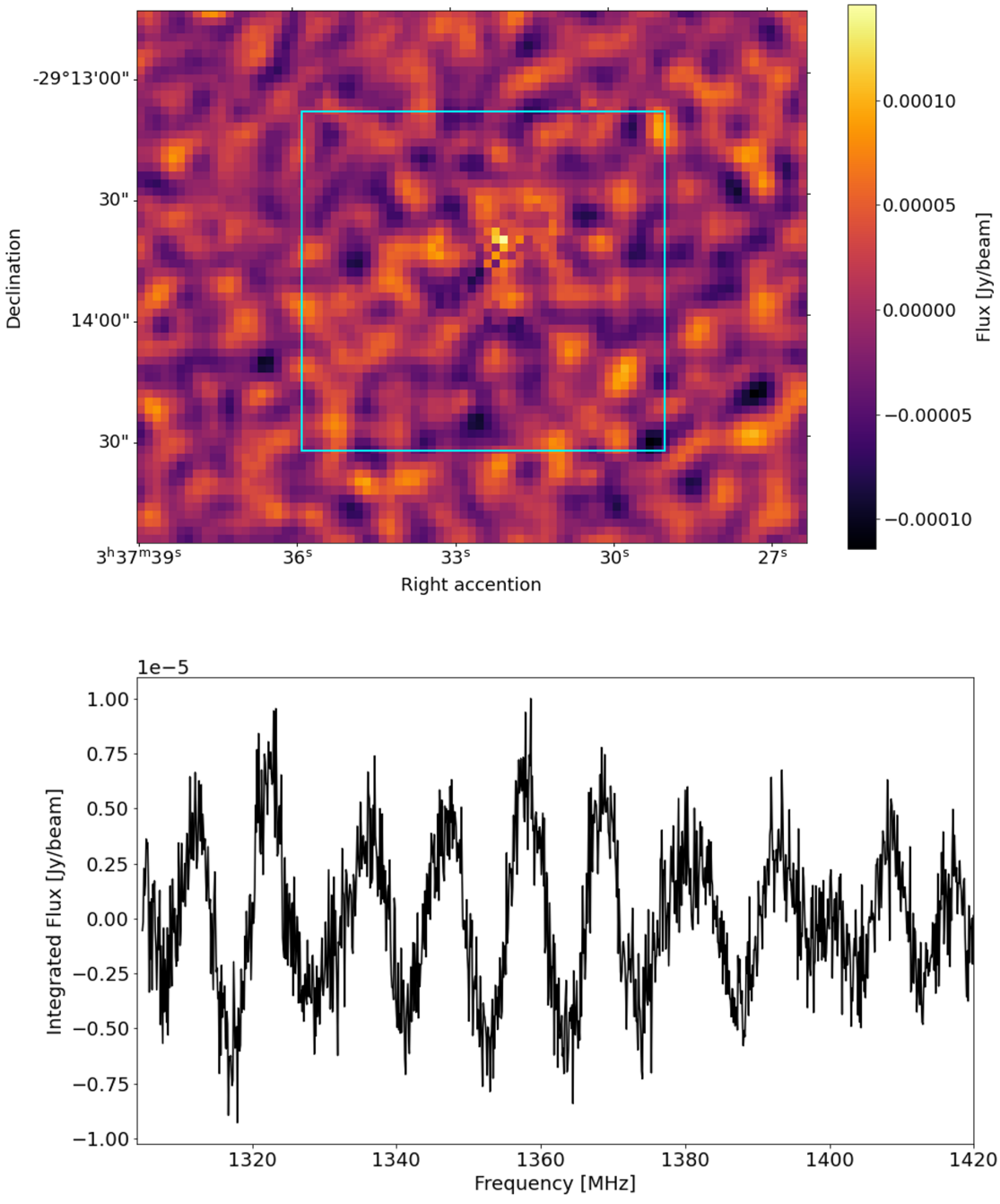


Figure 3.2: Continuum source PKS 0326-288 shown in the single 1304MHz frequency channel of the LADUMA ES cube (top panel) and its spectrum (bottom panel), showing the integrated flux within the blue box over all frequency channels. The physical coordinates of the box are: $x_{min} = 52.17°$, $x_{max} = 52.14°$, $y_{min} = -28.71°$, $y_{max} = -28.69°$, $z_{min} = 1304$ MHz, $z_{max} = 1420$ MHz.

(in red) pixels above the set threshold in 3D parameter space, which is defined by the log of the mean SNR vs integrated SNR vs peak SNR of each detection.

We found that the performance of the source finding algorithm was heavily affected by the input value of `scfind.threshold`. When this parameter was set to greater than $4\sigma_{RMS}$, we found that faint sources that exist in the data cube were excluded from the source mask, as their RMS value is below the set threshold. This results in an underpopulated reliability plot, as shown in Figure 3.4, where `scfind.threshold` is set to 5. In other words, fewer pixels are being marked as detection, resulting in fewer blue and red points in the reliability plot. When `scfind.threshold` was set to 5, we found that ~ 35 sources, found in the optimised SoFiA-2 run, were now undetected. When `scfind.threshold` was set to less than $4\sigma_{RMS}$, the reliability plot is overpopulated, as shown in Figure 3.5, where `scfind.threshold` is set to 2.5. In other words, more pixels are being labelled as detections, resulting in more blue and red points being included in the reliability plot. By inspecting the detected sources' spectral profile and optical counterpart (see Section 3.2 for further details), we found that setting `scfind.threshold`$< 4\sigma_{RMS}$, results in ~ 15 more noise peaks (i.e. false detections) being included in the produced SoFiA-2 catalogue, in comparison to the optimised SoFiA-2 run. This occurs because the threshold is set below the RMS of the stochastic noise in the data cube.

3.1.3 Linker Parameters

With a source mask defined, the next step is to group the pixels into individual sources. This process begins by implementing a friends-of-friends algorithm, which runs as follows:

1. The algorithm loops over the mask cube until it detects a pixel that has been marked as a detection and has not been assigned to an identification (ID) number.
2. This pixel is assigned a new source ID number.
3. The neighbouring pixels are examined. If these neighbours are within a given merging length, they are assigned the same ID number.
4. The process is repeated until all pixels that are considered a detection have been assigned ID numbers.

The merging lengths in the spatial and spectral domain are governed by the `linker.radiusXY` and `linker.radiusZ` parameters, respectively. These were set to `linker.radiusXY`=3 and `linker.radiusZ`=2 in this work, corresponding to the physical values of $6''$ in the spatial domain and 209.04 kHz in the spectral domain. In other words, if two pixels are separated by $6''$ spatially and 209.04 kHz spectrally, they are assigned the same ID number. The default setting for both these parameters is 1. Increasing these parameters may result in SoFiA-2 combining two separate sources into one source. On the other hand, setting this parameter lower may result in SoFiA-2 separating one source into two sources. After inspecting the optical counterparts of the sources detected by SoFiA-2 (see Section 3.2 for details), we found the values set in this work to be optimal, ensuring that smaller sources are not combined and larger sources are not split in the final source

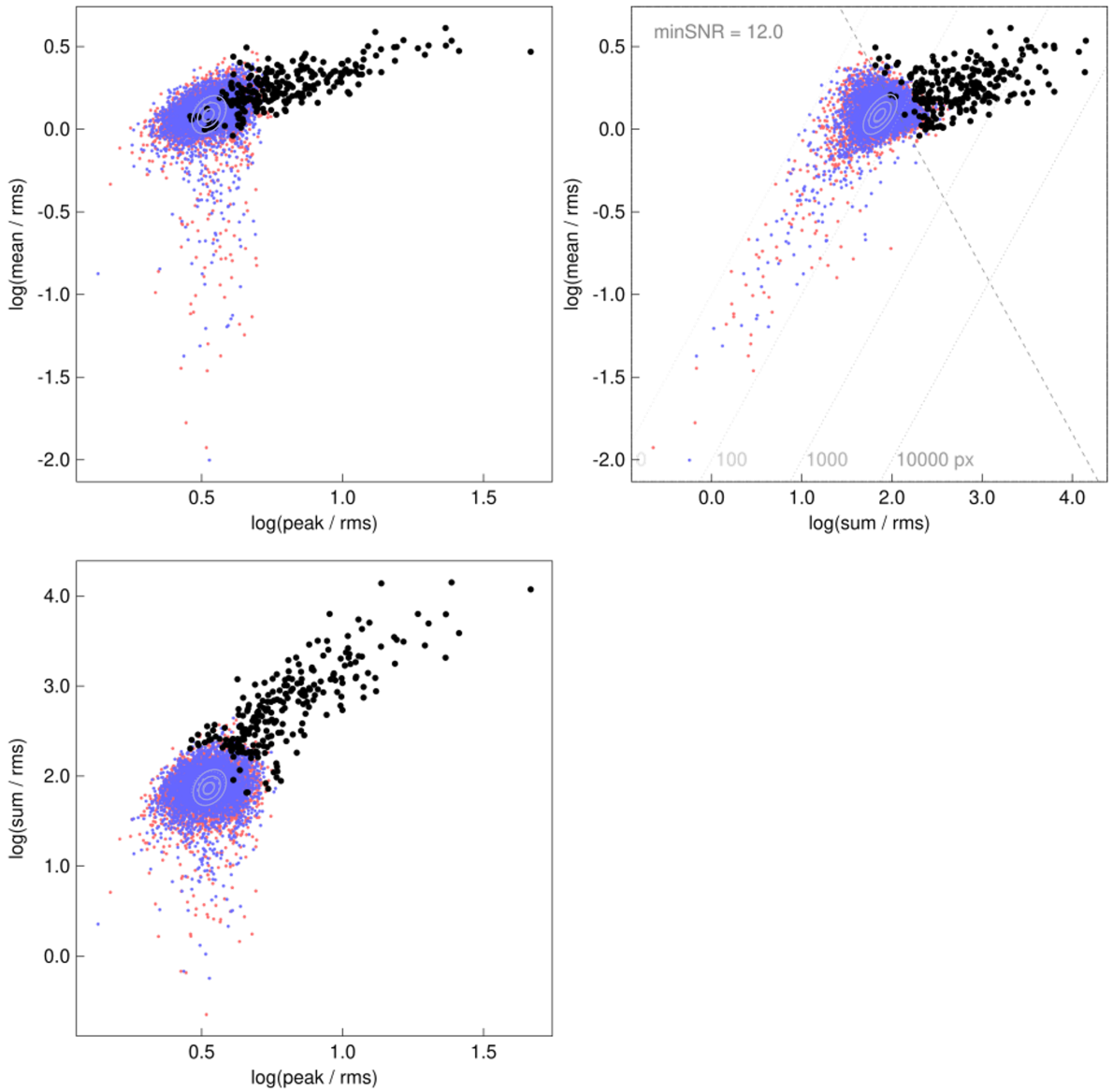


Figure 3.3: Reliability plot of the final SoFiA-2 run performed on the 1304-1420MHz SPW, cleaned LADUMA ES cube, with `scfind.threshold` set to 4 and `reliability.minSNR` set to 12. This shows the logarithmical 3D parameter space used by SoFiA-2. This consists of the mean SNR vs peak SNR (top left panel), mean SNR vs integrated SNR (top right panel) and integrated SNR vs peak SNR (bottom left panel). The blue points represent the positive pixels and the red points negative pixels that meet the RMS threshold set by the user. The black points are final detections which were added to the final catalogue outputted by SoFiA-2. The white contour ellipses (the thickness of which have been enhanced for visual purposes) represent 1σ, 2σ and 3σ of the 3D Gaussian function used in the kernel density estimation. The light-grey dotted diagonal lines in the top left panel depict the number of pixels that exceed the set threshold. From left-to-right, N=100 pixels, N= 1000 pixels and N=10,000 pixels. The grey dashed line represents the minimum SNR threshold, which is set by the `reliability.minSNR` parameter.

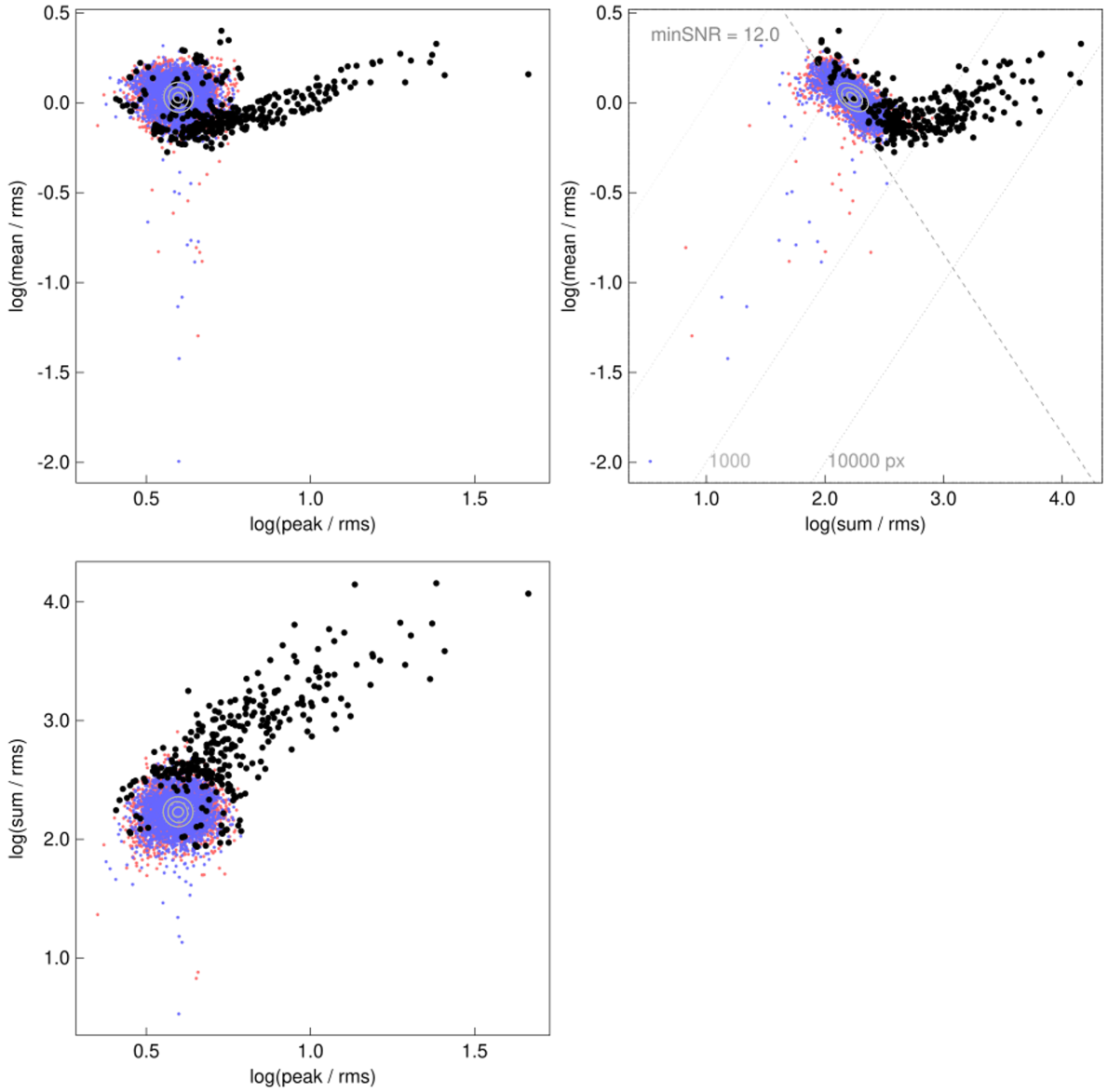


Figure 3.4: As in Figure 3.3 but with `scfind.threshold` set to 5.

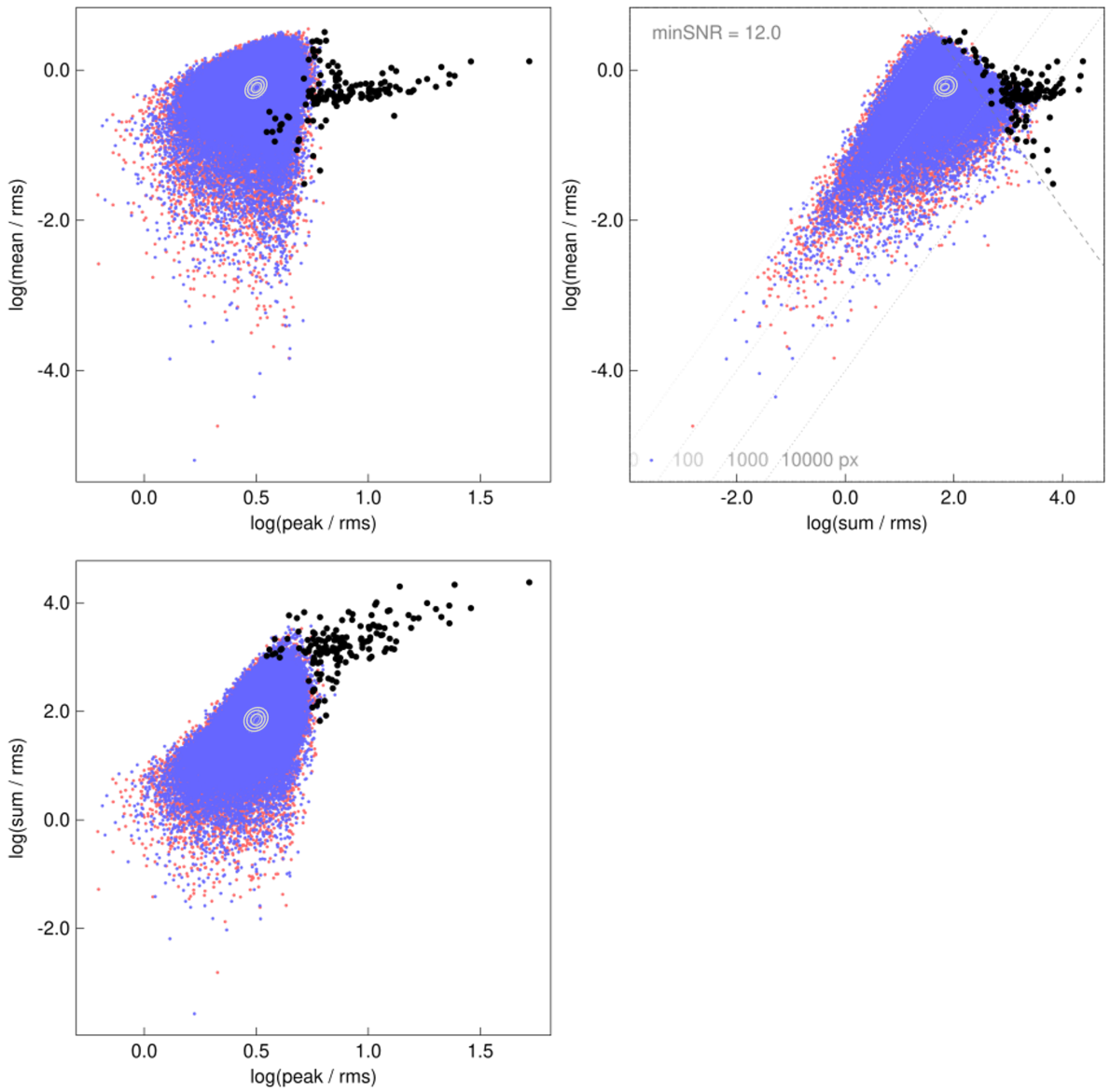


Figure 3.5: As in Figure 3.3 but with `scfind.threshold` set to 2.5.

catalogue.

Now that all the positive pixels have been assigned an ID number, they are grouped together by the `linker.minSizeXY` and `linker.minSizeZ` parameters. This dictates the minimum size requirement a detection needs to be to become a candidate source. In this work, both the `linker.minSizeXY` and `linker.minSizeZ` were set to 5. In other words, we required that there be at least 5 detected pixels next to each other, in either the spatial or spectral domain, for it to be considered a source. This corresponds to physical scales of at least 10″ wide in one of the spatial domains or 522.6 kHz wide in the spectral domain. The beam size of the LADUMA ES cube is $8'' \times 7.5''$ [Kazemi-Moridani 2024], meaning that a source needs to be at least 2″ larger (i.e. 25% larger) than the beam for it to be considered a source. We note that these parameters were left at the default setting, as they were found to be the optimal value when performing source finding on the LADUMA ES cube. We found when `linker.minSizeXY` and `linker.minSizeZ` were set to less than 5, fewer real sources were detected by SoFiA-2, as adjusting the parameters in this way causes SoFiA-2 to hit the resolution limit, where the level of detail required of the source mask can no longer be captured. When `linker.minSizeXY` and `linker.minSizeZ` were set to greater than 5, we found that small, faint sources were excluded from the list of candidate detections.

3.1.4 Reliability Parameters

With a list of candidate sources defined by the linker parameters, the next step is to evaluate which of these candidates are reliable enough to be included in the final output source catalogue. This is governed by the reliability parameters. The `reliability.minSNR` parameter sets the SNR threshold which candidates must exceed to be considered reliable. The SNR of candidate sources are calculated by:

$$\mathrm{SNR} = \frac{\Sigma \mathrm{S_i}}{\sigma_{\mathrm{rms}} \sqrt{\mathrm{N_{pix}} \Omega_{\mathrm{PSF}}}}, \tag{3.1}$$

where ΣS_i is the sum of flux densities across the source mask, σ_{rms} is the global RMS noise level of the data cube, N_{pix} is the total number of spatial and spectral pixels which form the detection and Ω_{PSF} is the beam solid angle (in pixels) [equation reproduced from the SoFiA-2 manual].

The `reliability.minSNR` parameter was set to 12 in this work, and is represented by the grey dashed line in Figure 3.3. In other words, candidates with an SNR$<$ 12 were considered unreliable and therefore removed from the final source mask. It is important to note that the `reliability.minSNR` is in pixel scale in this case since the LADUMA ES cube did not contain beam information in the header. Substituting Ω_{PSF} (as shown in Equation 3.1) with the beam size in pixel scale (~ 4 pixels), makes the true set value of `reliability.minSNR` ~ 3.

After examining the spectra and optical counterparts of the sources included in the final

catalogue, we found that setting `reliability.minSNR`< 12 resulted in a final catalogue with at least 10 more false detections than when `reliability.minSNR` is set to 12. This implies that more pixels which belong to noise peaks are being marked as detections in the reliability plot. This can be seen in Figure 3.6, where `reliability.minSNR` was set to 8, which has more black points than Figure 3.3. Conversely, when `reliability.minSNR` was set to greater than 12, it resulted in at least 10 real detections being excluded from the final catalogue when compared to the run where `reliability.minSNR` was set to 12. This implies that faint sources were labelled as unreliability due to their SNR being lower than the set threshold, resulting in them being excluded from the final source mask. This can be seen in Figure 3.7, where `reliability.minSNR` was set to 16, resulting in fewer black points in comparison to Figure 3.3.

It is expected that positive detections (established during the S+C finding phase) populate a different part of the parameter space than false detections caused by noise peaks. Therefore, the reliability (R_{sofia}) of a candidate source can be estimated by evaluating the number of positive versus negative detections in different regions of the parameter space. This is the approach implemented by the SoFiA-2 algorithm, with the reliability of each candidate source calculated by:

$$R_{\mathrm{sofia}} = \begin{cases} \frac{n_{pos}-n_{neg}}{n_{pos}} & \text{if } n_{pos} \geq n_{neg} \\ 0 & \text{otherwise} \end{cases}, \tag{3.2}$$

where n_{pos} and n_{neg} are the number of positive and negative detections that exist in that region, respectively [equation reproduced from the SoFiA-2 manual].

The number of positive and negative detections (required to calculate R_{sofia}) is measured via a Gaussian kernel density estimation (KDE) centred on the position of each candidate detection. The relative size of the 3-dimensional Gaussian kernel used in the KDE is controlled by the `reliability.scaleKernel` parameter and is represented by the white ellipses in Figure 3.3. To determine the optimal value of this parameter, we examined the cumulative Skellam distribution plot output by SoFiA-2 for diagnostics. An example of this is shown in Figure 3.8. This figure shows the cumulative Skellam distribution (K), evaluated at the positions of all negative detections, normalised to 1. We expect the shape of the normalised cumulative Skellam distribution at the positions of the negative detections to be roughly Gaussian, since this is the shape generated by stochastic noise in the data cube. K models the difference between two independent variables, which in this case is the number of positive and negative detections, and is defined as [reproduced by the SoFiA-2 manual]:

$$K = \frac{n_{pos} - n_{neg}}{\sqrt{n_{pos} + n_{neg}}}. \tag{3.3}$$

In this work, `reliability.scaleKernel` was set to 0.25. We found that setting the `reliability.scaleKernel` parameter to greater than 0.25 resulted in real sources be-

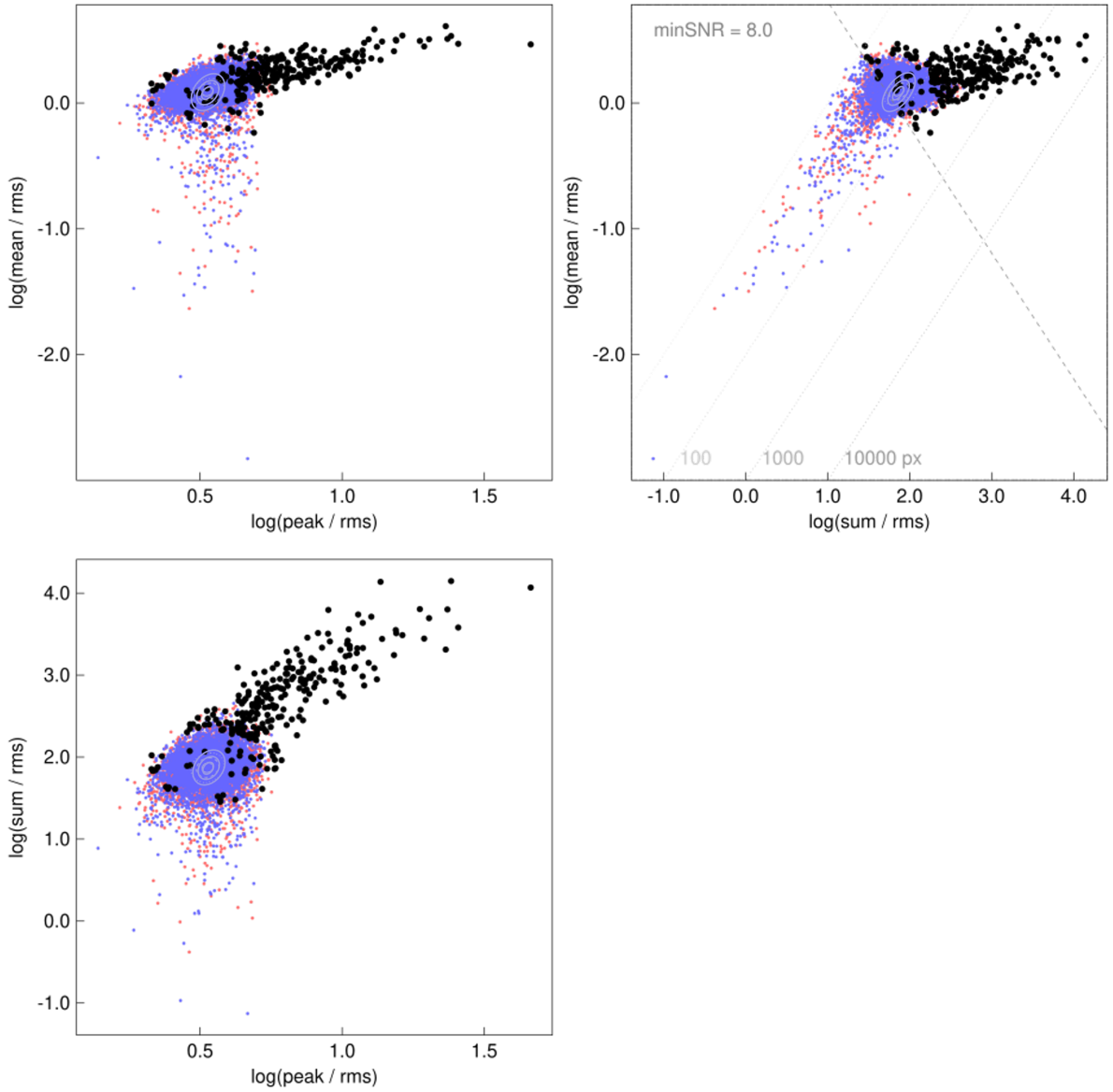


Figure 3.6: As in Figure 3.3 but with `reliability.minSNR` set to 8.

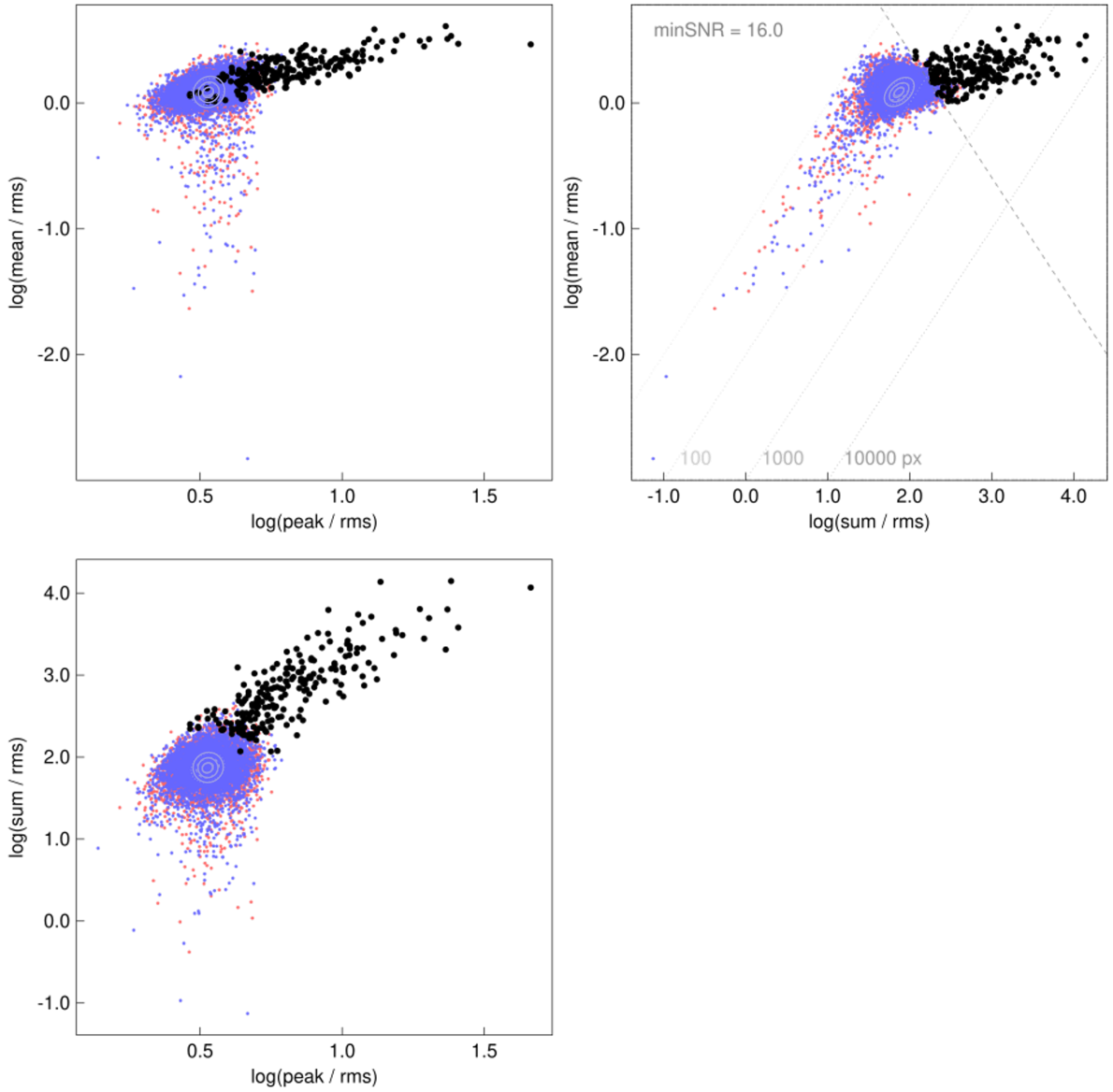


Figure 3.7: As in Figure 3.3 but with with `reliability.minSNR` set to 16.

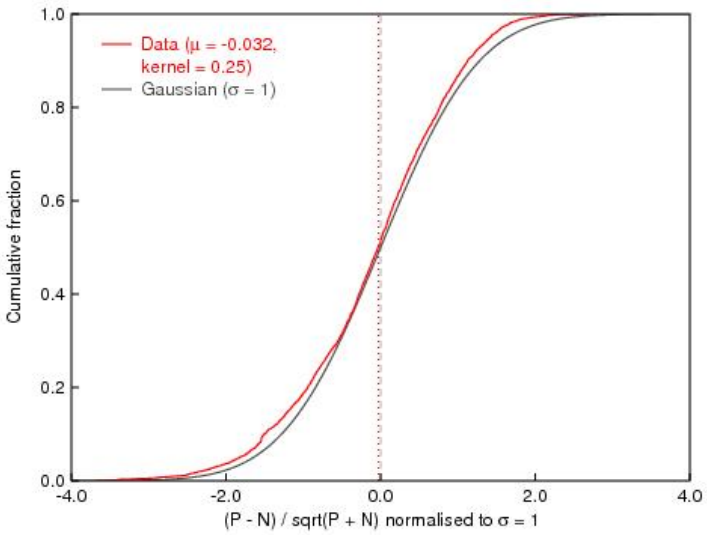


Figure 3.8: Cumulative Skellem distribution of perfect Gaussian noise (in grey) and of negative pixels in our data (in red) normalised to 1, where the `reliability.scaleKernel` is set to 0.25.

ing removed from the source mask. This is likely because setting this parameter higher results in the Gaussian kernel being much larger than the actual source, resulting in negative pixels influencing and inaccurately lowering the R_{sofia} value of real sources. When this is the case, median (μ, represented by the red line in Figure 3.8) is pushed into the positive regime (i.e. to the right of the grey line) which can be seen in Figure 3.9, where `reliability.scaleKernel` was set to 0.4. This occurs because a larger number of positive and/or negative pixels are being included when calculating K. This results in a normalised cumulative Skellam distribution with a $\mu > 0$, which is not comparable to the cumulative Skellam distribution produced by a Gaussian.

On the other hand, we found that setting `reliability.scaleKernel`< 0.25 results in noise peaks being marked as reliable, including them in the final source mask. This is likely caused by the Gaussian kernel being too small, resulting in inaccurate R_{sofia} values caused by the low sampling of positive and negative detections. In this case, μ is pushed to the negative regime (i.e. to the left of the grey line in Figure 3.8), which can be seen in Figure 3.10, where `reliability.scaleKernel` was set to 0.1. In this case, fewer positive and/or negative pixels are being included when calculating K, resulting in a normalised cumulative Skellam distribution with a $\mu < 0$. We therefore found that the `reliability.scaleKernel` parameter is optimal when μ is as close to zero as possible, as this will ensure that the scaled kernel is close to that produced by stochastic noise, which we expect to be Gaussian.

The `reliability.threshold` parameter, which was set to 0.5 in this work, sets the minimum reliability threshold. In other words, the reliability of a detected source (R_{sofia} from Equation 3.2) must be greater than 50 percent for it to be added to the final source

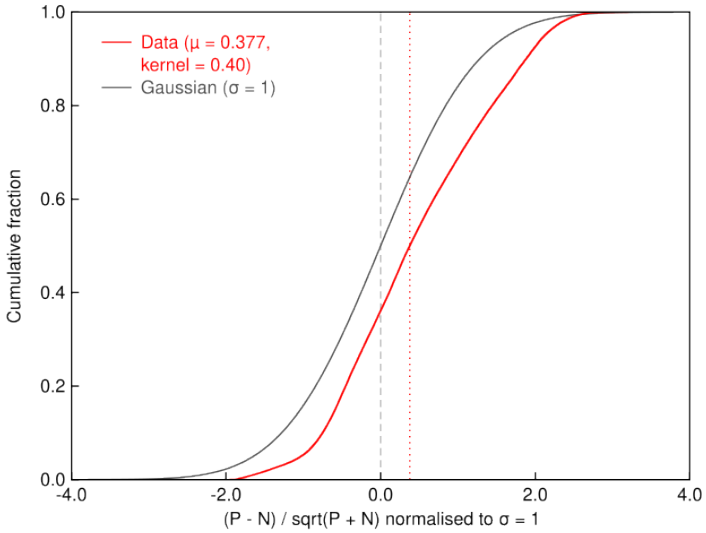


Figure 3.9: As in Figure 3.8 but with `reliability.scaleKernel` set to 0.4.

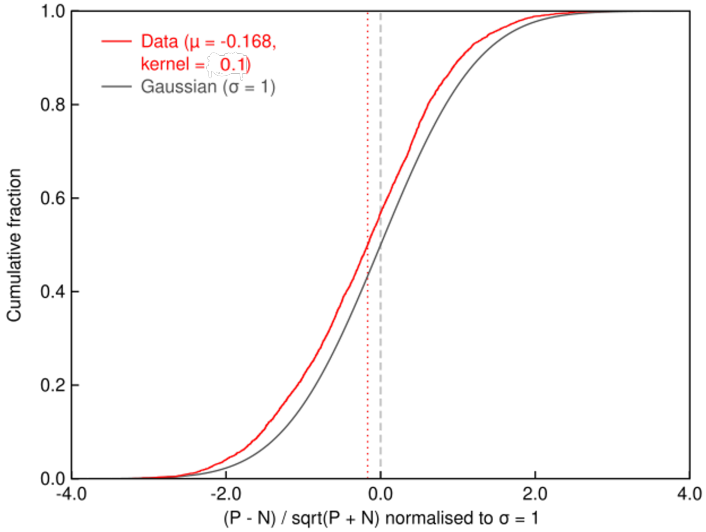


Figure 3.10: As in Figure 3.8 but with `reliability.scaleKernel` set to 0.1.

Key property	LADUMA	WALLABY	SMGPS
Beam size [″]	8×7.5	30	29.7×26.3
Pixel size [″]	2	6	3
Channel width [kHz]	104.52	18.5	210
Per channel flux sensitivity [mJybeam^{-1}]	0.033	1.6	0.36
Area sky coverage [deg^2]	0.8	180	90

Table 3.2: Comparison of the key properties of the LADUMA (utilised in this work), WALLABY (utilised in Westmeier et al. 2022), and SMGPS (utilised in Rajohnson et al. 2022) surveys.

Parameters	This work	Hydra	Vela
`S+C parameters`			
`scfind.kernelsXY` [arcsec]	0, 6, 10	0, 30, 60	0, 21, 45
`scfind.kernelsZ` [kHz]	0, 313.56, 522.6, 731.64	0, 55.5, 129.5, 277.5	0, 630, 1050, 1470
`scfind.threshold` [μJybeam^{-1} per channel]	132	1365	5600
Linker parameters			
`linker.minSizeXY` [arcsec]	10	48	30
`linker.minSizeZ` [kHz]	522.6	92.5	630
`Reliability parameters`			
`reliability.threshold`	0.5	0.8	0.9
`reliability.scaleKernel`	0.25	0.4	0.25
`reliability.minSNR`	3	2.8	3
Number of Detections	193	272	471
Number of False Detections	39	3	30

Table 3.3: The physical scales of the optimal SoFiA-2 parameters found in this work, Westmeier et al. (2022) (Hydra) and Rajohnson et al. (2024) (Vela).

mask. This parameter dictates whether to favour completeness or reliability. We find that setting the `reliability.threshold` parameter > 0.5 favours reliability, as noise peaks are likely to have a lower $R_{\rm sofia}$ value. However, doing so affects the completeness of the final catalogue produced, as real HI sources that exist close to the noise level of the data cube have a lower $R_{\rm sofia}$ value. $R_{\rm sofia}$ is included in the output catalogue produced by SoFiA-2, meaning that the catalogue can be filtered to favour reliability when appropriate. We found that setting `reliability.threshold` to 0.5 allows us to favour completeness while still ensuring a catalogue of reasonable reliability. Section 3.2 will discuss reliability and completeness in more detail.

3.1.5 Comparison to previous work

For further analysis of the optimal parameters found in this work, we compare it to those used in Westmeier et al. (2022) and Rajohnson et al. (2024), both of which utilised SoFiA-2 to perform blind HI extragalactic source finding on the Hydra cluster and the Vela supercluster, respectively. The Hydra cluster was observed by the WALLABY survey and the Vela supercluster was observed by the SMGPS survey. The key properties of these two surveys in comparison to the LADUMA survey are shown in Table 3.2. In Table 3.3, the physical scales of the SoFiA-2 parameters used in this work, along with those used in Westmeier et al. (2022) and Rajohnson et al. (2024) are given.

Larger values of `scfind.kernelsXY` were used in both Westmeier et al. (2022) and Rajohnson et al. (2024) in comparison to this work (where `scfind.kernelsXY` was set to 0, 2, 5). With `scfind.kernelsXY` set to 0, 5, 10 in Westmeier et al. (2022) and `scfind.kernelsXY` set to 0, 7, 15 in Rajohnson et al. (2024). This is likely due to the larger beam sizes of these data sets, with WALLABY and SMGPS having beam sizes that are $\sim$ 27% and $\sim$ 29% larger than the LADUMA beam size, respectively. Because of LAUDMA's higher spatial resolution, we could utilise smaller pixel smoothing iterations than those presented by Westmeier et al. (2022) and Rajohnson et al. (2024).

Westmeier et al. (2022) set the final two values of `scfind.kernelsZ` to higher than that set in this work. With `scfind.kernelsZ` set to 0, 3, 7, 15 in Westmeier et al. (2022) and `scfind.kernelsZ` set to 0, 3, 5, 7 in this work. However, when taking the physical scales into account (as shown in Table 3.3), Westmeier et al. (2022) utilised smaller spectral smoothing iterations than those used in this work. This is likely because the WALLABY data has a higher spectral resolution than the LADUMA data, with the channel width of WALLABY being $\sim$ 18% smaller than that of the LADUMA data. Because of this, Westmeier et al. (2022) could utilise smaller spectral smoothing iterations. On the other hand, Rajohnson et al. (2024) set `scfind.kernelsZ` to the same values used in this work (0, 3, 5, 7), however, the SMGPS data has a lower spectral resolution than that of the LADUMA data (as shown in Table 3.2. Since the channel resolution of SMGPS is $\sim$ 50% larger than LADUMA's, Rajohnson et al. (2024) had to utilise larger spectral smoothing values in comparison to this work.

Both Westmeier et al. (2022) and Rajohnson et al. (2024) set `scfind.threshold` to 3.5σ, whereas we set it to 4σ in this work. However, the per channel flux sensitivity is larger in these data cubes, with WALLABY and SMGPS having a per channel flux sensitivity that is $\sim$ 2% and $\sim$ 9% larger, respectively. This further emphasises the sensitivity of the LADUMA survey in comparison to other HI surveys. For this reason, we could probe deeper into the data with a lower set `scfind.threshold` to detect fainter HI galaxies.

In this work, we set the `scfind.threshold` parameter lower than that set in Westmeier et al. (2022) and Rajohnson et al. (2024). This was done to minimise the amount of artefacts included in the final catalogue, thereby prioritising completeness over reliability. The catalogues produced by Westmeier et al. (2022) and Rajohnson et al. (2024) had a reliability of 99% and 93%, respectively. This is higher than the reliability of the catalogue produced in this work ($\sim$ 83%). This is because we chose to prioritise completeness over reliability in this work to take full advantage of the low sensitivity of the LADUMA data, hence why we set `scfind.threshold` lower than that set in Westmeier et al. (2022) and Rajohnson et al. (2024). The completeness and reliability of our catalogue are discussed in detail in Section 3.2.

Both Westmeier et al. (2022) and Rajohnson et al. (2024) detected significantly more

sources than those presented in this work. This is not due to poorly chosen parameters on our part, but rather because of the differences in the field size. The WALLABY observation of Hydra is 225 times larger and the SMGPS observation of the Vela super cluster is ~ 112 times larger than the LADUMA sky area coverage. So it is expected that fewer sources were found in this work in comparison to Westmeier et al. (2022) and Rajohnson et al. (2024).

In conclusion, SoFiA-2 is highly customisable due to its various parameters. This is a strength, as it allows you to customise your output catalogue for your specified scientific case. However, this comparison with previous implementations of SoFiA-2 emphasises the importance of optimising your parameters to the data cube on which source finding is being performed on. Many of the parameters utilised in Westmeier et al. (2022) and Rajohnson et al. (2024) would not have given good results on the LADUMA ES cube and vice versa. This is due to the differences in beam size, flux sensitivity and spectral resolution from dataset to dataset.

3.2 Assessing performance of SoFiA-2 source finding

In this section, we evaluate the success of the SoFiA-2 source finding over the LADUMA ES cube. By implementing the optimised SoFiA-2 parameters summarised in Table 3.1, we identified 232 candidate sources. However, we must confirm the validity of each candidate source to create our final HI sample, as SoFiA-2 is not a perfect source finder.

A perfect source finder would be able to 1) identify every astronomical source detected in a data cube, regardless of its size or flux and 2) evaluate what flux in the data cube is caused by astronomical sources versus that from noise peaks or artefacts. In other words, a perfect source finder would have a reliability and completeness of 100%.

The reliability (R) of a source catalogue is defined by the fraction of sources detected by the source finder (N_{detect}) that are genuine detections (N_{real}), i.e.

$$R = \frac{N_{detect}}{N_{real}} \times 100\%. \tag{3.4}$$

The completeness (C) of a source catalogue is defined as the fraction of total sources detected in the data (N_{tot}) that are genuine detections (N_{real}), i.e.

$$C = \frac{N_{real}}{N_{tot}} \times 100\%. \tag{3.5}$$

In reality, a perfect source finder does not exist, and a trade-off between reliability and completeness is required. In the context of SoFiA-2, this is governed by the `reliability.threshold` parameter (see Chapter 3.1.4). Lower flux density HI sources can behave similarly to noise peaks that exist in the data cube, and so favouring reliability may result in a final catalogue that has a lower completeness and vice versa. In this work, we chose an implementation of SoFiA-2 that favours completeness to ensure the production of an extensive low-redshift HI catalogue for our sample. However, our SoFiA-2 catalogue will not be 100% complete close to the flux limit of the data. Therefore, further work would be required when using this catalogue for studies that require completeness correction.

Accurately calculating the true completeness and reliability of a produced catalogue is a far more complex task than it initially appears. Ideally, you would need to know the exact number of real galaxies that exist in the field of observation, which we do not know a priori for a blind survey like LADUMA. In practice, the correct way to assess the completeness and reliability of a catalogue produced by a source finder is to run it on a mock data cube with injected galaxies, where the exact number of galaxies injected into the data cube is known, and examine the fraction recovered by the source finder. This is beyond the scope of this project and is investigated by Kazemi-Moridani (2024) for the LADUMA survey.

Instead, to crudely estimate the reliability of the catalogue produced in this work, we examine whether optical counterparts could be identified for each SoFiA-2 HI candidate source. To do this, we made use of the SoFiA Imaging Pipeline[4] (SIP) created by Dr. Kelly Hess. SIP is a plotting software that overlays the HI contours of sources detected by SoFiA-2 onto images from a selection of multiwavelength surveys. Along with this, SIP produces useful diagnostic products from the HI data, such as moment 0 maps, pixel-by-pixel SNR maps, moment 2 maps, position-velocity (PV) plots and spectra. For this work, we created the overlays using images from the Dark Energy Camera Legacy Survey (DECaLS; Blum et al. 2016) and visually inspected them to identify candidate sources that have no visible optical counterpart. We then examined the HI spectra of these sources to confirm that there is no obvious HI detection. We found 39 out of the 232 candidate detections have no optical counterparts and low HI SNRs. Assuming these are false detections and the remaining 193 sources are real, the confirmed SoFiA-2 catalogue then has a "reliability" of $\sim 83\%$. Examples of SIP images of sources with and without an optical counterpart, along with their corresponding HI spectra, are given in Figure 3.11.

We note that using the above method to validate detections introduces several biases. By relying on optical counterparts to validate HI detection, we run the risk of eliminating genuine dark HI galaxies (galaxies that are bright in the HI but have no optical counterpart) from our sample. This bias should not affect our subsequent analysis, which requires multiwavelength detections for SED fitting purposes. That being said, it may, however, affect the overall robustness of our final HI catalogue. This method introduces

[4] https://github.com/kmhess/SoFiA-image-pipeline

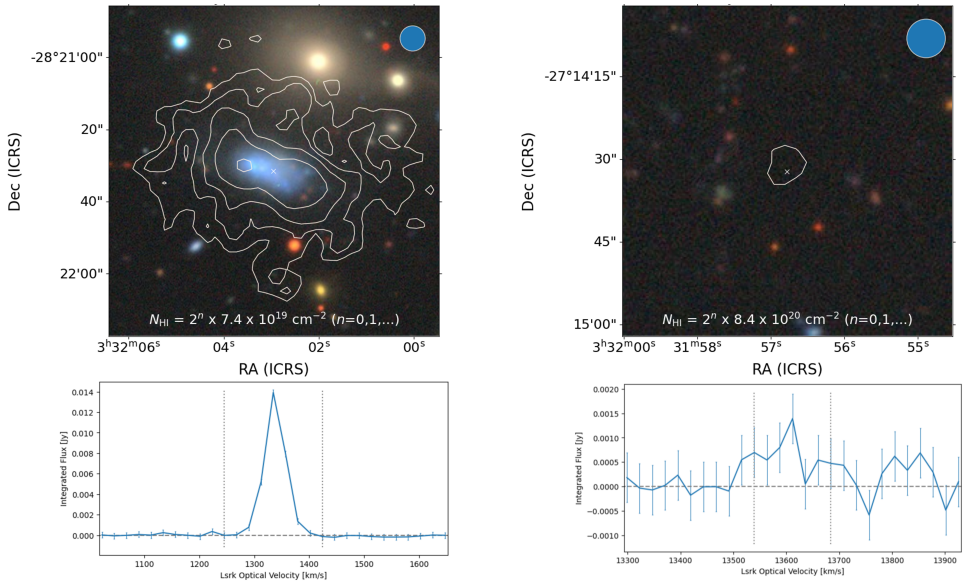


Figure 3.11: Examples of HI contours overlayed onto DECaLS optical colour images of detections made by SoFiA-2. (top row) with their corresponding HI spectra (bottom row), where the y-axis shows flux density [Jy] and the x-axis shows heliocentric optical velocity [km/s]. To the left is an example of a detection that has an optical counterpart. To the right is an example of a detection without an optical counterpart. Images generated by the SoFiA Imaging Pipeline (SIP).

another bias towards bright galaxies, as it is possible we discarded real, optically faint, low SNR HI galaxies.

The coordinates of the 39 false detections with no optical counterpart are represented by the crosses in Figure 3.12, and the distribution of their detected HI central frequencies is shown in Figure 3.15a. From these figures, we see that a majority of the detections without optical counterparts were found at frequencies between 1304 and 1340 MHz (corresponding to $0.06 < z_{HI} < 0.088$). Additionally, a majority of these false detections exist on the top right-hand section of the LADUMA ES cube. These correspond to areas and frequency channels of the LADUMA ES cube that were most affected by noise and artefacts caused by the bright continuum source (shown by the green square in Figure 3.12). It is therefore likely that a majority of these candidates are false detections of residual continuum emission or sidelobes that were not removed during the cleaning and continuum subtraction process. These candidates were therefore deemed to be unreliable and removed from our final catalogue, leaving us with a final HI sample of 193 confirmed detections.

The catalogue output by SoFiA-2 includes the bounding box of each source detected in the x, y and z direction in pixels. Using this and the pixel scale and spectral width of the LADUMA cube, we can calculate the spatial and spectral extent of each source. We analysed these for the 39 false detections and compared them to those of the confirmed

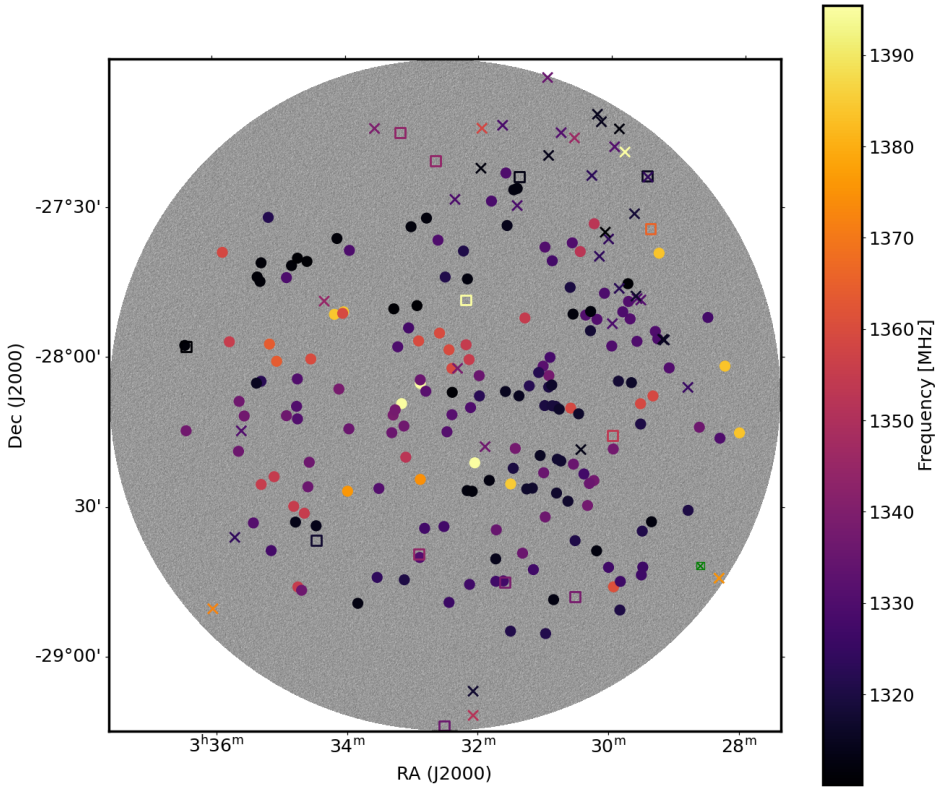


Figure 3.12: Positions of all sources found via SoFiA-2. The circles represent sources that were found by both SoFiA-2 and at least one other method of source finding. The squares represent sources that were found only by SoFiA-2. The crosses represent the 'false detections' with no optical counterpart. Colour indicates the central frequency of the source. The green square with a cross in the centre shows the region masked off during source finding to avoid the bright continuum source. The background image shows one channel of the LADUMA ES cube at 1420 MHz.

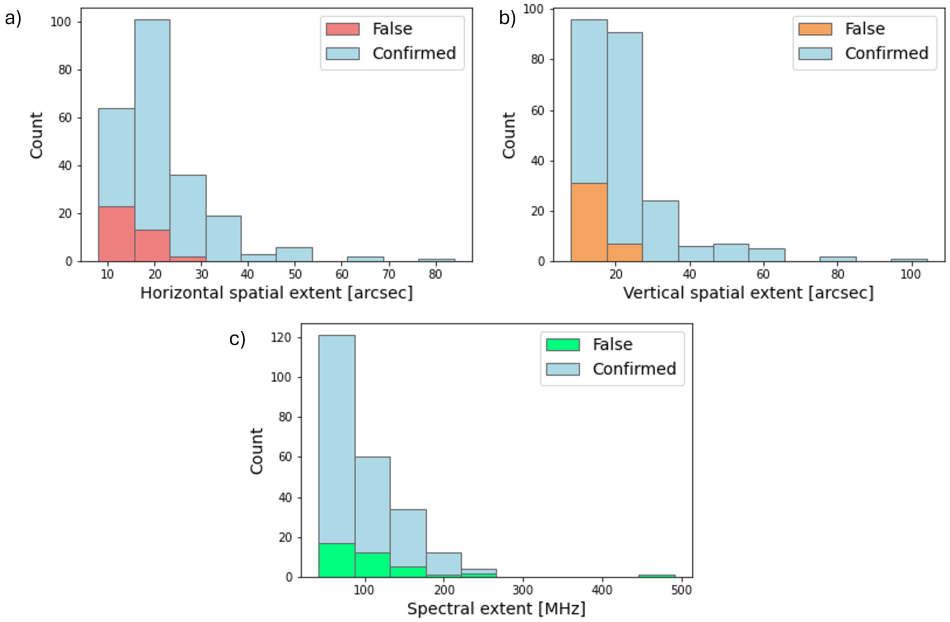


Figure 3.13: Histograms showing the spectral (in arcseconds) and spatial (in MHz) extent of the 39 'false detection' (marked by crosses in Figure 3.12) in comparison to the confirmed detections (shown in blue and marked by square and dots in Figure 3.12)). a) the horizontal spatial length, b) the vertical spatial length, and c) the spectral length.

real detections, to assess potential differences in size. This comparison is illustrated in Figure 3.13. The average spatial extent of the false detections is $\sim 14''$, which is $\sim 9''$ smaller than the average spatial extent of the confirmed real detections ($\sim 23''$). The average spectral length of the false detections is ~ 114 MHz, which is ~ 16 MHz (or 8 spectral channels) larger than the average spectral extent of the confirmed real detections (~ 98 MHz). These results suggest that many noise peaks in the LADUMA ES cube are, on average, smaller spatially and longer spectrally than real sources that exist in the data cube.

To crudely estimate the completeness of the confirmed SoFiA-2 catalogue, we compare it to manual and semi-automated methods of source finding that were performed on the LADUMA ES cube[5]. Manual source finding was performed in a similar manner to that described in Ponomareva et al. (2021) and Rajohnson et al. (2022). The semi-automated source finder applied utilises a 'matched filtering' algorithm[6] and will be detailed in Baker et al (in prep). To briefly summarise, this algorithm works by setting a filter slightly smaller than the size of the galaxies expected to be detected during source finding. This filter is then applied at every pixel of the data cube, with the mean flux extracted at each position. The distribution of the mean fluxes will then be Gaussian when produced by noise and outliers when produced by candidate detections. This method is considered semi-automated because, unlike SoFiA-2, the matched filtering does not do any reliability

[5]by other members of the LADUMA team
[6]written and implemented by Prof. Ed Elson (UWC)

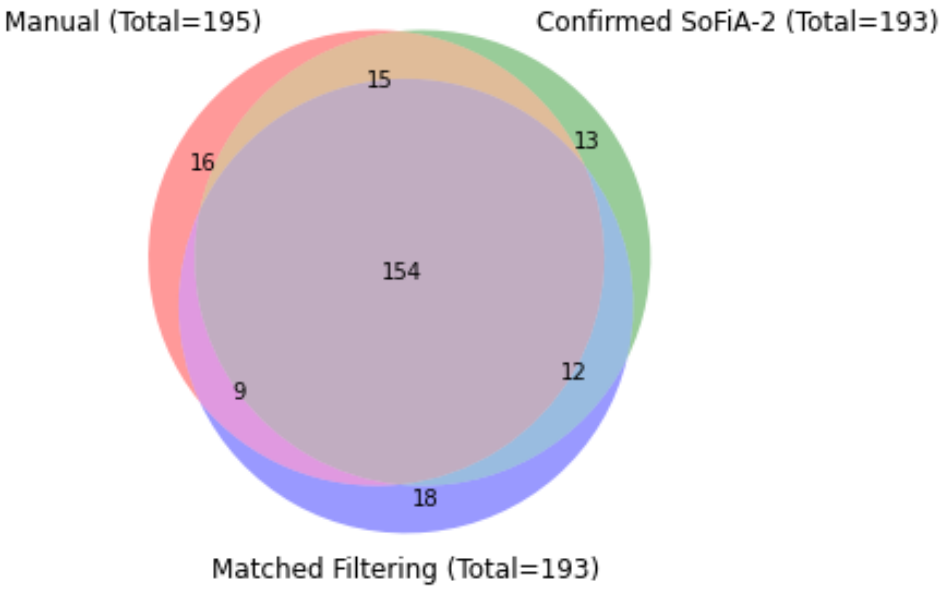


Figure 3.14: Venn diagram illustrating cross matching results between the confirmed SoFiA-2 catalogue (i.e. excluding source with no optical counterpart), the visual source finding catalogue and the matched filtering source finding catalogue.

calculations, meaning the validity of candidates must be confirmed by eye. Therefore, this method essentially prioritises completeness over reliability.

All three catalogues produced by the three different source finding methods contained a similar number of objects (193/195/193 for SoFiA-2/manual/matched filtering source finding). We perform a three-way cross-match between them. Since each source finding catalogue uses different methods to encapsulate the flux of each source, the central coordinates of one source can differ from catalogue to catalogue. For this reason, we chose a 15″ cross matching radius to account for this discrepancy. Cross-matching is done by identifying the nearest source in each of the other two catalogues centred on each SoFiA-2 source position. We then do the same but centred on each matched filtering source position, and lastly centred on the position of each visual source position[7]. Using this strategy, we determined which galaxies were found by all, two or one of the source finding methods. The results are shown in Figure 3.14. A total of 237 unique sources were identified, 65% of which were detected by all three methods. Crudely assuming all existing HI sources with sufficient SNR have been identified via at least one of the source finding methods, our SoFiA-2 catalogue has an estimated "completeness" of 82%.

We examined the 13 sources that were only detected by SoFiA-2 to determine why the alternative methods of source finding did not detect them. Their coordinates are represented by the squares in Figure 3.12 and their frequency distribution is shown in Figure 3.15b). From these figures, we see that the majority of sources detected by SoFiA-2 were

[7]executed via a Python code

only at frequencies <1350 MHz (corresponding to $z_{HI} > 0.05$). This is the higher z_{HI} region of the LADUMA ES cube, which is most affected by noise, therefore, sources in this region likely have a low SNR. This demonstrates how powerful SoFiA-2 can be at detecting low SNR sources that would likely go undetected by alternative methods of source finding. The spatial extent of these 13 sources is, on average, similar to those found by other methods of source finding. However, their average spectral extent is significantly smaller, at $\sim$ 72 kHz (or $\sim$ 7 spectral channels), in comparison to the average spectral extent of those found by other methods of source finding ($\sim$ 98 MHz). This emphasises that the spectral smoothing utilised by SoFiA-2 allows us to detect smaller galaxies.

Out of all the sources found by the various forms of source finders, 25% (58) of them were undetected by SoFiA-2. We found no obvious trend in the positions or frequencies of these sources. It is likely that SoFiA-2 rejected these sources due to their low reliability, i.e. their $R_{\text{sofia}} < 0.5$ (calculated via Equation 3.2), meaning they were indistinguishable from the noise. We chose not to lower the reliability threshold further to capture these sources, as this would result in a significant increase in false detections.

Overall, SoFiA-2 is not a perfect source finder. However, when optimised, it can be a powerful tool in detecting sources that can be found via other methods of source finding along with low SNR sources that would likely go undetected. Additionally, it is extremely efficient (with each run taking $\sim$ 45 minutes) in comparison to alternative methods. For these reasons, SoFiA-2 is a powerful tool that can be utilised by HI LSPs to detect higher z_{HI} galaxies. However, some checks will still need to be made to validate the reliability of the SoFiA-2 detections.

3.3 Properties of the HI sample

Our final SoFiA-2 detected sample consists of 193 galaxies of various different morphologies, including low brightness galaxies, mergers and several beautiful spiral galaxies. The sensitivity of LADUMA also reveals low column density gas (as low as $\sim 5 \times 10^{19}\text{cm}^{-1}$) between interacting pairs and at least one HI tail. Figure 3.16 shows the DECaLs optical colour image overlaid with the HI contours of three galaxies from our sample as an example. The optical images overlaid with HI contours for each galaxy in our final sample is given in Appendix A.3.

The coordinates of the 193 galaxies that make up our HI sample are represented by the circular and square points in Figure 3.12 and their frequency distribution are shown in Figure 3.15c). From Figure 3.15c), we see that a majority of the sources in our HI sample were detected at frequencies below 1340 MHz (corresponding to $z_{HI} > 0.06$). This can be explained by the larger sky volume that is probed by the LADUMA pointing at higher redshifts. It is therefore expected that more galaxies will be detected towards higher redshifts. Additionally, the LADUMA field was selected specifically to avoid nearby objects, explaining why fewer galaxies are found in the lower redshift regime.

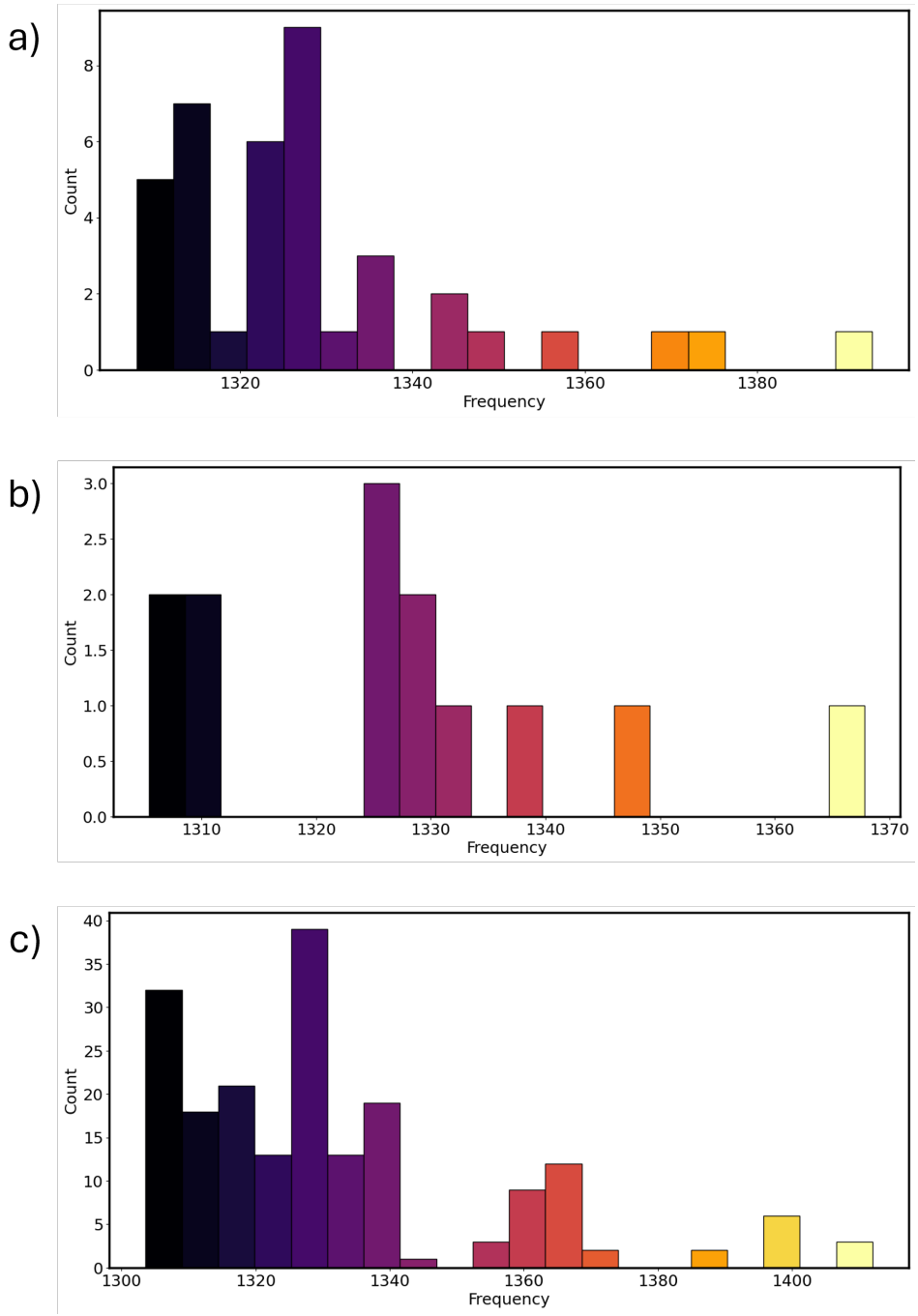


Figure 3.15: Frequency distributions of sources found via SoFiA-2: a) 'false detections' with no optical counterpart, b) detections found only by SoFiA-2, and c) all confirmed sources.

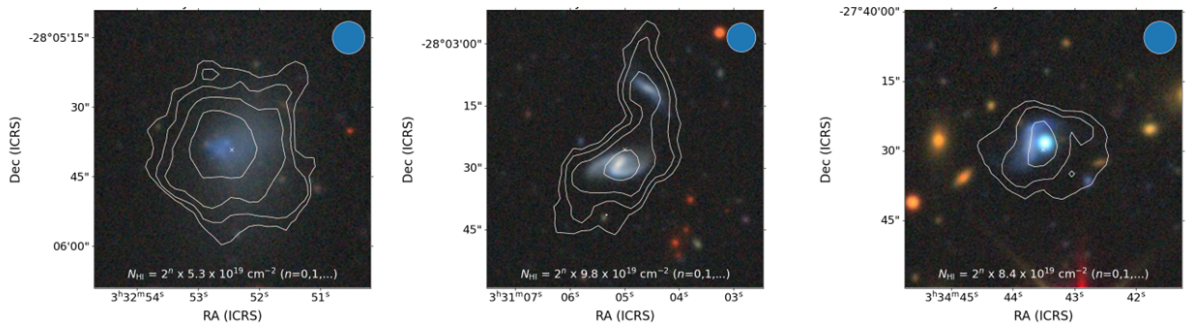


Figure 3.16: The optical images (from DECaLS) overlaid with HI contours of three galaxies (ID 1, 152 and 231) in our HI sample.

Using Equation 1.5, we calculated the HI mass (M_{HI}) of our final sample using the flux densities from the pb corrected cube. The HI mass-redshift parameter space and distributions of our sample are shown in Figure 3.17. Note that the HI mass uncertainties are not included in this figure as they are too small to be visible. A description of how the uncertainty values are estimated is discussed in Appendix A.2. Additionally, we present the HI mass and uncertainties of our final sample, along with their ID numbers, coordinates and redshifts in Table A.1.

Looking at Figure 3.17, we see apparent overdensities of galaxies at HI redshifts of approximately z_{HI} 0.04, 0.06 and 0.07. This is likely a result of existing large-scale structures in the field. Comparing Figure 3.17 to Figure 1.3, we see that LADUMA has probed lower HI masses at low redshifts than was previously projected by Maddox et al. (2021), with the lowest HI mass detection at $\log(M_{HI}) \sim 7.03 M_{\odot}$. Looking at the HI mass distribution of our sample, the majority of galaxies have an HI mass $8.5 M_{\odot} < \log(M_{HI}) < 10 M_{\odot}$, which falls in the higher mass regime. This is expected as a galaxy must have a significant amount of HI to be detected, due to the faint emission of the HI line. This creates a sample with a bias towards high HI mass galaxies.

3.4 Chapter Summary

In this chapter, we utilised SoFiA-2 to perform blind source finding on the low redshift ($0 < z_{HI} < 0.088$) LADUMA ES data cube. SoFiA-2 was run iteratively to optimise parameters for the data cube. The optimal parameters used in this work are summarised in Table 3.1, with the S+C parameters, linker parameters and reliability parameters having the most influence on the source finding performance. We compare the optimal parameters used in this work to those presented by Westmeier et al. (2022) and Rajohnson et al. (2024). This comparison emphasises how customisable SoFiA-2 is to your data set and how important it is to optimise the parameters accordingly.

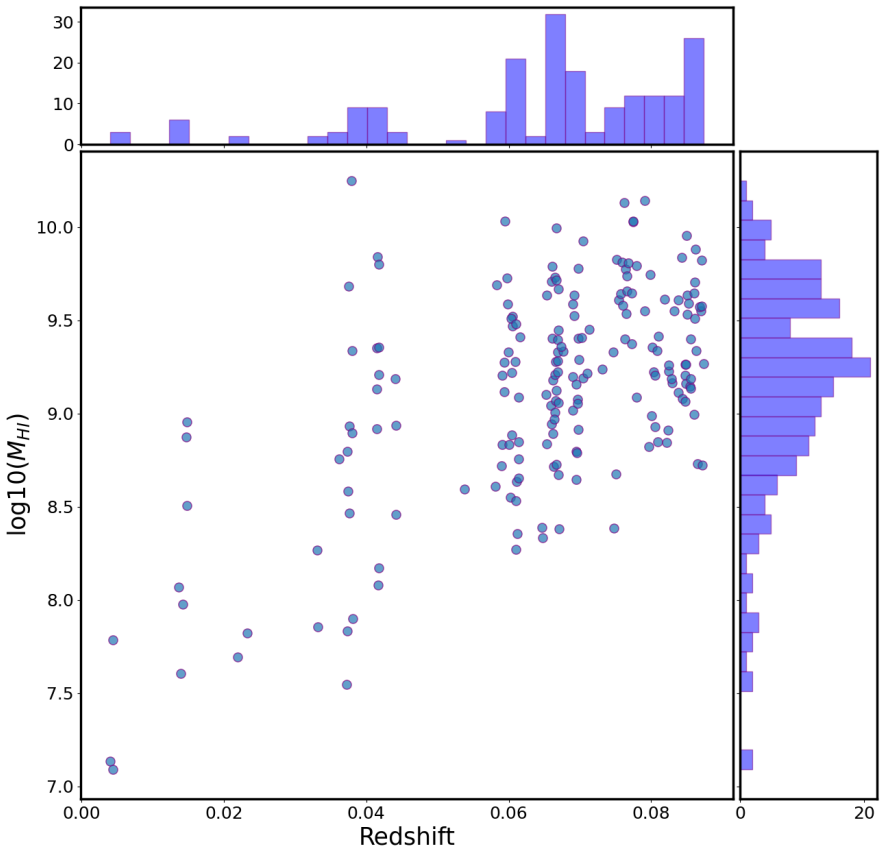


Figure 3.17: HI mass-redshift parameter space of our HI sample, with corresponding histograms along axes.

Using the parameters listed in Table 3.1, SoFiA-2 detected 232 candidate detections. To crudely evaluate the reliability of these sources, we visually inspected their optical counterparts and HI spectra with SIP to determine which candidate sources are trustworthy. We found that 39 candidate sources have no optical counterpart and low HI SNRs. These are likely false detections and were therefore removed from our sample, leaving us with 193 confirmed SoFiA-2 sources.

To crudely estimate the completeness of our HI sample, we compare our confirmed SoFiA-2 catalogue to those produced via manual source finding and matched filtering source finding. Assuming that all existing HI sources with sufficient SNR have been identified via at least one method, our confirmed SoFiA-2 catalogue has an estimated completeness of 82%. We emphasise that this is a crude estimation of completeness and that the HI catalogue compiled in this work should not be used for studies that require completeness corrections. The catalogue produced by Kazemi-Moridani (2024) would be more appropriate for this science case, as they used an injection and recovery method alongside a higher flux density limit to measure the HI mass function based on the LADUMA data. Out of the 237 unique sources found by all three methods of source finding, 13 of them were found only by SoFiA-2. After examining the spatial and spectral extent of these sources, we found that the sources found only by SoFiA-2 are, on average, $\sim$ 72 kHz smaller than the other confirmed sources. This emphasises that SoFiA-2 allows us to detect smaller galaxies.

We found that the majority of our HI sample has a $z_{HI} > 0.06$, likely due to the larger sky volume probed towards higher redshifts and the LADUMA field selection, which was specifically chosen to avoid nearby objects. We observe overdensities of galaxies at $z_{HI} \sim 0.04$, 0.06 and 0.07. This likely is due to the existence of large-scale structures. Examining the HI masses of the 193 confirmed sources reveals that the majority of our sample have HI masses $8.5M_{\odot} < \log(M_{HI}) < 10M_{\odot}$. This is expected, since the faintness of the HI line creates a bias towards higher HI mass galaxies.

4 | Multiwavelength Properties

We will now examine the multiwavelength properties of our HI selected sample. This involves obtaining the multiband photometry of our galaxy sample from the DEVILS catalogue, which will then be used to perform spectral energy distribution (SED) fitting. This will give us properties such as the stellar masses, star formation rates (SFRs) and star formation histories (SFHs) of our galaxy sample. In Section 4.1, we discuss how cross-matching was done between the final HI catalogue and the DEVILS photometric catalogue to obtain multiband photometry. Section 4.2 discusses how we implemented ProSpect, the free priors and the Markov Chain Monte Carlo (MCMC) set-up used to perform SED fitting. Section 4.3 discusses the SED fitting results.

4.1 Cross-matching with DEVILS

To obtain multiwavelength photometry of our HI sample, we cross-match our final confirmed SoFiA-2 detected HI catalogue with the DEVILS photometric catalogue, using a 10″ search radius. Unlike the catalogues produced by alternative HI source finding, the DEVILS catalogue was produced to ensure all flux of each galaxy is encapsulated. This means that there is less discrepancy between this catalogue's coordinates and those outputted by SoFiA-2, allowing us to use a smaller cross-match radius in this case. Cross matching was done by identifying the nearest DEVILS source within a 10″ search radius centred on each SoFiA-2 position[1]. As discussed in Section 2.2, the DEVILS photometric catalogue does not cover the entire LADUMA field, meaning that a portion of the galaxies in our HI sample do not appear in the DEVILS catalogue. Galaxies from our HI sample that are outside of the DEVILS field are therefore removed from our sample due to a lack of coverage by the DEVILS catalogue. After cross-matching, our sample of 193 galaxies was reduced to 135 galaxies. All confirmed SoFiA-2 detections within the DEVILS field have a counterpart in the DEVILS photometric catalogue.

Due to the lower sensitivity of the MIR-FIR imaging used to compile the DEVILS catalogue (discussed in Section 2.2.2), the fraction of detections made in this wavelength regime varies depending on the depth of imaging [Davies et al. 2021]. The number of objects in our sample with detections in the MIPS24, MIPS70, P100, P160, S250, S350 and S500 bands are shown in Table 4.1. In our sample of 135 galaxies, 19% have detections in all 7 MIR-FIR bands, 34% have detections in more than half of the MIR-FIR bands (corresponding to coverage in 4, 5 or 6 bands), 28% have detections in less than half of the MIR-FIR band (corresponding to coverage in 1, 2 or 3 bands) and 19% have no detections in any of the MIR-FIR bands. These statistics are also summarised in Table 4.1.

It is important to note that Davies et al. (2021) did not extract the MIR-FIR photometry for galaxies with a Y band magnitude > 21.2 mag (discussed in Section 2.2.2). Davies et al. (2021) reported that all objects with a Y band magnitude < 21.2 mag were detected

[1] executed via a Python code

Bands	Galaxies with detections	Corresponding percentage [%]
MIPS24	89	66
MIPS70	71	53
P100	65	48
P160	69	51
S250	78	58
S350	68	50
S500	53	39
All bands	26	19
4-6 bands	46	34
1-3 bands	38	28
No bands	25	19

Table 4.1: Number and corresponding percentage of the 135 HI galaxies in our final sample detected in; 1) each MIR-FIR band (upper panel), and 2) the given number of MIR-FIR bands (lower panel).

in all seven of the MIR-FIR bands. We found the same to be true in our sample, meaning the 25 galaxies undetected in every MIR-FIR band are galaxies with a Y band magnitude > 21.2 mag. We perform SED fitting on all 135 galaxies in our sample, regardless of their detection fraction, to avoid further reduction of our sample.

To account for offsets between the different facilities and instruments used to capture the multi-band photometry and fluctuations in zero-point offset, we assume a 10% flux uncertainty for each photometric measurement and used these to perform SED fitting on our final sample of 135 galaxies.

4.2 SED fitting with ProSpect

To perform SED fitting on the 135 HI galaxies with photometric coverage, we use the state-of-the-art SED fitting algorithm ProSpect [Robotham et al. 2020]. ProSpect makes use of BC03 [Bruzual and Charlot 2003] stellar libraries and the initial mass function (IMF) presented in Chabrier (2003) to model stellar components. For dust attenuation, ProSpect uses a two-component description of the interstellar medium presented in Charlot and Fall (2000), which makes use of 1) a diffuse dust component to attenuate all stellar emission, and 2) a birth cloud, which only attenuates emission from younger stars that are < 10Myr old. Along with this, ProSpect makes use of templates presented in Dale et al. (2014) to model the re-radiation of stellar emission absorbed by dust and emitted into the infrared.

Although ProSpect uses the same underlying templates as other SED modeling codes (eg. magphys [Da Cunha et al. 2008] and beagle [Chevallard and Charlot 2016]), it has exceptional flexibility in constructing star formation histories (SFHs), offering 13 built-in

Priors Category	Priors	Units	Scale	Fitting Range	Initial Condition
Star Formation History	mSFR	Myr^{-1}	log	[-3,4]	3.162
	mpeak	Gyr	linear	[-2,13.38]	6
	mperiod	Gyr	log	[log(0.3),2]	1.977
	mskew	Gyr	linear	[-0.5,1]	0
Metallicity	Zfinal	-	log	[-4,-1.3]	0.001
Dust	τ_{birth}	-	log	[-2.5,1.5]	0.316
	τ_{screen}	-	log	[-5,1]	0.01
	α_{birth}	-	linear	[0,4]	2
	α_{screen}	-	linear	[0,4]	2

Table 4.2: All free priors used when running ProSpect. Including their units, fitting regime (whether fitting in logarithmic or linear space), set range and initial guess used for the MCMC set-up.

algorithms for processing SFHs. Along with this, ProSpect incorporates an evolving metallicity, making it the first of its kind, as all other SED fitting codes assume a constant metallicity [Robotham et al. 2020]. This inclusion should result in more accurate star formation histories, as there is a known degeneracy between stellar age and metallicity [Worthey 1994].

In this work, we apply to our galaxy sample a similar implementation of ProSpect to that outlined in Bellstedt et al. (2020a) and Thorne et al. (2021). In the following sections, we discuss the method used to run ProSpect and the respective priors used. Table 4.2 summarises all free priors used for the SED modelling along with our chosen initial condition and the allowed ranges for the MCMC fit.

4.2.1 Modelling Star Formation History

As discussed above, ProSpect contains multiple algorithms to model the star formation history of galaxies. The SFH reflects the timeline of a galaxy's star formation activity. In other words, it traces the star formation rate (SFR) of a galaxy at a given lookback time, where 0 Gyrs is the time of observation.

ProSpect estimates the stellar mass and SFR of a galaxy via the parametrisation of the SFH. To model the star formation histories of our galaxy sample, we make use of the `massfunc_snorm_trunc` function, which uses a skewed Normal distribution with a truncation at the beginning of the Universe, which forces the SFR= 0 at a lookback time of 13.38 Gyrs. The `massfunc_snorm_trunc` function produces a diverse range of simulated SFHs (as illustrated in Figure 10 of Robotham et al. 2020). Additionally, the `massfunc_snor_trunc` produces SFHs which are consistent with measurements of the cosmic star formation history [Bellstedt et al. 2020a]. For these reasons, we chose to utilise this model to fit the SFHs of our HI sample.

The parameterization of the `massfunc_snorm_trunc` function contains six free priors:

1. `mSFR`- the peak star formation rate
2. `mpeak`- the lookback time of peak star formation
3. `mperiod`- the width of the normal distribution
4. `mskew`- the skewness of the normal distribution (where 0 is perfectly normal, positive skews to younger ages and negative skews to older ages)
5. `magemax`- the maximum lookback time of star formation history
6. `mtrunc`- the sharpness of the early-time truncation

The `massfunc_snorm_trunc` function returns the star formation rate (SFR) at a given lookback time (t) and is defined by [reproduced from Bellstedt et al. 2020a]:

$$\mathrm{SFR(t)_{snorm}} = \texttt{mSFR} \times e^{\frac{-X(t)^2}{2}}, \tag{4.1}$$

where

$$X(t) = \left(\frac{t - \texttt{mpeak}}{\texttt{mperiod}}\right)\left(e^{\texttt{mskew}}\right)^{\mathrm{asinh}\left(\frac{t-\texttt{mpeak}}{\texttt{mperiod}}\right)}. \tag{4.2}$$

This parameterisation is truncated to ensure a SFR$=0$ at the start of the Universe. This truncation is implemented as follows:

$$\mathrm{SFR(t)_{trunc}} = \mathrm{SFR(t)_{snorm}} \times \left[1 - \frac{1}{2}\left[1 + erf\left(\frac{t-\mu}{\sigma\sqrt{2}}\right)\right]\right], \tag{4.3}$$

where

$$\mu = \texttt{mpeak} + \frac{|(\texttt{magemax} - \texttt{mpeak})|}{\texttt{mtrunc}}, \tag{4.4}$$

and

$$\sigma = \frac{|(\texttt{magemax} - \texttt{mpeak})|}{2 \times \texttt{mtrunc}}. \tag{4.5}$$

In this work, we fix the values of `mtrunc` and `magemax` while the remaining four are free priors that are fit by an MCMC algorithm (see Section 4.2.4). `magemax` is set to 13.38 Gyr, which corresponds to the redshift of the most distant galaxy known (z$=$ 11, Oesch et al. 2016)[2]. `mtrunc` is set to 2 to ensure strong truncation at the beginning of the Universe [Thorne et al. 2021].

As shown in Table 4.2, `mpeak` can be negative, allowing the SFR peak to occur after the time of observation. This accounts for a rising SFR at the time of observations. The

[2]Since SED fitting work was performed, higher redshift galaxies have been confirmed by the *James Webb Space telescope* [Carniani et al. 2025]

lower limit of `mpeak` was selected due to the sampling of the BC03 templates and allows galaxies' SFR to peak up to 2 Gyr after the observation point.

4.2.2 Modelling Metallicity

As discussed above, one of the strengths of ProSpect is the implementation of an evolving metallicity. The SED produced by assuming a constant metallicity can vary by a factor of a few compared to an SED produced with evolving metallicity (this is well illustrated in Figure 2 of Thorne et al. 2021). Incorporating an evolving metallicity provides much more reasonable solutions that are comparable to the overall shape of the cosmic star formation history [Bellstedt et al. 2020a].

We chose not to assume a constant metallicity and instead implemented a 1:1 evolution of metal enrichment to stellar mass build-up. In other words, as a galaxy's stellar mass assembles, its chemical enrichment occurs at the same rate. This evolution implementation naturally introduces a low initial metallicity during the earliest phases of star formation and higher metallicity for later phases of star formation.

This metallicity model is implemented with ProSpect built-in `Zfunc_massmap_lin` function. This function was implemented by setting a fixed initial metallicity (`Zstart`) of 10^{-4} (which corresponds to the lowest metallicity template in BC03) and allowing the final metallicity (`Zfinal`) to be a free parameter which we fit using the range of metallicities available in the BC03 templates (from 10^{-4} to 0.05).

4.2.3 Modelling Dust

We make use of the Charlot and Fall (2000) two-phase model, built into ProSpect, to model the dust attenuation, where the flux at a given wavelength (λ) is modified by the attenuation factor A, which is defined by [reproduced from Thorne et al. 2021]:

$$A(\lambda) = e^{-\tau(\lambda/\lambda_0)^v}, \tag{4.6}$$

where λ_0 represents the pivot wavelength (which defaults to 5500Å), v is the modified power which we set to 0.7 [Thorne et al. 2021] and τ is the effective optical depth of attenuation.

To model the re-emission in the FIR, we make use of templates from Dale et al. (2014) which make use of an energy balance approach. These templates contain a free parameter, α, which specifies the power law of the radiation field heating the dust.

Overall, we make use of four free priors to model the dust, two of which control the reddening of the dust and two that model the radiation field heating of the dust. They are defined as follows [summarised from Thorne et al. 2021]:

1. τ_{birth}- models the reddening of the dust in birth clouds

2. τ_{screen}- models the reddening of the dust in the diffused ISM
3. α_{birth}- models the radiation field heating the dust in birth clouds
4. α_{screen}- models the radiation field heating the diffused ISM

The 'birth' priors only affect emission from stars that are less than 10 Myrs old while the 'screen' priors affect emission from all stars. Effectively, the τ prior represents the column density of the dust (where larger values of τ correspond to higher column densities of dust) and the α prior represents the temperature of the dust (where larger values of α correspond to colder temperatures). How the α and τ prior affect the overall shape of an SED is well illustrated in Figure 3 of Thorne et al. (2021).

4.2.4 MCMC set-up

In order to fit the free prior discussed above, we implemented ProSpect in a Baysian manner using Markov Chain Monte Carlo (MCMC) fitting. MCMC fitting constructs a chain of dependent samples which eventually converge to a 'best-fit solution'. There are multiple SED templates that would result in a 'good' fit for a given galaxy and using MCMC sampling generates a distribution of possible solutions (known as a posterior distribution), allowing us to quantify realistic uncertainties and generate well-fitted SEDs.

The fitting of our galaxy sample is done in a two-stage process, executed by the `HIGHLANDER`[3] R package. We implemented `HIGHLANDER` in the following manner:

1. The galaxy's HI redshift, the list of free priors, their allowed introduced priors fitting type, and initial conditions (listed in Table 4.2) are inputted into `HIGHLANDER`.
2. A genetic optimizer is then run with 500 steps to find approximate solutions within the allowed fitting range of the nine free priors.
3. These approximate solutions are then fed into an MCMC chain of 100 steps, procuring new approximate solutions for the nine free priors.
4. These approximate solutions are then run through the 500 step genetic optimizer for a second time.
5. The approximate solutions produced by the second run of the genetic optimizer are then fed into the final 200 step MCMC chain, producing the final best-fit solution, an estimate of the reduced χ^2 and the posterior chains.

The genetic optimisation steps are implemented by the `cmaeshpc`[4] package. This package makes use of the covariance matrix adapting evolutionary strategy (CMA-ES), an advanced optimisation algorithm designed to find the best solution to non-linear, complex problems. In each step of the CMA-ES, a best solution is generated based on their fitness while the covariance matrix is updated according to the results. The covariance matrix represents how different priors being fit relate to each other, which allows the algorithm

[3] https://github.com/asgr/highlander
[4] https://github.com/asgr/cmaeshpc

to determine how new solutions should be sampled. This makes CMA-ES adaptable and efficient in optimising highly complex fitting landscapes, such as SED fitting free priors.

The MCMC chain is implemented by the `LaplacesDemon`[5] package, which is set up to utilise the component-wise hit and run metropolis (CHARM) algorithm. The CHARM algorithm effectively creates a parameter space, where the number of dimensions in this space is equal to the number of free priors. In each step, it randomly picks a dimension (or component) and chooses a direction (or 'hit') and 'runs' a certain distance in that dimension, while keeping all other components constant. It then uses a Metropolis criterion to accept this new state and repeats the process. The Metropolis criterion evaluates whether the new state has a higher probability than the current state. The CHARM algorithm breaks down complex, multi-dimensional problems into 1D movements, resulting in efficient exploration of complex parameter spaces.

4.2.5 SED Modelling

By combining all the elements discussed above, we model the SEDs of our galaxy sample using their HI redshifts, multiband photometric measurements, and the outputs of `HIGHLANDER`. As noted above, `HIGHLANDER` determines the best-fit values of the properties derived by ProSpect (e.g., the SED, stellar mass, star formation rate, etc.) and provides the posterior chain for each property from the final 200 step MCMC chain. We found that the best-fit values for ~ 50 of our galaxies ($\sim 37\%$ of our sample) fall outside the 16th-84th percentile range. An example of this case is shown in Figure 4.1. This is problematic as the 16th-84th percentile range is used as uncertainty values. As discussed above, ProSpect was implemented in a similar method to that outlined by Bellstedt et al. (2020a) and Thorne et al. (2021). Both these studies had samples of thousands of galaxies, meaning cuts for poorly constrained SED fits could be made with little to no impact on the final result. Our sample is too small to do this, and so, to utilise the results we do have, we chose to use the median of the posterior chains rather than the best-fit value to model our SEDs. This ensures that the chosen values are within the 16th-84th percentile, avoiding unrealistic, negative uncertainties.

To ensure that this choice would not result in drastically different results, we calculated the z-value of the stellar mass (M_*), SFR, peak star formation rate (mSFR) and the lookback time at peak star formation (mpeak). The z-value (z_{val}) is a statistical measurement which we utilised to quantify the number of standard deviations (σ) between the best-fit value (x, represented by the blue dashed line in Figure 4.1) and the median value (M, represented by the purple dashed line in Figure 4.1). The following equation is used to calculate the z-value:

$$z_{val} = \frac{x - M}{P_{84} - P_{16}} \tag{4.7}$$

where x is the best-fit value, M is the median, P_{84} is the 84th percentile of the final

[5] https://cran.r-project.org/web/packages/LaplacesDemon/index.html

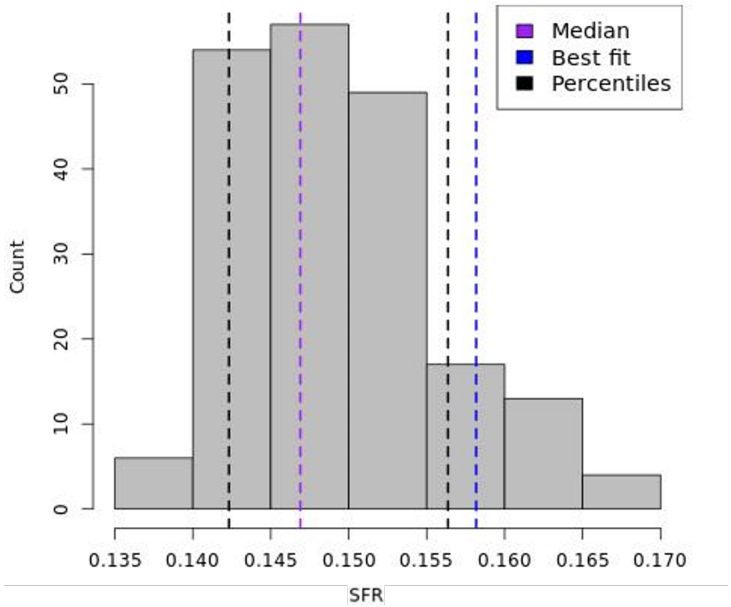


Figure 4.1: The distribution of the SFR posterior chain of a galaxy in our sample. The dashed lines represent the 16th and 84th percentiles (in black), the median of the posterior chain (in purple) and the best-fit value (in blue).

posterior chain and P_{16} is the 16th percentile.

Histograms of the distribution of the z-values for the stellar mass, SFR, mSFR and mpeak for galaxies in our sample are shown in Figure 4.2. After examining the distribution, we find that; 1) M_*, $z_{val} < 1\sigma$, 2) SFR, $z_{val} < 1.2\sigma$, 3) mSFR, $z_{val} < 1.72\sigma$, and 4) mpeak, $z_{val} < 2\sigma$. In other words, the median of the MCMC chain is never more than 2σ away from the best-fit value for all derived properties. This confirms that using the median values to produce the final SEDs of our sample will not result in drastically different results, with a majority of the best-fit values being within the uncertainty of our final fit.

4.3 SED Fitting Results

ProSpect was implemented on all 135 galaxies in our HI sample following the methodology discussed above. Four galaxies (ID 5, 28, 79 and 88) from our HI sample failed during the SED fitting phase. In other words, ProSpect was unable to model the shape of these four galaxies' SEDs, causing the program to terminate. This occurs when the combination of the priors found results in a non-physical or undefined SED model. This was likely caused by inaccurate photometry measurements due to the ProFound segmentation map either 1) not fully encapsulating the entire flux of these galaxies, or 2) the reverse effect, where the segmentation map is too large, therefore, capturing background sky emission and galaxies. These inaccurate photometry measurements, therefore, resulted in inaccurate and nonphysical SED profiles for these galaxies, which were im-

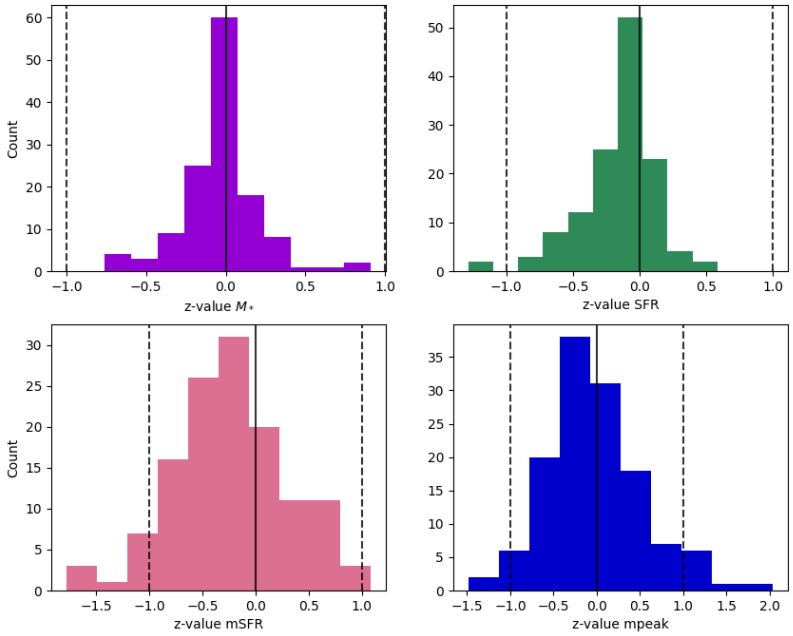


Figure 4.2: Distribution of z-values of (top left) stellar mass, (top right) star formation rate, (bottom left) peak star formation rate and (bottom right) lookback time of mSFR. The solid black line is where $x = M$, the dashed black lines sit at 1σ.

possible to model. For this reason, these four galaxies were excluded from our sample, leaving a total of 131 galaxies.

The SEDs and SFHs of all 131 galaxies are given in Appendix A.3 and are grouped by their reduced χ^2 value. While the vast majority of our SEDs have reasonable fits, some are poorer fits with high reduced χ^2 values. These poor fits are likely caused by a mismatch between the available model templates and the real galaxy's SEDs. This is a result of two things: 1) the confusion of FIR data due to its low resolution, and 2) ProSpect's oversimplification of how dust re-emits light.

The low resolution of FIR data introduces major challenges in SED fitting, as it causes confusion. This is especially true for the *Herschel* long-wavelength data, which has a resolution of 18″-36″. Unlike HI data (which also has low resolution), FIR confusion is caused by high-redshift background sources, which are difficult to detect in even the deepest optical observations. The DEVILS photometry, used in this work, attempts to mitigate this by using 24 and 70 μm data from the *SWIRE* telescope as positional priors to help interpret the FIR measurements. While this does help, it does not fully solve the problem, as FIR emission can come from dust at a wide range of temperatures (which emit at different FIR wavebands). This makes it difficult to accurately model dust, and so, all SED codes (including ProSpect) make use of oversimplified assumptions about how absorbed light is re-emitted in the FIR.

For these reasons, it is common in studies of high-redshift galaxies to implement a reduced χ^2 cut of galaxies where $\chi^2 \geq 10$ [e.g. Bowler et al. 2020, Varadaraj et al. 2023]. We chose to implement this in our work, and did not include the 27 galaxies in our sample that have a reduced $\chi^2 \geq 10$ in our analysis, leaving 104 galaxies.

Figure 4.3 shows examples of the SEDs and star formation histories (SFHs) for two galaxies in our sample (IDs 101 and 180) as output by ProSpect. These galaxies were chosen as examples because they are both relatively well-fit SEDs with photometric measurements in all wavebands. From the SFH plots, we see that ID 101's (shown in the top row) peak star formation rate occurred more recently than ID 180's (shown in the bottom row). As expected, this corresponds to the differences in colour we observe between the two galaxies: ID 101 has experienced more recent star formation, resulting in newer and bluer stars compared to the older and redder stars in ID 180.

4.4 Chapter Summary

In this chapter, to obtain multiwavelength photometry of our HI sample, we cross-match our confirmed SoFiA-2 detected HI catalogue with the DEVILS photometric catalogue. As discussed in Section 2.2, the DEVILS photometric catalogue does not cover the entire LADUMA field. For this reason, our sample was reduced to 135 galaxies after cross-matching. The MIR-FIR coverage of our sample varies due to the lower sensitivities of the images in this waveband. 19% of galaxies in our sample have detections in all MIR-FIR bands, 34% have detections in more than half of the bands, 28% have detections in less than half of the bands, and 19% have no detections in any of the bands. These statistics are summarised in Table 4.1.

To perform SED fitting, we utilise the state-of-the-art SED fitting code ProSpect. ProSpect is implemented such that there are nine free priors and two fixed priors, all of which model SFH, metallicity, and dust. The nine free priors are fit by an MCMC algorithm, which has been implemented by the `HIGHLANDER` R package. After performing the SED modelling, we found that the best-fit values for several of our galaxies fall outside the 16th-84th percentile range. Because of this, we chose to use the median of the posterior chain to model our SEDs rather than the best-fit values. To ensure that this decision does not result in drastically different results, we examined the z-value of the M_*, SFR, mSFR and mpeak. We found that the z-value of all these properties is never greater than 2σ, confirming that using the median value over the best fit value does not drastically change the result and avoids unrealistic, negative uncertainty values.

ProSpect was implemented on all 135 galaxies in our sample. ProSpect was unable to fit four galaxies' SEDs, due to inaccurate photometry measurements. These galaxies were therefore removed from our sample, leaving us with 131 galaxies. 104 of these galaxies have good SED fits, as their reduced $\chi^2 \leq 10$. The remaining 27 galaxies do not have good fits, likely due to the SED template mismatching the data of said galaxies. These

galaxies were therefore excluded from our analysis.

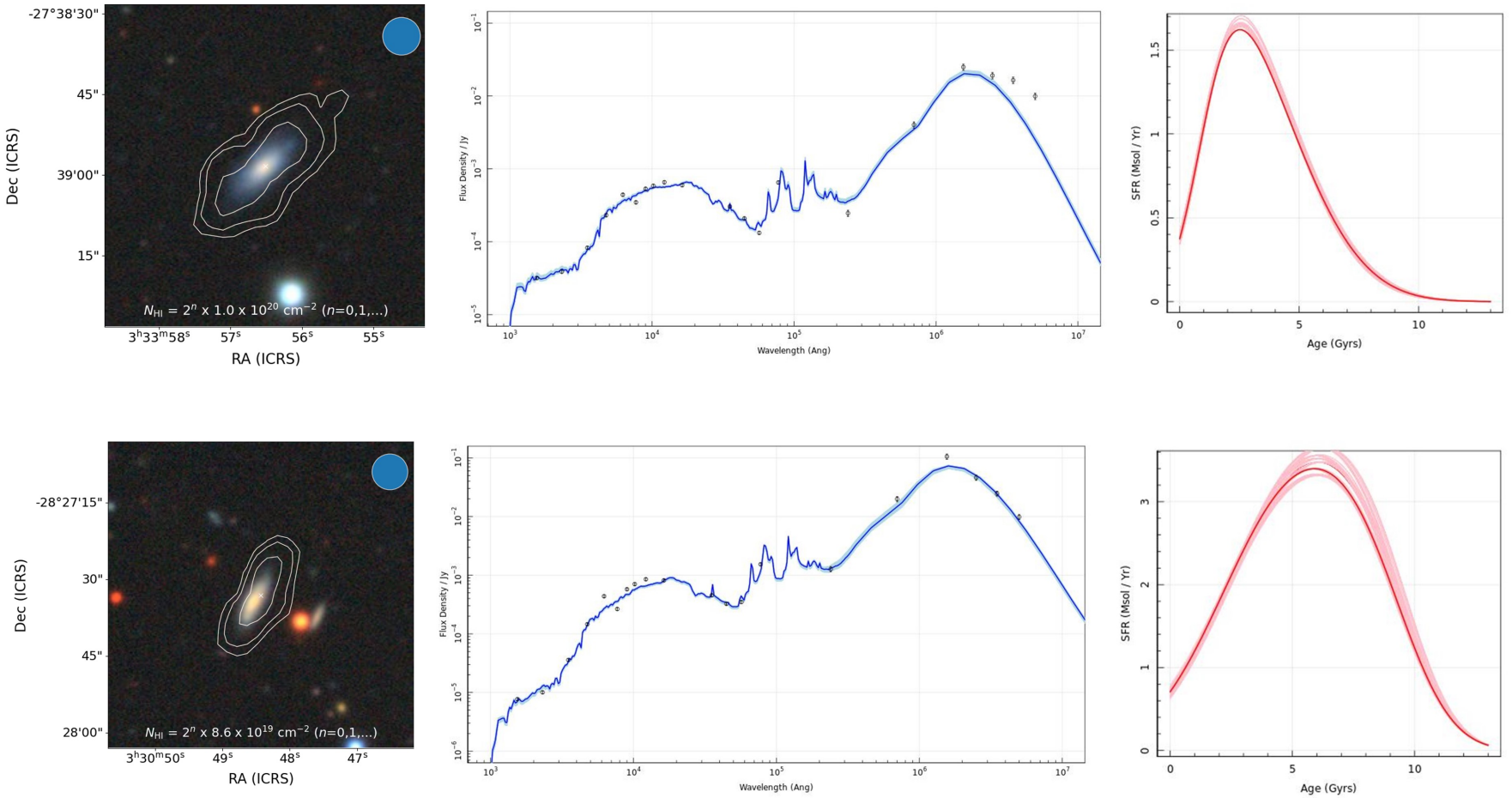


Figure 4.3: Example of the SEDs and SFHs of two galaxies, ID 101 (top row) and ID 180 (bottom row). The left hand panel shows the HI contours overlayed on the DECaLS optical image (produced by the SoFiA Imaging Pipeline). The middle panel shows the SED produced by ProSpect. The dark blue line represents the final SED fit, the light blue lines represent the SEDs generated from each step of the final MCMC chain. The black points show the input photometry (with uncertainties). The left panel is the SFH produced by ProSpect. The red line is the final SFH and the pink lines are the SFH generated from each step of the final MCMC chain.

5 | The Scaling Relations

In this chapter, we will examine the scaling relations using the HI properties obtained via SoFiA-2 and the stellar properties derived via ProSpect. In Section 5.1, we examine the previously well-studied relation of our HI sample of 131 galaxies. In Section 5.2 we examine the relationship between our sample's gas fraction (M_{HI}/M_*) and star formation history (SFH) properties. The full catalogue of all 131 galaxies in our sample, which includes all properties discussed below and their associated uncertainty, is given in Appendix A.1. Note that all figures in this chapter do not include uncertainties of the individual data points, as they are too small to be visible.

5.1 Comparison to low redshift literature scaling relations

Before examining the star formation history properties of our HI sample, we examine its stellar properties, which were derived via SED fitting. In particular, we investigate the stellar mass (M_*) versus the HI mass (M_{HI}) and the star formation rate (SFR) versus M_* scaling relations. These are well-studied relations, meaning that we can use them as diagnostic plots by comparing our relation to those from previous works. This will allow us to judge the robustness of our SED fitting results and their associated properties.

5.1.1 M_{HI} versus M_*

We begin by examining M_{HI} versus M_* of our HI sample. This relation, along with the binned averages of our M_{HI} versus M_* relation, is shown in Figure 5.1. We compare this to the scaling relation of MIGHTEE-HI (at $z < 0.84$; presented by Maddox et al. 2021). We group our data into M_* bins such that each bin contains approximately an equal number of data points, to ensure robust statistics. The median and standard deviation of M_{HI} of each bin are shown by the black points and error bars in Figure 5.1, respectively, and are presented in Table 5.1.

From Figure 5.1, we see that our M_*-M_{HI} relation does agree with the one produced by Maddox et al. (2021), within the 1σ spread, with the median HI mass increasing with M_*. Examining this relation further, we see that our median M_{HI} in each bin is lower than those found by Maddox et al. (2021). This offset is not caused by poorly fit SEDs, since there is no obvious trend between the reduced χ^2 and the M_*-M_{HI} relation (as shown by the colour of the points in Figure 5.1). It is important to note that Maddox et al. (2021) utilised a different SED fitting code (*Le Phare*; Arnouts et al. 1999, Ilbert et al. 2006) than that used in this work. Pacifici et al. (2023) found that M_* values can vary by $0.1 - 0.2$ dex when obtained by different SED fitting codes. However, the M_*-M_{HI} relation presented in this work and that presented in Maddox et al. (2021) differ by ~ 0.7 dex, meaning that the offset observed in Figure 5.1 is not accounted for by the different SED codes. LADUMA is more sensitive than the MIGHTEE survey, and so it is expected that LADUMA would detect lower HI mass galaxies in comparison to MIGHTEE. This

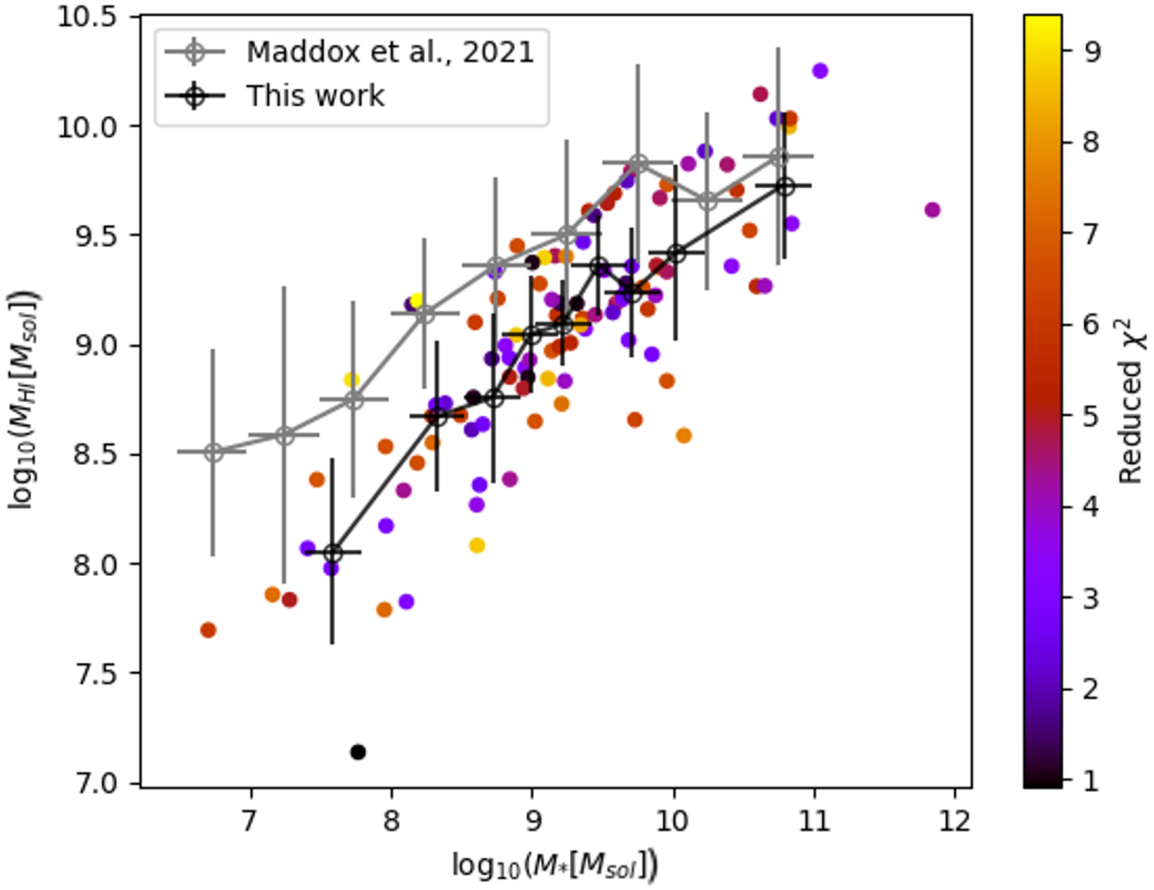


Figure 5.1: M_{HI} as a function of M_* of our HI sample. The points are coloured by the reduced χ^2 of the SED fit. The black circles and bars represent the median and 1σ spread of our sample within each M_* bin, respectively. The grey points and bars represent the same for the MIGHTEE-HI survey [Maddox et al. 2021]. The average uncertainties of $\log_{10}(M_{HI})$ is ± 0.03. The average uncertainty of $\log_{10}(M_*)$ is ± 0.01.

means that we could be detecting green valley and red cloud galaxies, which we do not expect to be sampled in the work of Maddox et al. (2021). Sampling of lower HI mass galaxies has been seen to cause $\sim$ 0.5 dex difference in previous studies [e.g., Parkash et al. 2018].

5.1.2 SFR versus M_*

Figure 5.2 presents the SFR-M_* relation of our HI sample, as well as the binned averages. We compare this relation to the star-forming main sequences (SFMSs) presented in Fraser-McKelvie et al. (2021) (represented by the grey dashed lines in Figure 5.2). Fraser-McKelvie et al. (2021) examined the SFMS of galaxies from the Sydney-Australian Astronomical Observatory Multi-object Integral field spectrograph (SAMI; Croom et al. 2012) galaxy survey and the Mapping Nearby Objects at Apache Point Observatory (MaNGA; Bundy et al. 2015) galaxy survey (note that neither of these surveys are HI surveys, meaning that this is an optically selected sample). Both the SAMI and MaNGA surveys focus

Bin centers	Number of objects	Median $\log_{10}(M_{HI})$	1σ $\log_{10}(M_{HI})$
7.4	12	8.05	0.43
8.35	11	8.67	0.35
8.73	12	8.76	0.39
9	11	9.04	0.27
9.24	12	9.09	0.20
9.49	11	9.36	0.23
9.74	12	9.24	0.30
10.13	11	9.42	0.40
11.12	12	9.73	0.34

Table 5.1: Bin properties of the median stellar mass versus HI mass for all data points (shown as the black points and error bars in Figure 5.1).

on low redshift galaxies, with SAMI at $0.04 < z < 0.128$ and MaNGA at $0.01 < z < 0.15$. For this reason, we found the SFMSs presented in Fraser-McKelvie et al. (2021) to be a good comparison to our galaxy sample, since it is also at low redshift ($0 < z_{HI} < 0.088$).

Fraser-McKelvie et al. (2021) used a linear and curved fit to determine the SFMS of galaxies in their sample. The curved SFMS presented by Fraser-McKelvie et al. (2021) is defined by:

$$\log_{10}(\mathrm{SFR}) = 0.256 - \log_{10}\left(1 + \frac{10^{10.064}}{10^{\mathrm{M}}}\right), \tag{5.1}$$

and the linear SFMS is defined by:

$$\log_{10}(\mathrm{SFR}) = (0.674\mathrm{M}) - 6.836, \tag{5.2}$$

where $M = \log_{10}(M_*)$.

For comparison purposes, we fit a linear relation to our sample:

$$\log_{10}(\mathrm{SFR}) = ((0.78 \pm 0.01)\mathrm{M}) - (8.11 \pm 0.1). \tag{5.3}$$

This fit is represented by the black dashed line and grey shaded region in Figure 5.2. From this figure, we see that the SFR-M_* relation of our HI-selected sample sits below the SFMSs presented by Fraser-McKelvie et al. (2021). Additionally, we do not observe the drop-off on the curved SFMS at $\log_{10}(M_*) > 10.5 M_\odot$. It is also important to note that this is an HI-selected sample, whereas the Fraser-McKelvie et al. (2021) sample is stellar mass selected. It has been found that sample selection significantly impacts the SFR-M_* relation. HI-selected samples are biased toward higher HI masses than stellar mass-selected samples, since HI surveys often lack the sensitivity needed to detect HI-poor systems. These HI-poor systems are ellipticals with stellar masses $> 10^{10} M_\odot$ [Parkash et al. 2018]. This could explain why we do not observe the drop-off depicted by Fraser-McKelvie et al. (2021) curved SFMS. Stellar mass-selected samples include a

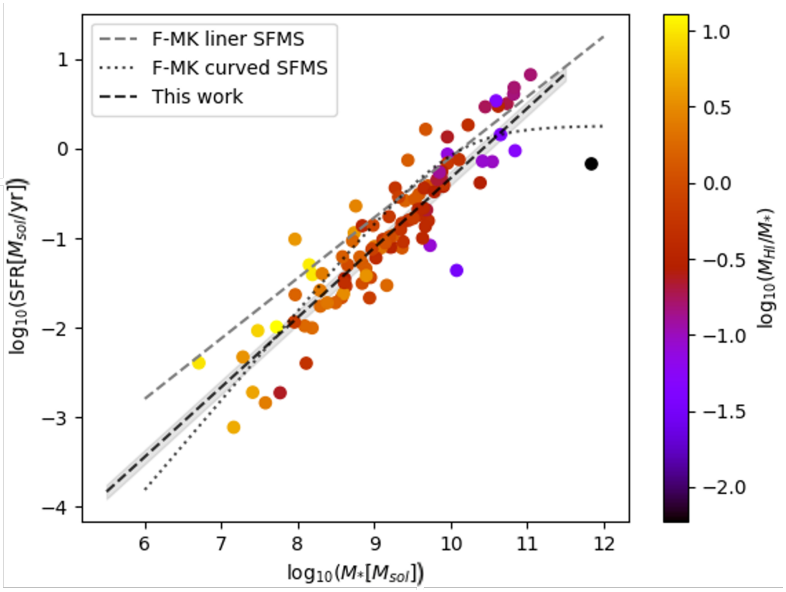


Figure 5.2: SFR-M_* relation of our HI sample represented by the dots, which are coloured by $\log_{10}(M_{HI})$. The black dashed line is the median relation, and the shaded region represents the 1σ spread of our sample. The grey lines represent the two star-forming main sequences (SFMS) presented in Fraser-McKelvie et al. (2021) (F-MK): the dashed line is the linear SFMS and the dotted line is the curved SFMS. The average uncertainty of $\log_{10}$ SFR is ± 0.02. The average uncertainty of $\log_{10}(M_*)$ is ± 0.003.

wider range of SFRs, which can lead to different interpretations of the SFR-M_* relation, since higher stellar mass galaxies (which are HI-poor) tend to have higher SFRs [Bauer et al. 2011]. Additionally, the low redshift LADUMA field is not a good representation of the local Universe, as this field was selected to avoid nearby objects to avoid foreground contamination. These elements could explain the offset observed between our SFMS and the linear SFMS presented by Fraser-McKelvie et al. (2021).

We observe that galaxies with a higher SFR and higher M_* have baryonic masses that are dominated by HI. The M_{HI}/M_* of galaxies subsequently decreases as SFR and M_* increase. This is shown by the coloured points in Figure 5.2. This suggests that galaxies whose baryonic mass is dominated by HI have higher SFR, which are likely being fueled by their large HI reservoirs.

5.2 The star formation histories of HI galaxies

With the above in mind, we examine the star formation history (SFH) properties of our HI sample. The SFH reflects the timeline of a galaxy's star formation activity. In other words, we can examine the star-forming properties of said galaxy as it evolves over time. Two useful properties of the SFH are the peak star formation rate of a galaxy's history (referred to as $\mathrm{SFR_{peak}}$) and the lookback time at that peak star formation (referred to as τ_{peak}). We therefore examine how these properties relate to the HI gas mass to stellar

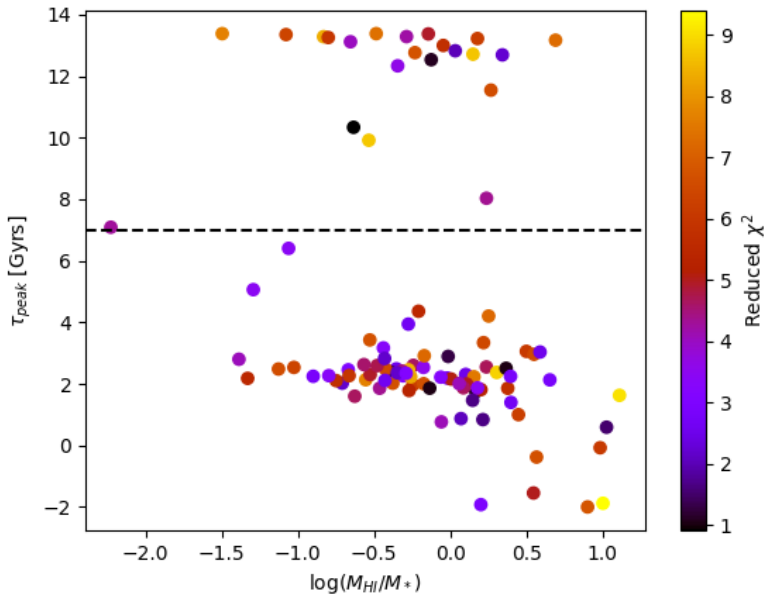


Figure 5.3: τ_{peak} vs M_{HI}/M_* of our galaxy sample. Colour scale indicates the reduced χ^2 of the SED fits. All points above the horizontal dashed line (at $\tau_{peak} = 7$ Gyr) are ignored in Figure 5.4. The average uncertainty of $\log_{10}(\tau_{peak})$ is $^{+0.14}_{-0.25}$. The average uncertainty of $\log_{10}(M_{HI}/M_*)$ is ± 0.02.

mass fraction (M_{HI}/M_*) to determine how the HI reservoirs of galaxies impact their star formation history.

5.2.1 τ_{peak} versus M_{HI}/M_*

In Figure 5.3 we present τ_{peak} versus M_{HI}/M_* of this work. We find that 22 galaxies in our sample have 12 Gyrs$<\tau_{peak}<$13.38 Gyrs. This does not agree with Tudorache et al. (2024), who investigated the same relation for the MIGHTEE-HI galaxies at $0.004 < z_{HI} < 0.084$, a similar redshift range examined in this work ($0 < z_{HI} < 0.088$). It is unlikely that the τ_{peak} of these galaxies is physical, as this would imply that these galaxies' star formation rate peaked close to the beginning of the Universe. From the colouring of the points in Figure 5.3, we see that this is not caused by poor SED fits, as these 22 galaxies have varying reduced χ^2 measurements. This could be caused by two things: 1) these galaxies have experienced multiple bursts of star formation throughout their history, or 2) the SFHs of these galaxies do not have a well-defined peak and have formed stars more gradually. In both these cases, the function we chose to model SFH(`massfunc_snorm_trunc`, discussed in Section 4.2.1) would not accurately represent this, as it assumes a galaxy has a singular, well-defined peak. The SFHs of these galaxies are therefore inaccurate, with meaningless τ_{peak} values.

For this reason, we chose to exclude these galaxies from further analysis and examine the 82 galaxies with $\tau_{peak} < 7$ Gyr. This allows us to compare with the relation reported in Tudorache et al. (2024). This is shown in Figure 5.4. To calculate the median values,

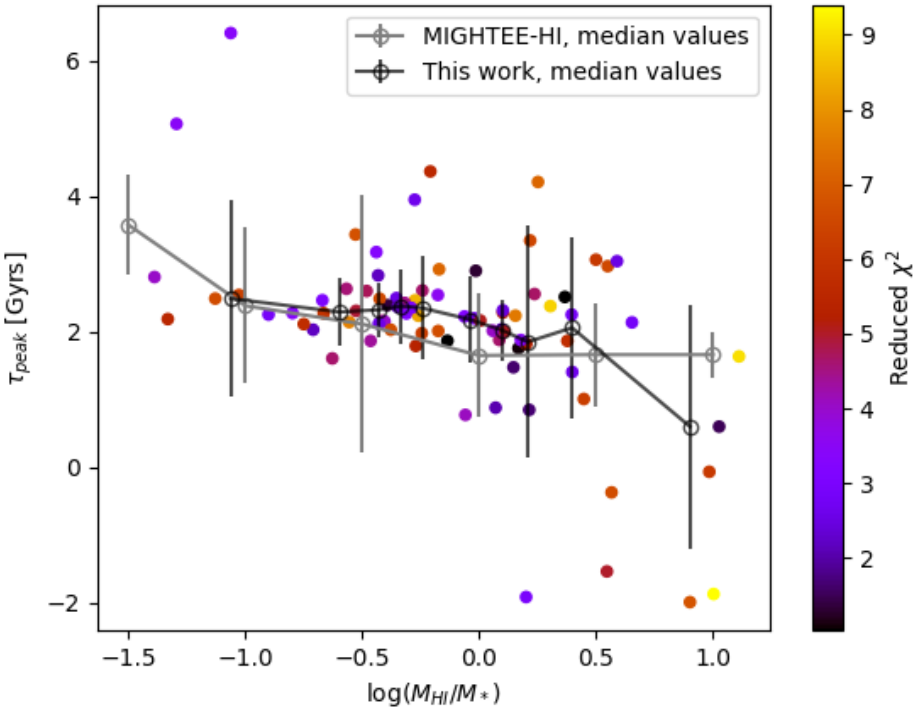


Figure 5.4: τ_{peak} vs M_{HI}/M_* of our galaxy sample where $\tau_{peak} < 7$ Gyrs. The black points and error bars represent the median relation and 1σ spread of our sample. The grey points represent the MIGHTEE-HI median relation [Tudorache et al. 2024]. Colour scale indicates the reduced χ^2 of the SED fits.

we group our data into M_{HI}/M_* bins such that each bin has approximately equal data points, to ensure robust statistics, and calculate the median and standard deviation of τ_{peak} in each bin. These values are given in Table 5.2 and are represented by the black points and error bars shown in Figure 5.4, respectively. From Figure 5.4, we see the τ_{peak} vs M_{HI}/M_* median derived from our HI sample agrees with that reported by Tudorache et al. (2024).

5.2.2 $\mathrm{SFR_{peak}}$ versus M_{HI}/M_*

To further investigate the correlation between galaxies' HI properties and their star formation histories, we examine the relation between the peak star formation rate ($\mathrm{SFR_{peak}}$) and gas fraction (M_{HI}/M_*). This is shown in Figure 5.5 and, to the best of our knowledge, is studied here for the first time. From this figure, we see a strong anti-correlation between $\mathrm{SFR_{peak}}$ and M_{HI}/M_*, which has the linear relation of:

$$\log_{10}(\mathrm{SFR_{peak}}) = -(0.20 \pm 0.01)\mathrm{M} - 0.71 \pm 0.1 \tag{5.4}$$

Bin centers	Number of objects	Median τ_{peak} [Gyr]	1σ τ_{peak} [Gyr]
-1.07	9	2.48	1.45
-0.63	8	2.29	0.5
-0.46	8	2.32	0.41
-0.34	8	2.37	0.55
-0.22	8	2.34	0.76
-0.05	8	2.18	0.63
0.12	8	2.01	0.46
0.23	8	1.84	1.71
0.42	8	2.05	1.34
0.83	9	0.59	1.8

Table 5.2: Bin properties of the median values of τ_{peak} versus M_{HI}/M_* (represented by the black points and error bars in Figure 5.4).

where M = $\log_{10}(\mathrm{M_{HI}}/\mathrm{M_*})$. This fit line is represented by the black dashed line in Figure 5.5.

The anti-correlation observed between $\mathrm{SFR_{peak}}$ and M_{HI}/M_* implies that galaxies with a stellar-dominated mass composition have experienced a more aggressive star formation period in their history compared to galaxies whose baryonic mass is dominated by HI. This is somewhat expected, as galaxies with a lower M_{HI}/M_* value are likely to have formed a significant amount of stars over their lifetime to account for their high M_*. Additionally, their smaller HI reservoirs can be attributed to the depletion of gas during these heightened periods of star formation.

In the case of galaxies with higher M_{HI}/M_* values, likely, these galaxies have not yet reached their peak star formation rate. This is further emphasised by the τ_{peak} values decreasing towards larger M_{HI}/M_* values. We see that the majority of galaxies with $\log_{10}(M_{HI}/\mathrm{M_*}) > 0.5$ have a $\tau_{peak} < 1$ Gyrs. This reiterates that galaxies whose mass is dominated by HI are likely either currently experiencing, or have yet to experience, their peak star formation and will undergo a more aggressive star formation period later in their evolution, which will be fueled by their large HI reservoirs.

5.3 Chapter Summary

Before examining the SFH properties of our HI sample, we investigate the M_{HI}-M_* relation and SFR-M_* relation. Doing this allows us to judge the robustness of our SED fitting results and their associated properties, since these are well-studied relations. We find that the M_{HI}-M_* relation agrees with the one presented in Maddox et al. (2021), within 1σ spread. However, we observe a ~ 0.7 dex offset between the relation presented in this work and the MIGHTEE relation. This is likely caused by 1) differing SED codes, which has been seen to produce an offset of 0.1-0.2 dex in M_* measurements, and 2) LADUMA being more sensitive than MIGHTEE, meaning that lower HI mass galaxies are detected by LADUMA in comparison to MIGHTEE. This means that the LADUMA

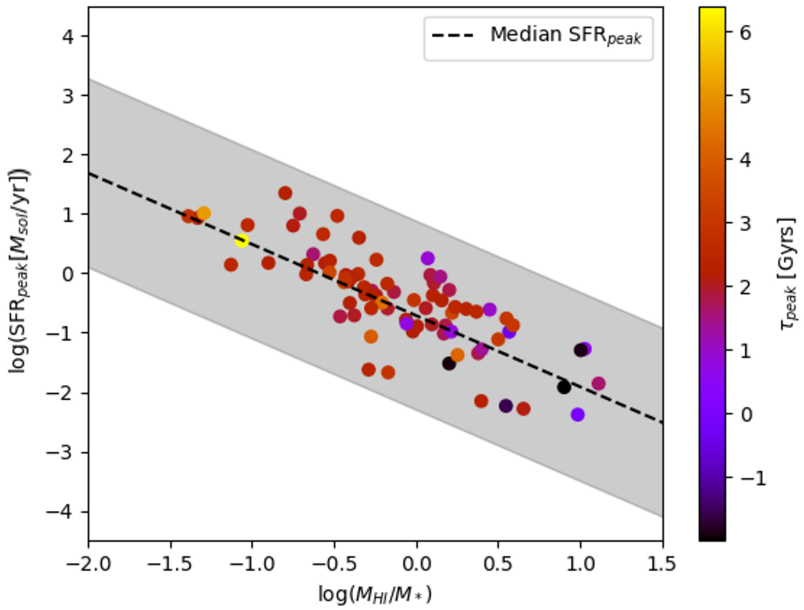


Figure 5.5: $\mathrm{SFR_{peak}}$ versus M_{HI}/M_* of our galaxy sample with $\tau_{peak} < 7$ Gyrs. Colour scale indicated τ_{peak}. The black dashed line and shaded region show the median relation and 1σ spread of our sample, respectively. The average uncertainty of $\log_{10}(\mathrm{SFR_{peak}})$ is ± 0.01.. The average uncertainty of $\log_{10}(M_{HI}/M_*)$ is ± 0.003.

sample could be detecting green valley and red cloud galaxies, which we do not expect to be sampled by the MIGHTEE survey.

Examining the SFR-M_* relation, we found that the SFMS produced from our sample sits below the SFMS presented by Fraser-McKelvie et al. (2021). Additionally, we do not observe the drop-off shown on the curved SFMS at $\log_{10}(M_*) > 10.5\ M_\odot$. This is likely due to the sampling difference, with our sample being HI-selected and the Fraser-McKelvie et al. (2021) sample being stellar mass-selected. This results in our sample being biased to HI-rich galaxies, excluding HI-poor, stellar mass-heavy, high star-forming galaxies, which are sampled by the stellar mass-selected sample. We observe a decrease in M_{HI}/M_* as SFR and M_* increase. This suggests that galaxies whose baryonic mass is dominated by HI have higher SFRs, which are likely being fueled by their large HI reservoirs.

To investigate the SFHs properties of our HI selected sample, we examine the τ_{peak}-M_{HI}/M_* relation and $\mathrm{SFR_{peak}}$-M_{HI}/M_* relation, which, to the best of our knowledge, is studied here for the first time. Looking at the τ_{peak}-M_{HI}/M_* relation, we find that 22 galaxies in our sample peak at 12 Gyrs$<\tau_{peak}<$13.38 Gyrs. This is unlikely to be physical and is likely caused by galaxies that have: 1) experienced multiple bursts of star formation through their history, or 2) formed stars gradually and, therefore, do not have a well-defined τ_{peak} value. Since the SFHs of these galaxies are likely modelled inaccurately, we chose to exclude these galaxies from further analysis and examine the 82 galaxies with $\tau_{peak} < 7$. With this limitation, the τ_{peak}-M_{HI}/M_* relation of our sample agrees with that

presented in Tudorache et al. (2024).

We find a strong anti-correlation between $\mathrm{SFR}_{\mathrm{peak}}$ and M_{HI}/M_*. This implies that galaxies with stellar-dominated mass compositions have experienced a more aggressive star formation period in their history compared to galaxies whose baryonic mass is dominated by HI. This would also explain their smaller HI reservoirs, as their gas was depleted during these aggressive star formation periods. This is further emphasised by the lower τ_{peak} values at higher M_{HI}/M_*, implying that galaxies with an HI-dominated baryonic mass have yet to experience their peak star formation and will undergo a more aggressive star formation period later in their evolution, which will be fueled by their HI reservoirs.

6 | Summary and Future Work

6.1 Summary

In this study, we have investigated the star formation history properties of low-redshift ($0 < z_{HI} < 0.088$) HI galaxies detected by the LADUMA survey. LADUMA is an ongoing ultra-deep and narrow extragalactic HI emission survey with a sky area coverage of 0.8 deg^2 (at 1420 MHz) in the ECDFS field. We began our exploration by performing automated HI source finding, with the SoFiA-2 algorithm, on the high-frequency (1304-1420 MHz) early science LADUMA data cube. This involved identifying input parameters for SoFiA-2 that produced an HI source catalogue of optimal reliability and completeness when compared to catalogues produced by manual and semi-automated source finding methods. Our implementation of SoFiA-2 generated an output catalogue of 232 candidate sources.

We inspected optical images of these candidates and identified counterparts for 83% of our sample. We consider the 39 sources without counterparts to be false detections. Next, we compared the remaining 193 sources with source catalogues produced by manual and semi-automated techniques. We found that SoFiA-2 identified 87% of the sources found via manual source finding and 86% of the sources found via matched filtering. Through this comparison, we found that our final HI catalogue of 193 sources has an estimated "completeness" of 82%.

Once we obtained our HI sample, we cross-matched our SoFiA-2 detected sample with the DEVILS multiband photometric catalogue. Due to the partial coverage of the DEVILS catalogue over the LADUMA field, our sample was reduced to 135 galaxies after cross-matching. Using the multiband photometry and HI redshift of our sample, we performed SED fitting with ProSpect to obtain their stellar masses (M_*), star formation rates (SFRs) and star formation histories (SFHs). The SED fitting of four galaxies failed, likely due to inaccurate photometry measurements. These galaxies were therefore removed from our sample, leaving 131 galaxies. 104 of these galaxies have well-fit SEDs, as their reduced $\chi^2 \leq 10$. The remaining 27 galaxies are, therefore, bad fits that were likely due to the SED templates mismatching the data of these galaxies. For this reason, these galaxies were excluded from further analysis.

To judge the robustness of our SED fits and their associated properties, we compare the HI mass (M_{HI}), SFR and M_* of the 104 galaxies in our sample to relations from the literature. We found that M_{HI}-M_* and M_*-SFR relations produced with our data are in agreement with those from the literature, and demonstrated that the LADUMA survey is, by design, probing to lower HI masses than the MIGHTEE-HI survey. Additionally, we found that galaxies with higher SFRs have HI-dominated baryonic mass compositions, suggesting that their SFRs are being fueled by their large HI reservoirs.

To further examine the correlation between galaxies' HI properties and their star formation properties, we examined their star formation history properties. We first investigated the correlation between the lookback time of peak star formation history (τ_{peak}) and gas fraction (M_{HI}/M_*). We found that 22 galaxies in our sample have 12 Gyrs$<\tau_{peak}<$13.38 Gyrs, which is unlikely to be physical. These galaxies likely have: 1) multiple bursts of star formation throughout there SFH, or 2) formed stars more gradually and, therefore, do not have well-defined τ_{peak} values. Since the SFHs of these galaxies are likely inaccurate, they are excluded from further analysis. We compared the remaining 82 galaxies with $\tau_{peak} < 7$Gyrs to a similar sample studied in the MIGHTEE-HI survey. We find that our τ_{peak}-M_{HI}/M_* relation is in broad agreement with that found for MIGHTEE-HI.

Next, we examined the relation between the peak star formation rate ($\mathrm{SFR_{peak}}$) and the gas fraction (M_{HI}/M_*). To the best of our knowledge, this is the first time such a relation has been examined. We found a strong anti-correlation between $\mathrm{SFR_{peak}}$ and M_{HI}/M_*. This suggests galaxies with lower M_{HI}/M_* have experienced more aggressive periods of star formation in their history. Additionally, we found that galaxies with higher M_{HI}/M_* have lower τ_{peak}. This implies that galaxies with an HI-dominated baryonic mass have likely not yet experienced their peak star formation period. These galaxies will likely experience a much more aggressive star-forming period as they evolve, which will be fueled by their present HI content.

6.2 Future work

The first public release of the LADUMA data is imminent and will include dirty cubes, clean cubes and the masks used for cleaning, along with a comprehensive source catalogue compiled using multiple different source finding methods. The source finding mask of the low-redshift data cube created in this work will be used in the cleaning process of DR1. This will be discussed in detail in a forthcoming paper [Baker er al., in prep].

Our work was limited to low-redshift galaxies, as cleaning and continuum subtraction were not yet complete in the high-redshift spectral windows LADUMA data cubes at the time. For future work, our study can be extended to higher redshifts, which can be compared to our $z_{HI} \sim 0$ results to identify any observed evolution. This, however, will introduce challenges for source finding, since high redshift HI galaxies have significantly lower SNRs than those of low redshift galaxies. To mitigate these challenges, these low-frequency cubes must undergo thorough cleaning and continuum subtraction. Since we find that SoFiA-2 performs poorly on dirty cubes, we suggest first cleaning with masks created by guided source finding (where sources are searched for at the position of already known galaxies) before implementing blind source finding. Additionally, the implementation of SoFiA-X for source finding may assist in efficiently finding the extremely faint sources that exist at higher redshifts. SoFiA-X is a parallelised extension of SoFiA-2, which performs source finding over subsets of large data cubes and compiles the results into one final catalogue. By performing source finding on these smaller sub-cubes, SoFiA-X becomes more sensitive to small, faint sources that are hidden among the noise.

The DEVILS multiband photometric catalogue used in this work does not cover the entire LADUMA field. Along with this, higher sensitivity imaging of the ECDFS field now exists, such as the data from the Hyper Suprime-Cam (HSC; Aihara et al. 2018). Our work can therefore be extended by compiling a new photometric catalogue, using similar techniques implemented by Davies et al. (2021) with ProFound on newer, more sensitive multiwavelength images which cover more of the LADUMA field.

To further improve the multiwavelength results of HI galaxies, the quality of the SED fits can be improved by investigating different initial priors of ProSpect to find those that produce higher-quality SED fits on an HI sample, since this SED code has previously only been used on a stellar-mass selected sample. This would include analysis of the MCMC chains and aperture cross-checking. Additionally, testing other SED fitting softwares on this sample may be beneficial, as each program has different intricacies which could be beneficial in different scientific cases. Along with this, stellar mass limits can be calculated for the HI source candidates that have no optical counterparts. By also including the investigation of dark HI galaxies, the final sample and diversity of galaxies studied widens, broadening the scope of this work.

A | Appendix

A.1 SoFiA-2 parameter file

Below we provide the SoFiA-2 parameter file used to produce the final HI source catalogue. SoFiA-2 was run on the ilifu cluster for this work. Each run took ~ 45 minutes.

```
# Global settings

pipeline.verbose            =  false
pipeline.pedantic           =  false
pipeline.threads            =  0

# Input

input.data                  =
/idia/projects/laduma/DR1.1/cleaned/laduma_dr1.1_image.1304~1420MHz.clean.fits
input.region                =
input.gain                  =
input.noise                 =
input.weights               =
input.mask                  =
input.invert                =  false

# Flagging

flag.region                 =  3533,3578,974,1016,0,1102
flag.catalog                =
flag.radius                 =  20
flag.auto                   =  false
flag.threshold              =  5.0
flag.log                    =  false

# Continuum subtraction

contsub.enable              =  false
contsub.order               =  0
contsub.threshold           =  2.0
contsub.shift               =  4
contsub.padding             =  3

# Noise scaling

scaleNoise.enable           =  false
scaleNoise.mode             =  spectral
```

```
scaleNoise.statistic       =  mad
scaleNoise.fluxRange       =  negative
scaleNoise.windowXY        =  25
scaleNoise.windowZ         =  100
scaleNoise.gridXY          =  0
scaleNoise.gridZ           =  0
scaleNoise.interpolate     =  false
scaleNoise.scfind          =  false

# Ripple filter

rippleFilter.enable        =  true
rippleFilter.statistic     =  median
rippleFilter.windowXY      =  31
rippleFilter.windowZ       =  15
rippleFilter.gridXY        =  0
rippleFilter.gridZ         =  0
rippleFilter.interpolate   =  false

# S+C finder

scfind.enable              =  true
scfind.kernelsXY           =  0, 3, 5
scfind.kernelsZ            =  0, 3, 5, 7
scfind.threshold           =  4.0
scfind.replacement         =  2.0
scfind.statistic           =  mad
scfind.fluxRange           =  negative

# Threshold finder

threshold.enable           =  false
threshold.threshold        =  5.0
threshold.mode             =  relative
threshold.statistic        =  mad
threshold.fluxRange        =  negative

# Linker

linker.enable              =  true
linker.radiusXY            =  3
linker.radiusZ             =  2
linker.minSizeXY           =  5
linker.minSizeZ            =  5
linker.maxSizeXY           =  0
```

```
linker.maxSizeZ            =  0
linker.minPixels           =  0
linker.maxPixels           =  0
linker.minFill             =  0
linker.maxFill             =  0
linker.positivity          =  false
linker.keepNegative        =  false

# Reliability

reliability.enable         =  true
reliability.parameters     =  peak, sum, mean
reliability.threshold      =  0.5
reliability.scaleKernel    =  0.25
reliability.minSNR         =  12.0
reliability.minPixels      =  0
reliability.catalog        =
reliability.plot           =  true
reliability.debug          =  false

# Mask dilation

dilation.enable            =  false
dilation.iterationsXY      =  10
dilation.iterationsZ       =  5
dilation.threshold         =  0.001

# Parameterisation

parameter.enable           =  true
parameter.wcs              =  true
parameter.physical         =  false
parameter.prefix           =  SoFiA
parameter.offset           =  false

# Output

output.directory           =  /idia/users/leyya/sofia/outFeb23
output.filename            =  2023Feb_real_cleaned_dr1.1_8k_4t_1304-1420MHz_r_v20
output.writeCatASCII       =  false
output.writeCatXML         =  true
output.writeCatSQL         =  false
output.writeNoise          =  false
output.writeFiltered       =  false
output.writeMask           =  true
```

```
output.writeMask2d          =  false
output.writeRawMask         =  false
output.writeMoments         =  true
output.writeCubelets        =  true
output.marginCubelets       =  10
output.thresholdMom12       =  0.0
output.overwrite            =  false
```

A.2 Catalogue of final galaxy sample

The catalogue of our final sample of 131 galaxies is given Table A.1. These are the confirmed HI galaxies, detected by SoFiA-2 (i.e. with optical counterparts), that had multiband photometric coverage (in the DEVILS photometric catalogue), and successfully ran on ProSpect to produce SED fits, star formation histories, stellar masses and star formation rates.

The table contains the following columns:
Column 1: The SoFiA identification number.
Column 2: The HI Right Ascension (RA; produced by SoFiA-2)
Column 3: The HI Declination (Dec; produced by SoFiA-2)
Column 4: The HI redshift (z_{HI}; produced by SoFiA-2)
Column 5: The logarithmic value of the HI mass and uncertainty ($\log(M_{HI})$; produced by ProSpect)
Column 7: The logarithmic value of the stellar masses and uncertainty ($\log(M_*)$; produced by ProSpect)
Column 8: The logarithmic value of the peak star formation rate of the star formation history and uncertainty ($\log(\mathrm{SFR}_{\mathrm{peak}})$; produced by ProSpect)
Column 9: The lookback time of the peak star formation rate and uncertainty (τ_{peak}; produced by ProSpect
Column 10: The reduced χ^2 of the SED fit (produced by ProSpect)

It is important to note that the HI mass uncertainties stated in this catalogue are approximations only. To calculate the HI mass uncertainty, we co-added the error caused by statistical error, which is caused by stochastic noise, and a systematic error of 5 percent for higher mass galaxies ($M_{HI} > 10^8$) and 20 percent for the lower mass objects.

The statistical flux uncertainty caused by stochastic noise (σ) is outputted as part of the SoFiA-2 source finding catalogue and is calculated with the following equation (reproduced from the SoFiA-2 manual [T.Westmeier[1]]):

$$\sigma = \sqrt{\frac{N_{pix}}{\Omega_{PSF}}} \Delta z \frac{S_{sum}}{\sigma_{rms}} \tag{A.1}$$

[1] https://gitlab.com/SoFiA-Admin/SoFiA-2/-/wikis/documents/SoFiA-2_User_Manual.pdf

where S_{sum} is the integrated flux, N_{pix} is the total number of spatial and spectral pixels that make up the detections, Ω_{PSF} is the solid angle of the beam (in pixels), Δz the total spectral width of the source and σ_{rms} is the global RMS noise level of the data cube.

This statistical uncertainty is small in comparison to that contributed by systematic errors. These are errors introduced by calibration and continuum subtraction. To calculate statistical error accurately, one could employ the empirical approach of placing the source mask at different emission-free positions around the detection and use the mean RMS scatter of these flux measurements as the statistical flux uncertainty. However, this is beyond the scope of this project. Instead, we used the approximate error percentage found in Ponomareva et al. (2023), which utilised the method outlined above. Ponomareva et al. (2023) found galaxies with a ~ 5 percent error for higher mass galaxies ($M_{HI} > 10^8 M_\odot$) and $\sim$ 20 percent error for lower mass galaxies ($M_{HI} \leq 10^8 M_\odot$), which we applied to calculate the HI uncertainty of our sample.

Table A.1: Catalogue of final sample of HI galaxies, detected by SoFiA-2 (with optical counter parts), that had multiband photometric coverage (in the DEVILS photometric catalogue), and successfully ran on ProSpect

ID	RA	Dec	z_{HI}	$\log(M_{HI})$ $[M_\odot]$	$\log(\mathrm{SFR})$ $[M_\odot/\mathrm{yr}]$	$\log(M_*)$ $[M_\odot]$	$\log(\mathrm{SFR}_{\mathrm{peak}})$ $[M_\odot/\mathrm{yr}]$	τ_{peak} [Gyr]	χ^2
1	53.22	-28.09	0.00	7.14 ± 0.09	$-2.72^{+0.02}_{-0.02}$	$7.77^{+0.02}_{-0.02}$	$-2.11^{+0.02}_{-0.02}$	$10.34^{+0.43}_{0.61}$	0.90
2	53.29	-28.16	0.00	7.09 ± 0.09	$-3.95^{+0.01}_{-0.03}$	$6.77^{+0.01}_{-0.01}$	$-3.0^{+0.01}_{-0.01}$	$13.38^{+0.0}_{-0.01}$	76.34
3	53.01	-28.36	0.00	7.79 ± 0.09	$-1.94^{+0.02}_{-0.02}$	$7.96^{+0.01}_{-0.02}$	$-1.67^{+0.01}_{-0.01}$	$2.92^{+0.16}_{-0.14}$	7.21
4	52.88	-28.43	0.01	8.07 ± 0.02	$-2.72^{+0.03}_{-0.02}$	$7.41^{+0.02}_{-0.01}$	$-2.28^{+0.01}_{-0.01}$	$2.13^{+0.04}_{-0.09}$	2.81
6	53.54	-27.86	0.01	7.98 ± 0.09	$-2.84^{+0.03}_{-0.07}$	$7.58^{+0.02}_{-0.03}$	$-2.15^{+0.02}_{-0.03}$	$2.25^{+0.12}_{-0.06}$	2.82
7	52.32	-27.66	0.01	8.95 ± 0.02	$-0.26^{+0.02}_{-0.02}$	$9.85^{+0.01}_{-0.01}$	$0.17^{+0.01}_{-0.01}$	$2.25^{+0.03}_{-0.1}$	2.83
11	53.50	-28.45	0.02	7.69 ± 0.09	$-2.39^{+0.01}_{-0.01}$	$6.71^{+0.01}_{-0.02}$	$-2.38^{+0.01}_{-0.01}$	$-0.07^{+0.08}_{-0.07}$	6.10
12	53.22	-28.41	0.02	7.82 ± 0.09	$-2.4^{+0.02}_{-0.02}$	$8.11^{+0.01}_{-0.01}$	$-1.62^{+0.01}_{-0.01}$	$2.35^{+0.03}_{-0.02}$	3.11
14	53.79	-27.96	0.03	8.27 ± 0.03	$-1.5^{+0.01}_{-0.01}$	$8.61^{+0.01}_{-0.03}$	$-1.49^{+0.01}_{-0.01}$	$12.33^{+0.89}_{-2.84}$	3.60
15	53.76	-28.02	0.03	7.86 ± 0.1	$-3.11^{+0.01}_{-0.01}$	$7.16^{+0.01}_{-0.02}$	$-2.8^{+0.03}_{-0.03}$	$13.17^{+0.21}_{-0.76}$	7.25
18	53.15	-27.93	0.04	9.68 ± 0.02	$-0.69^{+0.03}_{-0.04}$	$10.63^{+0.01}_{-0.02}$	$0.84^{+0.02}_{-0.03}$	$13.38^{+0.0}_{-0.1}$	15.20
19	52.34	-28.13	0.04	10.25 ± 0.02	$0.83^{+0.01}_{-0.02}$	$11.05^{+0.01}_{-0.01}$	$1.35^{+0.01}_{-0.01}$	$2.27^{+0.03}_{-0.04}$	3.35
20	53.05	-27.82	0.04	7.55 ± 0.1	$-2.04^{+0.01}_{-0.02}$	$6.68^{+0.04}_{-0.02}$	$-1.83^{+0.01}_{-0.03}$	$-1.06^{+0.01}_{-0.11}$	25.01
21	53.22	-27.95	0.04	7.83 ± 0.09	$-2.33^{+0.01}_{-0.01}$	$7.28^{+0.03}_{-0.02}$	$-2.23^{+0.01}_{-0.01}$	$-1.54^{+0.16}_{-0.46}$	5.00
22	53.63	-28.01	0.04	8.8 ± 0.02	$-1.67^{+0.01}_{-0.01}$	$8.94^{+0.01}_{-0.01}$	$-0.87^{+0.01}_{-0.01}$	$13.37^{+0.01}_{-0.23}$	4.94

Continued on the next page...

ID	RA [°]	Dec [°]	z_{HI}	$\log(M_{HI})$ $[M_\odot]$	$\log(\mathrm{SFR})$ $[M_\odot/\mathrm{yr}]$	$\log(M_*)$ $[M_\odot]$	$\log(\mathrm{SFR_{peak}})$ $[M_\odot/\mathrm{yr}]$	τ_{peak} [Gyr]	χ^2
23	53.11	-27.98	0.04	8.58 ± 0.02	$-1.36^{+0.01}_{-0.02}$	$10.08^{+0.01}_{-0.01}$	$0.5^{+0.01}_{-0.01}$	$13.38^{+0.0}_{-0.01}$	7.66
24	53.51	-27.86	0.04	8.93 ± 0.02	$-1.03^{+0.01}_{-0.01}$	$8.72^{+0.04}_{-0.02}$	$-0.98^{+0.01}_{-0.01}$	$0.84^{+0.54}_{-0.13}$	1.66
25	53.10	-28.04	0.04	8.46 ± 0.02	$-1.92^{+0.01}_{-0.02}$	$8.35^{+0.01}_{-0.02}$	$-1.39^{+0.01}_{-0.01}$	$2.07^{+0.05}_{-0.03}$	23.96
26	52.38	-28.16	0.04	9.34 ± 0.02	$-0.46^{+0.01}_{-0.02}$	$10.04^{+0.01}_{-0.01}$	$0.3^{+0.01}_{-0.01}$	$2.34^{+0.03}_{-0.01}$	11.14
27	53.96	-27.65	0.04	8.9 ± 0.03	$-3.75^{+0.02}_{-0.01}$	$6.78^{+0.01}_{-0.01}$	$-3.0^{+0.01}_{-0.01}$	$13.38^{+0.0}_{-0.03}$	31.67
30	53.82	-28.43	0.04	9.35 ± 0.02	$-0.42^{+0.03}_{-0.02}$	$9.9^{+0.01}_{-0.01}$	$0.17^{+0.01}_{-0.01}$	$2.14^{+0.04}_{-0.04}$	7.68
32	53.94	-27.95	0.04	9.13 ± 0.02	$-1.04^{+0.01}_{-0.01}$	$9.18^{+0.01}_{-0.01}$	$-0.79^{+0.01}_{-0.01}$	$13.0^{+0.33}_{-0.76}$	5.74
34	53.77	-28.40	0.04	9.36 ± 0.02	$-0.8^{+0.01}_{-0.03}$	$9.71^{+0.01}_{-0.01}$	$-0.01^{+0.02}_{-0.01}$	$2.49^{+0.02}_{-0.02}$	2.55
35	52.82	-27.88	0.04	9.21 ± 0.02	$-0.64^{+0.01}_{-0.02}$	$8.76^{+0.03}_{-0.02}$	$-0.61^{+0.01}_{-0.01}$	$1.0^{+0.14}_{-0.19}$	6.24
36	53.03	-28.01	0.04	8.08 ± 0.03	$-1.45^{+0.01}_{-0.01}$	$8.61^{+0.02}_{-0.03}$	$-1.44^{+0.01}_{-0.01}$	$9.92^{+0.82}_{-2.26}$	8.73
37	53.05	-27.96	0.04	8.17 ± 0.03	$-1.63^{+0.01}_{-0.01}$	$7.97^{+0.01}_{-0.01}$	$-1.52^{+0.02}_{-0.01}$	$-1.92^{+0.29}_{-0.08}$	2.94
38	53.27	-28.34	0.04	9.18 ± 0.02	$-0.74^{+0.01}_{-0.02}$	$9.6^{+0.01}_{-0.01}$	$-0.14^{+0.01}_{-0.01}$	$2.15^{+0.03}_{-0.01}$	4.65
39	52.56	-27.56	0.04	8.94 ± 0.03	$-1.28^{+0.01}_{-0.02}$	$8.84^{+0.01}_{-0.01}$	$-0.86^{+0.01}_{-0.01}$	$2.03^{+0.08}_{-0.05}$	3.07
40	52.62	-27.65	0.04	8.46 ± 0.03	$-2.0^{+0.01}_{-0.01}$	$8.19^{+0.01}_{-0.04}$	$-1.79^{+0.01}_{-0.01}$	$11.54^{+0.7}_{-1.77}$	6.61
43	52.35	-27.58	0.05	8.6 ± 0.05	$-2.39^{+0.02}_{-0.03}$	$8.35^{+0.01}_{-0.02}$	$-1.37^{+0.01}_{-0.02}$	$2.66^{+0.05}_{-0.06}$	15.59
46	53.53	-28.11	0.06	9.69 ± 0.02	$-0.51^{+0.01}_{-0.01}$	$9.59^{+0.01}_{-0.01}$	$-0.09^{+0.01}_{-0.01}$	$2.28^{+0.01}_{-0.09}$	5.21
47	53.32	-28.20	0.06	8.61 ± 0.02	$-1.66^{+0.04}_{-0.02}$	$8.57^{+0.01}_{-0.01}$	$-1.43^{+0.01}_{-0.04}$	$12.82^{+0.35}_{-0.46}$	2.02

Continued on the next page...

ID	RA [°]	Dec [°]	z_{HI}	$\log(M_{HI})$ $[M_\odot]$	$\log(\mathrm{SFR})$ $[M_\odot/\mathrm{yr}]$	$\log(M_*)$ $[M_\odot]$	$\log(\mathrm{SFR_{peak}})$ $[M_\odot/\mathrm{yr}]$	τ_{peak} [Gyr]	χ^2
48	52.73	-28.07	0.06	9.21 ± 0.02	$-0.34^{+0.01}_{-0.01}$	$10.12^{+0.01}_{-0.01}$	$0.33^{+0.01}_{-0.01}$	$13.38^{+0.0}_{-0.03}$	13.56
49	53.31	-28.18	0.06	8.72 ± 0.02	$-1.86^{+0.02}_{-0.02}$	$8.21^{+0.02}_{-0.02}$	$-1.49^{+0.01}_{-0.03}$	$1.91^{+0.05}_{-0.11}$	59.65
50	52.86	-28.31	0.06	8.83 ± 0.02	$-0.06^{+0.01}_{-0.01}$	$9.96^{+0.01}_{-0.02}$	$0.14^{+0.01}_{-0.01}$	$2.48^{+0.09}_{-0.13}$	6.61
51	52.58	-28.50	0.06	9.28 ± 0.02	$-1.11^{+0.01}_{-0.02}$	$9.06^{+0.01}_{-0.01}$	$-0.67^{+0.01}_{-0.01}$	$3.34^{+0.12}_{-0.06}$	6.47
52	52.49	-28.31	0.06	10.03 ± 0.02	$0.5^{+0.01}_{-0.01}$	$10.74^{+0.01}_{-0.01}$	$1.0^{+0.01}_{-0.01}$	$2.02^{+0.03}_{-0.02}$	2.18
53	52.56	-28.42	0.06	9.12 ± 0.02	$-0.83^{+0.03}_{-0.01}$	$9.36^{+0.01}_{-0.01}$	$-0.37^{+0.01}_{-0.01}$	$1.97^{+0.01}_{-0.02}$	6.25
54	53.89	-28.20	0.06	9.59 ± 0.02	$-0.6^{+0.02}_{-0.03}$	$9.78^{+0.01}_{-0.01}$	$0.05^{+0.01}_{-0.01}$	$2.23^{+0.02}_{-0.04}$	11.73
56	53.28	-28.24	0.06	9.33 ± 0.02	$0.13^{+0.01}_{-0.02}$	$9.96^{+0.04}_{-0.03}$	$0.32^{+0.02}_{-0.01}$	$1.6^{+0.17}_{-0.07}$	4.63
57	53.90	-28.15	0.06	8.83 ± 0.03	$-1.1^{+0.01}_{-0.01}$	$9.23^{+0.01}_{-0.01}$	$-0.5^{+0.01}_{-0.01}$	$2.15^{+0.01}_{-0.02}$	3.98
58	53.91	-28.32	0.06	9.52 ± 0.02	$-0.14^{+0.01}_{-0.01}$	$10.54^{+0.01}_{-0.01}$	$0.81^{+0.01}_{-0.01}$	$2.54^{+0.02}_{-0.01}$	6.36
59	53.64	-28.36	0.06	9.47 ± 0.02	$-1.11^{+0.02}_{-0.01}$	$9.36^{+0.01}_{-0.01}$	$-0.37^{+0.02}_{-0.01}$	$2.31^{+0.02}_{-0.01}$	2.52
60	53.73	-28.20	0.06	8.55 ± 0.03	$-1.76^{+0.02}_{-0.02}$	$8.3^{+0.03}_{-0.01}$	$-1.38^{+0.02}_{-0.01}$	$4.21^{+0.28}_{-0.15}$	7.20
65	53.65	-28.44	0.06	9.28 ± 0.02	$-0.87^{+0.02}_{-0.01}$	$9.67^{+0.01}_{-0.01}$	$-0.07^{+0.01}_{-0.01}$	$2.39^{+0.01}_{-0.02}$	1.37
68	53.49	-28.25	0.06	8.27 ± 0.03	$-1.63^{+0.01}_{-0.03}$	$9.01^{+0.01}_{-0.01}$	$-0.72^{+0.01}_{-0.01}$	$2.49^{+0.04}_{-0.02}$	11.49
69	52.49	-28.27	0.06	8.53 ± 0.04	$-1.01^{+0.02}_{-0.02}$	$7.96^{+0.02}_{-0.01}$	$-0.99^{+0.03}_{-0.02}$	$-0.38^{+0.03}_{-0.04}$	6.65
70	53.33	-28.26	0.06	8.63 ± 0.03	$-1.3^{+0.02}_{-0.02}$	$8.65^{+0.01}_{-0.01}$	$-0.98^{+0.01}_{-0.01}$	$2.19^{+0.1}_{-0.07}$	2.70
72	53.00	-28.07	0.06	8.36 ± 0.03	$-1.53^{+0.02}_{-0.02}$	$8.63^{+0.01}_{-0.01}$	$-1.07^{+0.01}_{-0.03}$	$3.94^{+0.15}_{-0.14}$	2.62

Continued on the next page...

ID	RA [°]	Dec [°]	z_{HI}	$\log(M_{HI})$ $[M_\odot]$	$\log(\mathrm{SFR})$ $[M_\odot/\mathrm{yr}]$	$\log(M_*)$ $[M_\odot]$	$\log(\mathrm{SFR_{peak}})$ $[M_\odot/\mathrm{yr}]$	τ_{peak} [Gyr]	χ^2
73	52.64	-28.36	0.06	9.09 ± 0.02	$-1.23^{+0.02}_{-0.01}$	$7.93^{+0.02}_{-0.03}$	$-1.22^{+0.03}_{-0.01}$	$0.03^{+0.16}_{-0.47}$	52.66
74	53.22	-28.08	0.06	8.76 ± 0.02	$-1.21^{+0.02}_{-0.03}$	$8.59^{+0.03}_{-0.01}$	$-1.02^{+0.01}_{-0.02}$	$1.75^{+0.3}_{-0.08}$	1.07
75	52.58	-28.43	0.06	8.85 ± 0.03	$-1.5^{+0.02}_{-0.02}$	$8.84^{+0.01}_{-0.01}$	$-0.89^{+0.01}_{-0.01}$	$2.16^{+0.03}_{-0.03}$	5.06
76	52.75	-28.39	0.06	8.65 ± 0.03	$-1.08^{+0.01}_{-0.01}$	$9.73^{+0.01}_{-0.01}$	$0.03^{+0.01}_{-0.01}$	$13.35^{+0.03}_{-0.19}$	6.37
80	53.20	-28.12	0.06	8.33 ± 0.03	$-1.98^{+0.02}_{-0.01}$	$8.09^{+0.01}_{-0.06}$	$-1.88^{+0.01}_{-0.01}$	$8.03^{+0.66}_{-2.07}$	4.35
81	53.10	-28.20	0.07	9.1 ± 0.02	$-1.62^{+0.01}_{-0.02}$	$8.6^{+0.01}_{-0.01}$	$-1.11^{+0.01}_{-0.01}$	$3.06^{+0.06}_{-0.15}$	6.10
83	53.12	-28.26	0.07	8.84 ± 0.02	$-1.99^{+0.03}_{-0.02}$	$7.72^{+0.02}_{-0.02}$	$-1.86^{+0.01}_{-0.01}$	$1.63^{+0.18}_{-0.26}$	9.04
85	52.28	-28.04	0.07	9.71 ± 0.02	$0.1^{+0.01}_{-0.01}$	$10.55^{+0.01}_{-0.01}$	$0.81^{+0.01}_{-0.01}$	$2.28^{+0.02}_{-0.02}$	10.10
86	52.55	-27.88	0.07	9.4 ± 0.02	$-1.53^{+0.02}_{-0.04}$	$9.16^{+0.01}_{-0.01}$	$-0.57^{+0.01}_{-0.01}$	$2.55^{+0.04}_{-0.02}$	4.68
87	53.38	-28.44	0.07	9.04 ± 0.02	$-1.34^{+0.01}_{-0.01}$	$8.89^{+0.01}_{-0.01}$	$-1.07^{+0.01}_{-0.01}$	$12.71^{+0.49}_{-1.41}$	8.70
89	52.59	-27.86	0.07	8.89 ± 0.03	$-1.44^{+0.02}_{-0.02}$	$8.95^{+0.02}_{-0.01}$	$-0.78^{+0.02}_{-0.01}$	$2.22^{+0.05}_{-0.02}$	3.20
90	52.43	-27.88	0.07	8.94 ± 0.03	$-0.74^{+0.01}_{-0.01}$	$8.13^{+0.02}_{-0.01}$	$-0.67^{+0.01}_{-0.01}$	$-0.63^{+0.06}_{-0.03}$	60.18
91	52.75	-27.64	0.07	9.73 ± 0.02	$-0.16^{+0.01}_{-0.01}$	$9.96^{+0.01}_{-0.01}$	$-0.09^{+0.01}_{-0.03}$	$12.76^{+0.52}_{-2.1}$	6.72
93	52.40	-27.95	0.07	9.99 ± 0.02	$0.61^{+0.01}_{-0.01}$	$10.82^{+0.01}_{-0.01}$	$0.91^{+0.01}_{-0.01}$	$13.28^{+0.1}_{-0.17}$	8.35
94	53.68	-28.21	0.07	9.18 ± 0.02	$-1.3^{+0.01}_{-0.03}$	$8.15^{+0.05}_{-0.05}$	$-1.27^{+0.01}_{-0.01}$	$0.59^{+0.53}_{-0.25}$	1.50
96	52.43	-27.82	0.07	9.21 ± 0.03	$-1.0^{+0.01}_{-0.02}$	$9.63^{+0.01}_{-0.01}$	$-0.1^{+0.01}_{-0.01}$	$2.48^{+0.01}_{-0.01}$	6.55
97	53.68	-28.08	0.07	9.01 ± 0.03	$-0.44^{+0.02}_{-0.03}$	$9.28^{+0.02}_{-0.02}$	$-0.3^{+0.01}_{-0.01}$	$1.78^{+0.2}_{-0.17}$	5.34

Continued on the next page...

ID	RA [°]	Dec [°]	z_{HI}	$\log(M_{HI})$ $[M_\odot]$	$\log(\mathrm{SFR})$ $[M_\odot/\mathrm{yr}]$	$\log(M_*)$ $[M_\odot]$	$\log(\mathrm{SFR_{peak}})$ $[M_\odot/\mathrm{yr}]$	τ_{peak} [Gyr]	χ^2
98	53.69	-28.17	0.07	8.97 ± 0.03	$-1.07^{+0.01}_{-0.01}$	$9.14^{+0.01}_{-0.01}$	$-0.59^{+0.01}_{-0.01}$	$2.01^{+0.03}_{-0.02}$	6.66
99	53.03	-28.17	0.07	9.07 ± 0.02	$-1.04^{+0.03}_{-0.03}$	$9.38^{+0.02}_{-0.01}$	$-0.36^{+0.02}_{-0.01}$	$2.26^{+0.05}_{-0.04}$	3.17
101	53.49	-27.65	0.07	9.67 ± 0.02	$-0.4^{+0.03}_{-0.01}$	$9.91^{+0.01}_{-0.01}$	$0.23^{+0.01}_{-0.01}$	$2.6^{+0.04}_{-0.02}$	4.65
102	52.50	-27.97	0.07	8.73 ± 0.04	$-1.1^{+0.01}_{-0.01}$	$9.21^{+0.01}_{-0.01}$	$-0.63^{+0.01}_{-0.01}$	$13.38^{+0.0}_{-0.09}$	7.35
104	53.15	-27.62	0.07	9.22 ± 0.02	$-0.72^{+0.01}_{-0.01}$	$8.49^{+0.02}_{-0.02}$	$-0.71^{+0.01}_{-0.01}$	$0.25^{+0.13}_{-0.09}$	46.72
105	53.72	-27.74	0.07	9.06 ± 0.03	$-0.62^{+0.01}_{-0.01}$	$9.12^{+0.03}_{-0.03}$	$-0.46^{+0.01}_{-0.01}$	$1.71^{+0.06}_{-0.12}$	10.23
106	52.52	-27.79	0.07	9.45 ± 0.02	$-1.42^{+0.02}_{-0.02}$	$8.9^{+0.01}_{-0.01}$	$-0.76^{+0.01}_{-0.01}$	$2.96^{+0.1}_{-0.12}$	6.53
107	52.33	-27.92	0.07	9.4 ± 0.02	$-1.11^{+0.01}_{-0.02}$	$9.09^{+0.01}_{-0.01}$	$-0.6^{+0.02}_{-0.02}$	$2.37^{+0.18}_{-0.22}$	8.86
111	52.45	-27.85	0.07	8.67 ± 0.03	$-1.58^{+0.01}_{-0.02}$	$8.29^{+0.01}_{-0.03}$	$-1.35^{+0.01}_{-0.01}$	$1.86^{+0.1}_{-0.08}$	5.77
114	52.32	-27.94	0.07	9.36 ± 0.03	$-0.25^{+0.01}_{-0.02}$	$9.88^{+0.01}_{-0.01}$	$0.21^{+0.01}_{-0.02}$	$2.3^{+0.03}_{-0.12}$	4.94
115	53.30	-27.97	0.07	8.38 ± 0.03	$-2.03^{+0.01}_{-0.01}$	$7.48^{+0.04}_{-0.05}$	$-1.92^{+0.01}_{-0.01}$	$-2.0^{+0.14}_{-0.01}$	6.80
120	52.72	-27.68	0.07	9.52 ± 0.02	$-1.33^{+0.01}_{-0.01}$	$9.44^{+0.01}_{-0.01}$	$-0.22^{+0.01}_{-0.01}$	$4.52^{+0.04}_{-0.02}$	10.05
121	53.26	-27.91	0.07	9.02 ± 0.02	$-0.68^{+0.02}_{-0.01}$	$9.69^{+0.01}_{-0.02}$	$-0.02^{+0.01}_{-0.01}$	$2.46^{+0.03}_{-0.03}$	3.04
125	52.73	-28.01	0.07	9.2 ± 0.02	$-1.41^{+0.01}_{-0.01}$	$8.19^{+0.02}_{-0.01}$	$-1.3^{+0.02}_{-0.01}$	$-1.88^{+0.2}_{-0.12}$	9.40
130	52.60	-28.40	0.07	8.65 ± 0.04	$-1.22^{+0.02}_{-0.01}$	$9.02^{+0.01}_{-0.01}$	$-0.71^{+0.01}_{-0.01}$	$2.02^{+0.05}_{-0.03}$	6.67
150	53.82	-28.08	0.07	9.33 ± 0.02	$-0.94^{+0.03}_{-0.01}$	$8.74^{+0.04}_{-0.09}$	$-0.88^{+0.02}_{-0.02}$	$3.04^{+0.54}_{-1.06}$	2.53
151	52.99	-28.14	0.07	8.38 ± 0.04	$-0.86^{+0.01}_{-0.02}$	$8.85^{+0.02}_{-0.03}$	$-0.73^{+0.01}_{-0.01}$	$1.86^{+0.35}_{-0.31}$	4.35

Continued on the next page...

ID	RA [°]	Dec [°]	z_{HI}	$\log(M_{HI})$ $[M_\odot]$	$\log(\mathrm{SFR})$ $[M_\odot/\mathrm{yr}]$	$\log(M_*)$ $[M_\odot]$	$\log(\mathrm{SFR_{peak}})$ $[M_\odot/\mathrm{yr}]$	τ_{peak} [Gyr]	χ^2
152	52.77	-28.06	0.08	9.82 ± 0.02	$-0.12^{+0.01}_{-0.01}$	$10.11^{+0.01}_{-0.01}$	$0.21^{+0.01}_{-0.01}$	$13.28^{+0.1}_{-0.3}$	4.20
153	52.81	-28.10	0.08	8.67 ± 0.03	$-1.72^{+0.01}_{-0.01}$	$8.49^{+0.01}_{-0.01}$	$-1.43^{+0.02}_{-0.01}$	$13.22^{+0.16}_{-0.34}$	6.18
154	53.12	-27.74	0.08	9.81 ± 0.02	$0.34^{+0.01}_{-0.01}$	$10.53^{+0.01}_{-0.01}$	$0.81^{+0.01}_{-0.01}$	$2.07^{+0.07}_{-0.01}$	10.75
155	53.79	-27.54	0.08	9.61 ± 0.03	$-0.58^{+0.02}_{-0.02}$	$9.4^{+0.01}_{-0.01}$	$-0.28^{+0.01}_{-0.01}$	$1.82^{+0.05}_{-0.03}$	5.30
159	53.06	-27.65	0.08	9.4 ± 0.02	$-0.97^{+0.01}_{-0.01}$	$9.24^{+0.01}_{-0.01}$	$-0.46^{+0.01}_{-0.01}$	$2.23^{+0.03}_{-0.03}$	7.54
162	52.38	-28.23	0.08	9.74 ± 0.02	$-0.53^{+0.02}_{-0.02}$	$9.56^{+0.01}_{-0.01}$	$-0.17^{+0.01}_{-0.01}$	$1.8^{+0.05}_{-0.06}$	18.21
165	52.73	-28.11	0.08	10.03 ± 0.02	$-0.0^{+0.01}_{-0.01}$	$10.68^{+0.01}_{-0.01}$	$1.0^{+0.01}_{-0.01}$	$13.38^{+0.0}_{-0.03}$	17.16
166	52.75	-28.17	0.08	10.03 ± 0.02	$0.68^{+0.01}_{-0.01}$	$10.83^{+0.01}_{-0.01}$	$0.82^{+0.01}_{-0.03}$	$13.25^{+0.13}_{-0.35}$	5.97
167	52.65	-27.77	0.08	9.37 ± 0.02	$-1.09^{+0.01}_{-0.02}$	$9.0^{+0.01}_{-0.02}$	$-0.65^{+0.01}_{-0.01}$	$2.51^{+0.05}_{-0.04}$	1.03
168	52.87	-28.38	0.08	9.64 ± 0.02	$-0.56^{+0.03}_{-0.03}$	$9.53^{+0.02}_{-0.01}$	$-0.17^{+0.01}_{-0.01}$	$1.99^{+0.05}_{-0.03}$	5.04
169	52.71	-28.17	0.08	9.09 ± 0.03	$-0.9^{+0.02}_{-0.02}$	$9.35^{+0.01}_{-0.01}$	$-0.36^{+0.01}_{-0.01}$	$2.23^{+0.02}_{-0.05}$	8.27
171	52.72	-28.17	0.08	9.79 ± 0.02	$-0.41^{+0.01}_{-0.01}$	$9.7^{+0.01}_{-0.01}$	$-0.03^{+0.01}_{-0.01}$	$1.87^{+0.1}_{-0.06}$	4.65
174	52.58	-27.92	0.08	10.14 ± 0.02	$0.47^{+0.01}_{-0.01}$	$10.62^{+0.01}_{-0.01}$	$0.96^{+0.01}_{-0.01}$	$2.6^{+0.03}_{-0.04}$	4.74
175	52.47	-28.08	0.08	9.55 ± 0.02	$-2.53^{+0.01}_{-0.01}$	$5.79^{+0.01}_{-0.01}$	$-1.12^{+0.01}_{-0.02}$	$-2.0^{+0.01}_{-0.01}$	36.09
178	52.69	-28.35	0.08	9.75 ± 0.02	$0.22^{+0.02}_{-0.01}$	$9.67^{+0.03}_{-0.04}$	$0.25^{+0.01}_{-0.02}$	$0.87^{+0.38}_{-0.36}$	2.05
179	52.70	-28.35	0.08	8.82 ± 0.03	$-2.53^{+0.01}_{-0.01}$	$6.24^{+0.01}_{-0.02}$	$-2.39^{+0.01}_{-0.01}$	$-0.92^{+0.04}_{-0.02}$	32.04
180	52.70	-28.46	0.08	9.36 ± 0.02	$-0.14^{+0.01}_{-0.02}$	$10.42^{+0.01}_{-0.02}$	$0.55^{+0.01}_{-0.02}$	$6.4^{+0.11}_{-0.08}$	3.41

Continued on the next page...

ID	RA [°]	Dec [°]	z_{HI}	$\log(M_{HI})$ $[M_\odot]$	$\log(\mathrm{SFR})$ $[M_\odot/\mathrm{yr}]$	$\log(M_*)$ $[M_\odot]$	$\log(\mathrm{SFR_{peak}})$ $[M_\odot/\mathrm{yr}]$	τ_{peak} [Gyr]	χ^2
181	52.66	-28.49	0.08	8.99 ± 0.03	$-0.76^{+0.04}_{-0.02}$	$9.2^{+0.04}_{-0.03}$	$-0.5^{+0.01}_{-0.02}$	$4.36^{+0.48}_{-0.32}$	5.52
183	52.79	-28.44	0.08	9.22 ± 0.03	$-0.33^{+0.01}_{-0.01}$	$9.88^{+0.01}_{-0.01}$	$-0.06^{+0.01}_{-0.01}$	$13.12^{+0.26}_{-1.02}$	4.01
185	52.82	-28.45	0.08	9.21 ± 0.02	$-0.66^{+0.01}_{-0.02}$	$9.64^{+0.01}_{-0.01}$	$-0.15^{+0.02}_{-0.01}$	$3.17^{+0.03}_{-0.04}$	3.30
186	52.62	-28.19	0.08	8.93 ± 0.03	$-0.86^{+0.01}_{-0.01}$	$8.99^{+0.03}_{-0.02}$	$-0.85^{+0.01}_{-0.02}$	$0.77^{+0.98}_{-1.07}$	4.09
188	52.42	-28.09	0.08	9.34 ± 0.02	$-0.67^{+0.01}_{-0.02}$	$9.29^{+0.01}_{-0.01}$	$-0.34^{+0.01}_{-0.01}$	$2.39^{+0.08}_{-0.16}$	11.40
189	52.69	-28.18	0.08	9.41 ± 0.02	$-1.29^{+0.01}_{-0.01}$	$7.79^{+0.05}_{-0.03}$	$-1.28^{+0.02}_{-0.01}$	$-0.17^{+0.07}_{-0.14}$	76.58
190	52.72	-28.10	0.08	8.85 ± 0.03	$-1.12^{+0.01}_{-0.01}$	$8.97^{+0.01}_{-0.02}$	$-1.11^{+0.01}_{-0.01}$	$12.53^{+0.7}_{-2.6}$	1.09
192	52.76	-28.33	0.08	9.61 ± 0.02	$-0.17^{+0.03}_{-0.05}$	$11.84^{+0.01}_{-0.01}$	$2.18^{+0.01}_{-0.01}$	$7.09^{+0.06}_{-0.11}$	4.27
194	52.90	-28.12	0.08	9.23 ± 0.02	$-0.44^{+0.03}_{-0.02}$	$9.65^{+0.01}_{-0.02}$	$-0.03^{+0.01}_{-0.01}$	$2.11^{+0.04}_{-0.05}$	2.59
195	52.85	-28.14	0.08	8.84 ± 0.03	$-1.01^{+0.01}_{-0.03}$	$9.12^{+0.01}_{-0.03}$	$-0.59^{+0.01}_{-0.01}$	$2.46^{+0.1}_{-0.06}$	8.48
198	53.84	-28.09	0.08	9.26 ± 0.03	$-0.48^{+0.01}_{-0.02}$	$9.79^{+0.01}_{-0.01}$	$0.01^{+0.01}_{-0.01}$	$3.43^{+0.18}_{-0.13}$	6.74
200	52.89	-27.57	0.08	9.19 ± 0.03	$-0.99^{+0.01}_{-0.02}$	$9.2^{+0.01}_{-0.01}$	$-0.45^{+0.01}_{-0.01}$	$2.89^{+0.04}_{-0.08}$	1.29
207	52.96	-28.42	0.08	9.2 ± 0.02	$-1.05^{+0.03}_{-0.02}$	$9.14^{+0.01}_{-0.02}$	$-0.59^{+0.01}_{-0.01}$	$2.01^{+0.02}_{-0.04}$	4.03
209	53.04	-28.45	0.08	9.26 ± 0.02	$0.53^{+0.02}_{-0.01}$	$10.59^{+0.03}_{-0.01}$	$0.93^{+0.01}_{-0.01}$	$2.18^{+0.05}_{-0.05}$	5.48
210	53.04	-27.75	0.08	9.16 ± 0.02	$-0.34^{+0.02}_{-0.01}$	$9.82^{+0.01}_{-0.01}$	$0.14^{+0.01}_{-0.01}$	$2.27^{+0.03}_{-0.02}$	5.92
215	52.64	-27.86	0.09	9.59 ± 0.02	$-0.13^{+0.02}_{-0.02}$	$9.44^{+0.02}_{-0.04}$	$-0.06^{+0.01}_{-0.01}$	$1.47^{+0.39}_{-0.4}$	1.64
216	53.19	-27.54	0.09	9.4 ± 0.02	$-0.7^{+0.01}_{-0.02}$	$9.45^{+0.01}_{-0.02}$	$-0.28^{+0.01}_{-0.01}$	$1.96^{+0.04}_{-0.04}$	10.54

Continued on the next page...

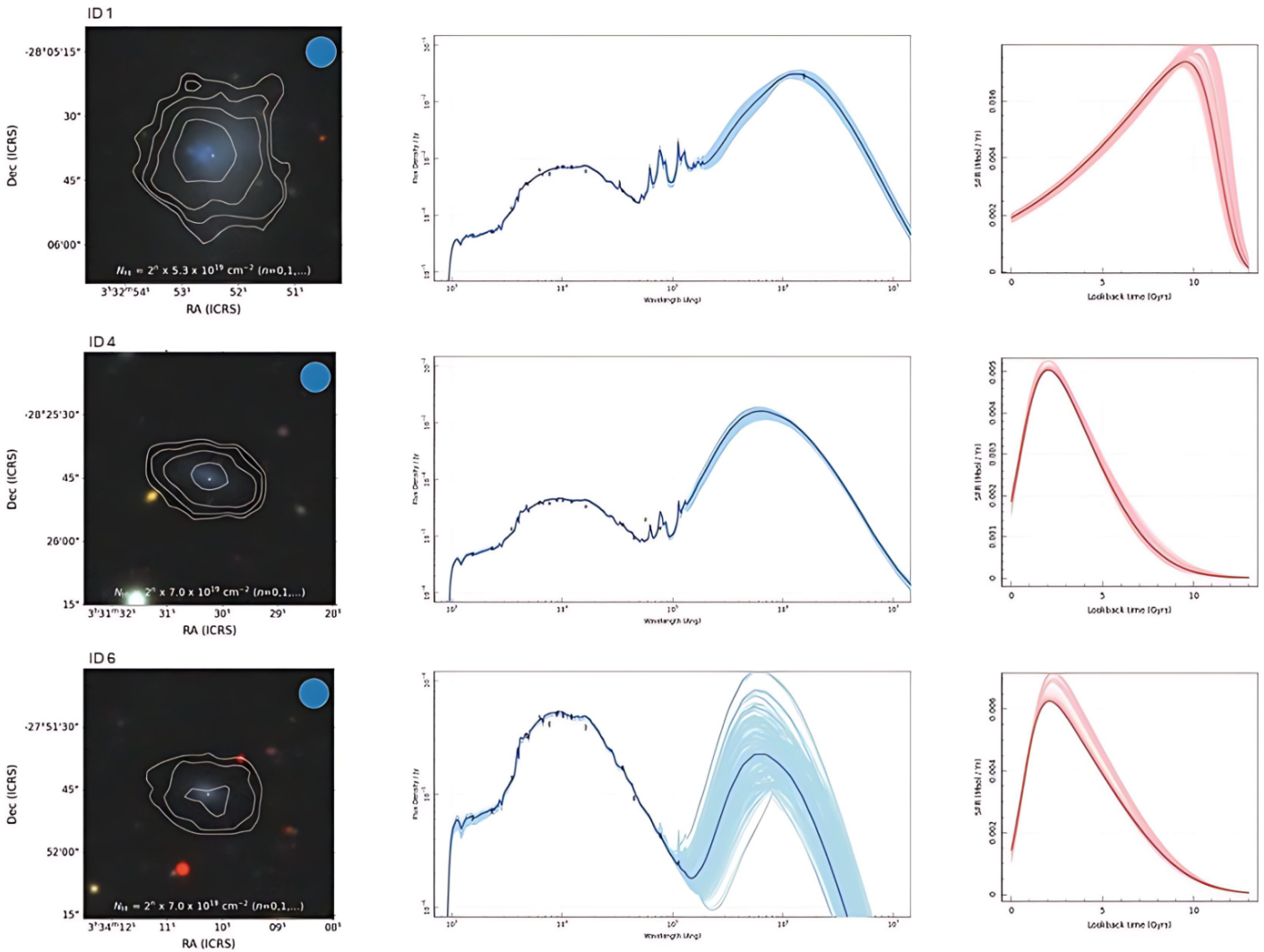
ID	RA [°]	Dec [°]	z_{HI}	$\log(M_{HI})$ $[M_\odot]$	$\log(\mathrm{SFR})$ $[M_\odot/\mathrm{yr}]$	$\log(M_*)$ $[M_\odot]$	$\log(\mathrm{SFR_{peak}})$ $[M_\odot/\mathrm{yr}]$	τ_{peak} [Gyr]	χ^2
217	53.02	-28.45	0.09	9.14 ± 0.03	$-0.68^{+0.03}_{-0.03}$	$9.57^{+0.01}_{-0.01}$	$-0.09^{+0.01}_{-0.01}$	$2.83^{+0.05}_{-0.02}$	2.19
218	53.25	-27.57	0.09	9.13 ± 0.03	$-0.8^{+0.02}_{-0.01}$	$9.45^{+0.01}_{-0.01}$	$-0.24^{+0.01}_{-0.01}$	$2.42^{+0.05}_{-0.03}$	4.44
219	52.58	-27.85	0.09	9.19 ± 0.03	$-0.54^{+0.06}_{-0.02}$	$9.32^{+0.01}_{-0.05}$	$-0.32^{+0.01}_{-0.01}$	$1.86^{+0.06}_{-0.13}$	1.03
220	53.82	-27.69	0.09	9.65 ± 0.03	$-1.81^{+0.02}_{-0.04}$	$8.31^{+0.01}_{-0.02}$	$-1.67^{+0.02}_{-0.01}$	$13.36^{+0.02}_{-0.19}$	18.39
221	53.83	-27.74	0.09	9.88 ± 0.02	$0.26^{+0.01}_{-0.01}$	$10.23^{+0.01}_{-0.01}$	$0.6^{+0.01}_{-0.01}$	$2.36^{+0.04}_{-0.04}$	2.18
222	53.82	-27.75	0.09	9.7 ± 0.03	$0.47^{+0.01}_{-0.01}$	$10.45^{+0.01}_{-0.01}$	$0.8^{+0.01}_{-0.01}$	$2.11^{+0.06}_{-0.11}$	5.72
223	53.70	-27.70	0.09	8.99 ± 0.03	$-1.21^{+0.02}_{-0.01}$	$8.81^{+0.02}_{-0.02}$	$-0.88^{+0.01}_{-0.01}$	$1.87^{+0.07}_{-0.06}$	3.42
225	53.53	-27.61	0.09	9.34 ± 0.03	$-0.78^{+0.02}_{-0.02}$	$9.51^{+0.01}_{-0.02}$	$-0.18^{+0.02}_{-0.01}$	$2.54^{+0.08}_{-0.07}$	3.55
226	53.10	-28.12	0.09	8.73 ± 0.03	$-1.72^{+0.02}_{-0.02}$	$8.39^{+0.01}_{-0.01}$	$-1.68^{+0.02}_{-0.02}$	$12.69^{+0.55}_{-1.66}$	2.23
227	53.65	-27.69	0.09	9.55 ± 0.03	$-0.02^{+0.01}_{-0.02}$	$10.84^{+0.01}_{-0.01}$	$1.01^{+0.01}_{-0.01}$	$5.06^{+0.04}_{-0.02}$	3.46
229	52.44	-27.76	0.09	9.58 ± 0.03	$-0.54^{+0.02}_{-0.02}$	$9.83^{+0.01}_{-0.01}$	$-0.06^{+0.01}_{-0.01}$	$5.06^{+0.07}_{-0.09}$	11.94
230	53.32	-27.84	0.09	9.27 ± 0.02	$0.16^{+0.01}_{-0.02}$	$10.65^{+0.01}_{-0.01}$	$0.96^{+0.01}_{-0.01}$	$2.8^{+0.03}_{-0.06}$	4.04
231	53.68	-27.67	0.09	9.82 ± 0.02	$-0.38^{+0.03}_{-0.03}$	$10.39^{+0.01}_{-0.01}$	$0.66^{+0.01}_{-0.01}$	$2.63^{+0.04}_{-0.03}$	4.58
232	53.23	-27.83	0.09	8.72 ± 0.03	$-1.4^{+0.02}_{-0.01}$	$8.32^{+0.03}_{-0.03}$	$-1.28^{+0.02}_{-0.01}$	$1.4^{+0.09}_{-0.11}$	2.65

A.3 SED fits

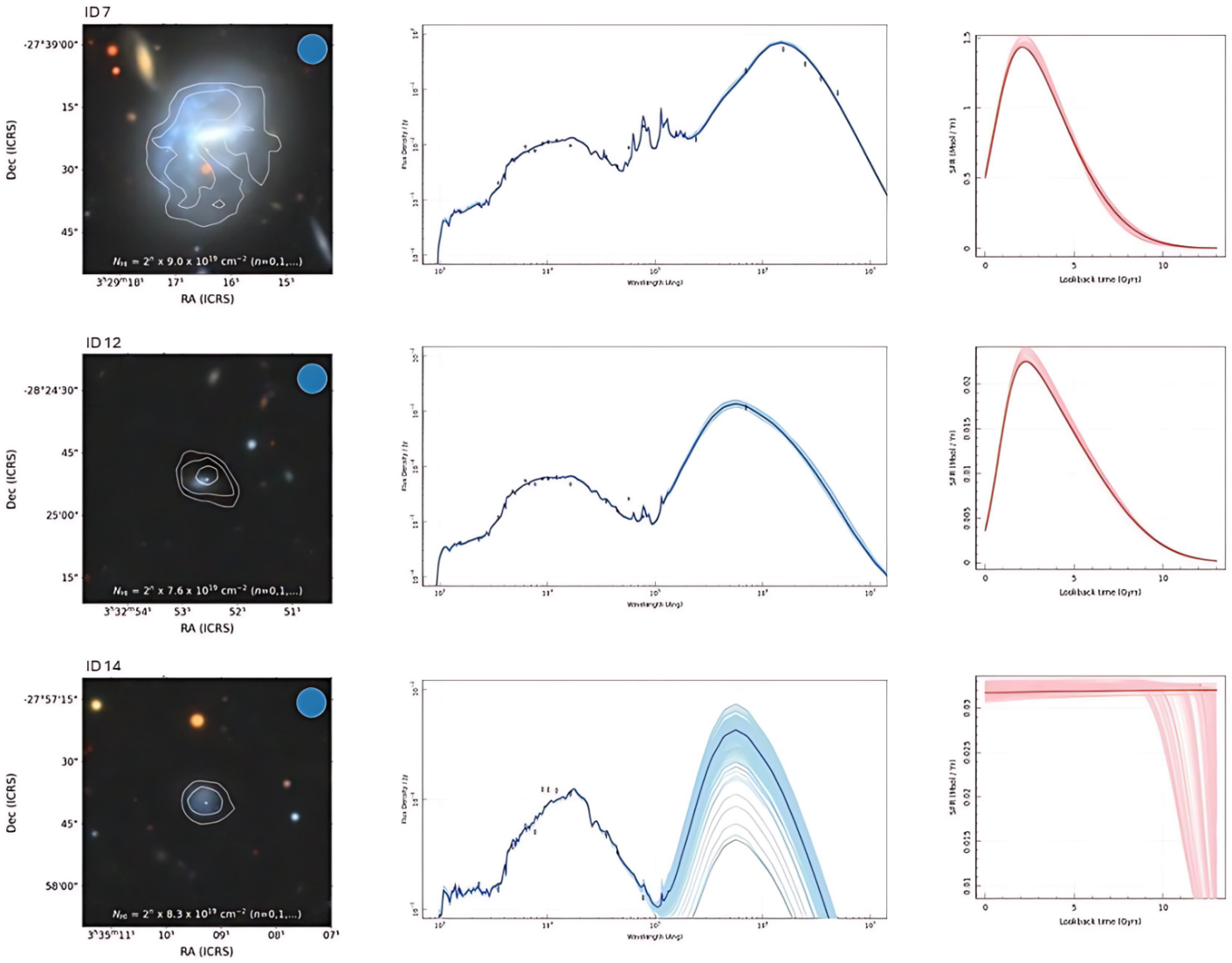
Below we provide the SED fits of all galaxies listed in Table A.1. The figures are grouped by their χ^2 value, with Figure A.1 showing fits where $\chi^2 < 5$, Figure A.2 where $5 < \chi^2 < 10$ and Figure A.3 where $\chi^2 > 10$. Each Figure is ordered in ascending ID numbers.

Each figure is structured as follows:
Left panel: The DECaLS optical colour image overlaid with the HI contour (produced by SIP).
Middle panel: The SED generated by ProSpect. The dark blue line shows the final SED fit, the light blue lines show the SEDs generated from each step of the final MCMC chain. The black points show the input phorometry (with uncertainties).
Right panel: the SFH generated by ProSpect. The red line shows the final SFH and the pink lines show the SFHs generated from each step of the final MCMC chain.

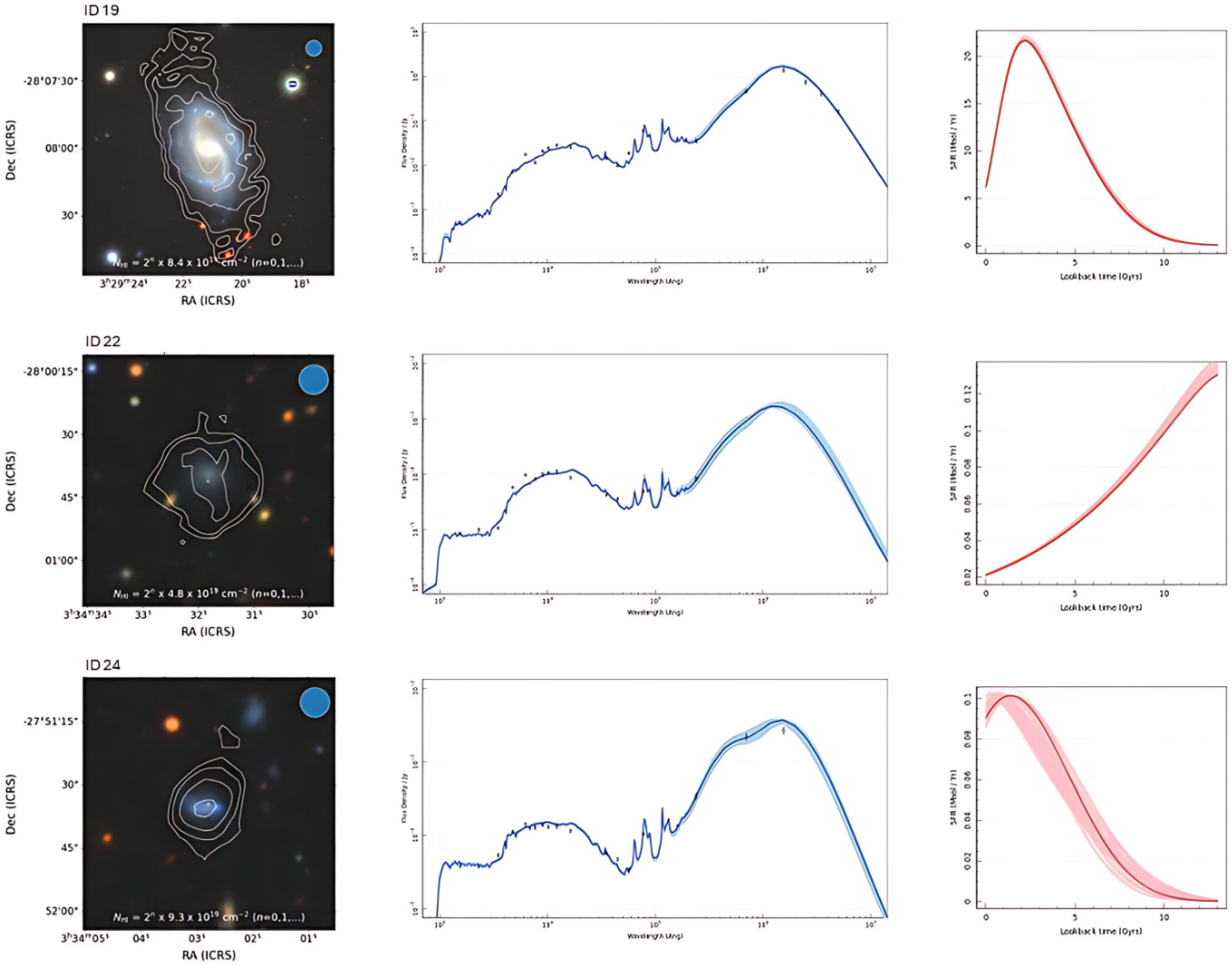
Figure A.1: DECaLS image overlaid with HI contours (left panel), SED (middle panel) and SFH (right panel) of galaxies where $\chi^2 < 5$.



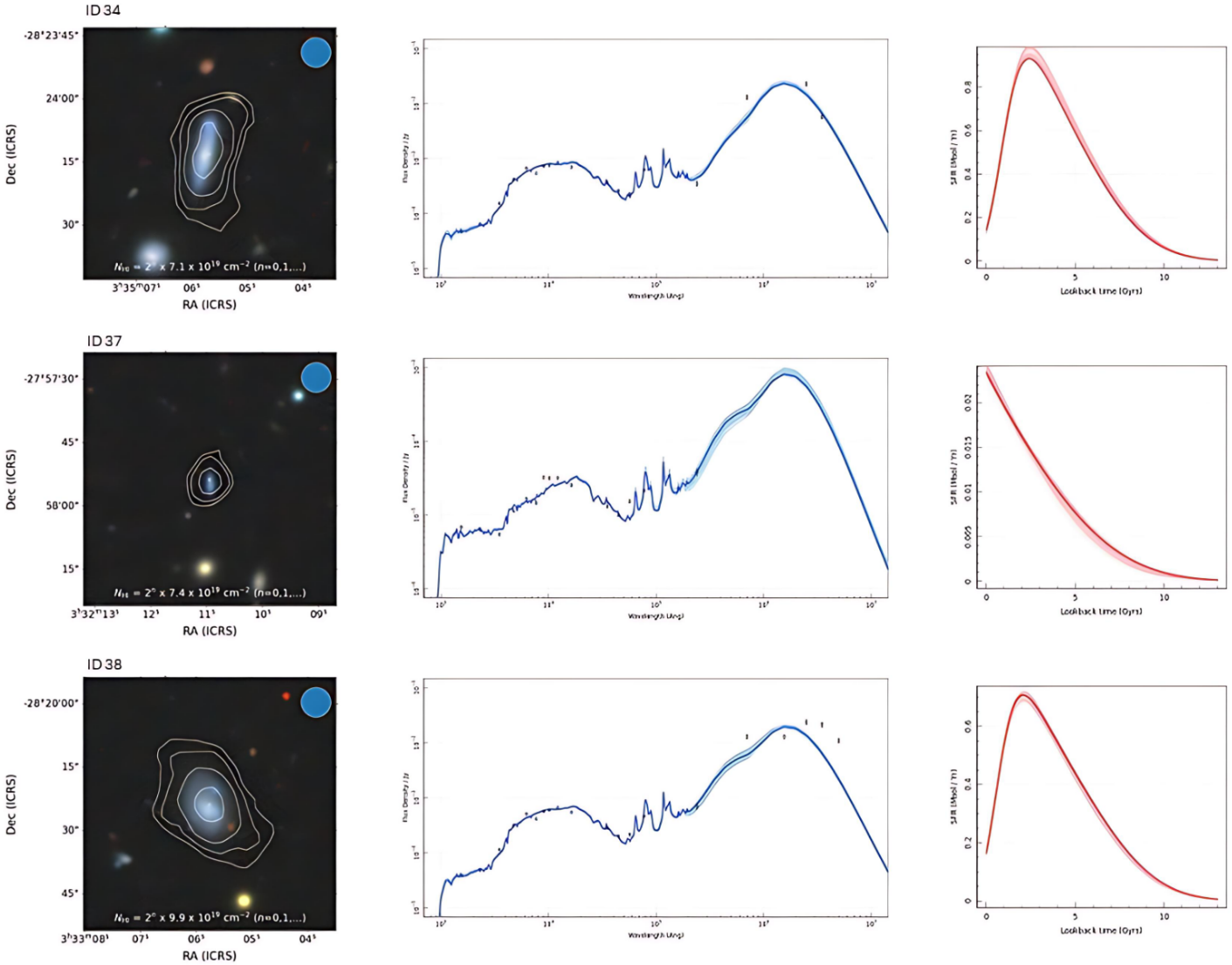
Continued on next page...



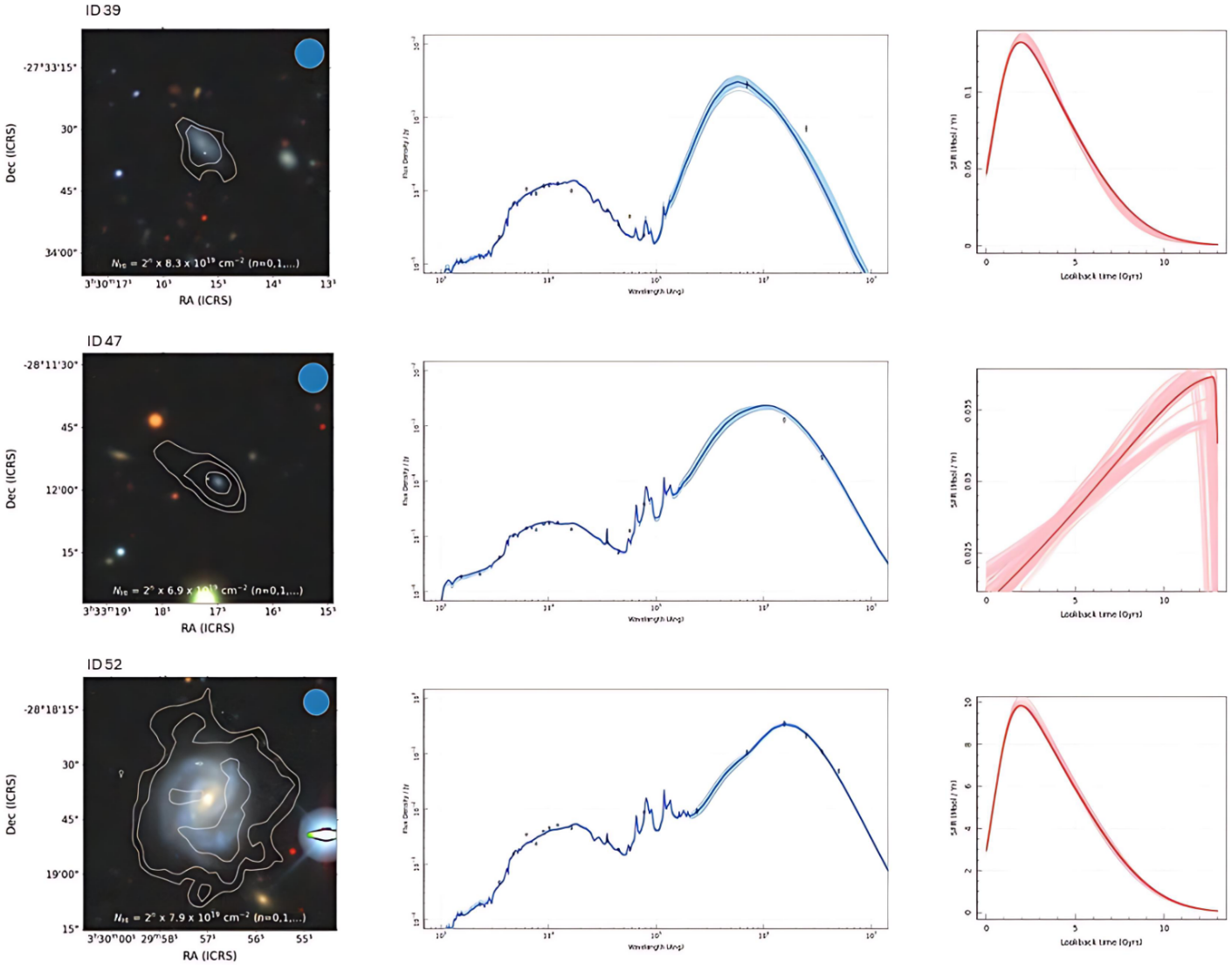
Continued on next page...



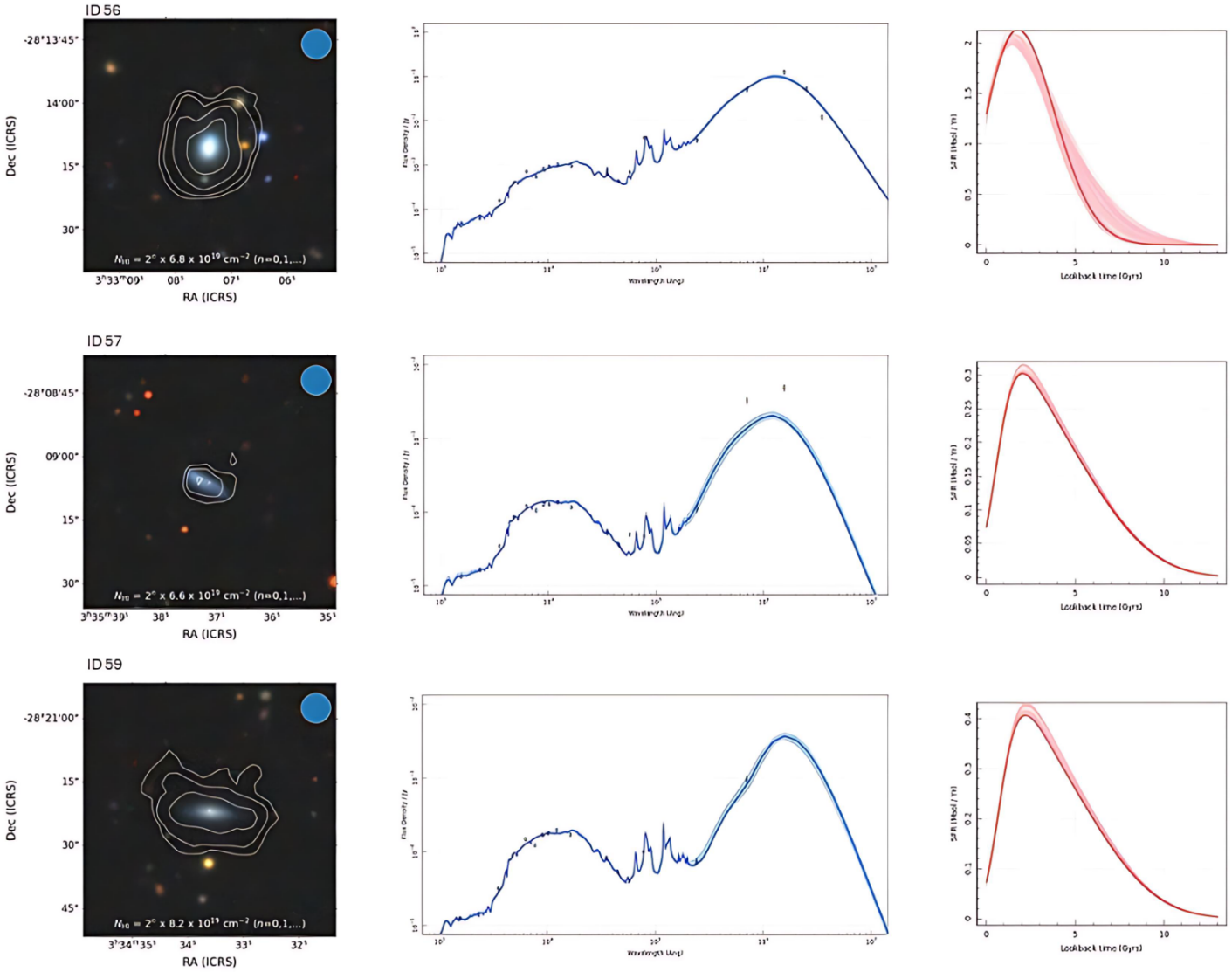
Continued on next page...



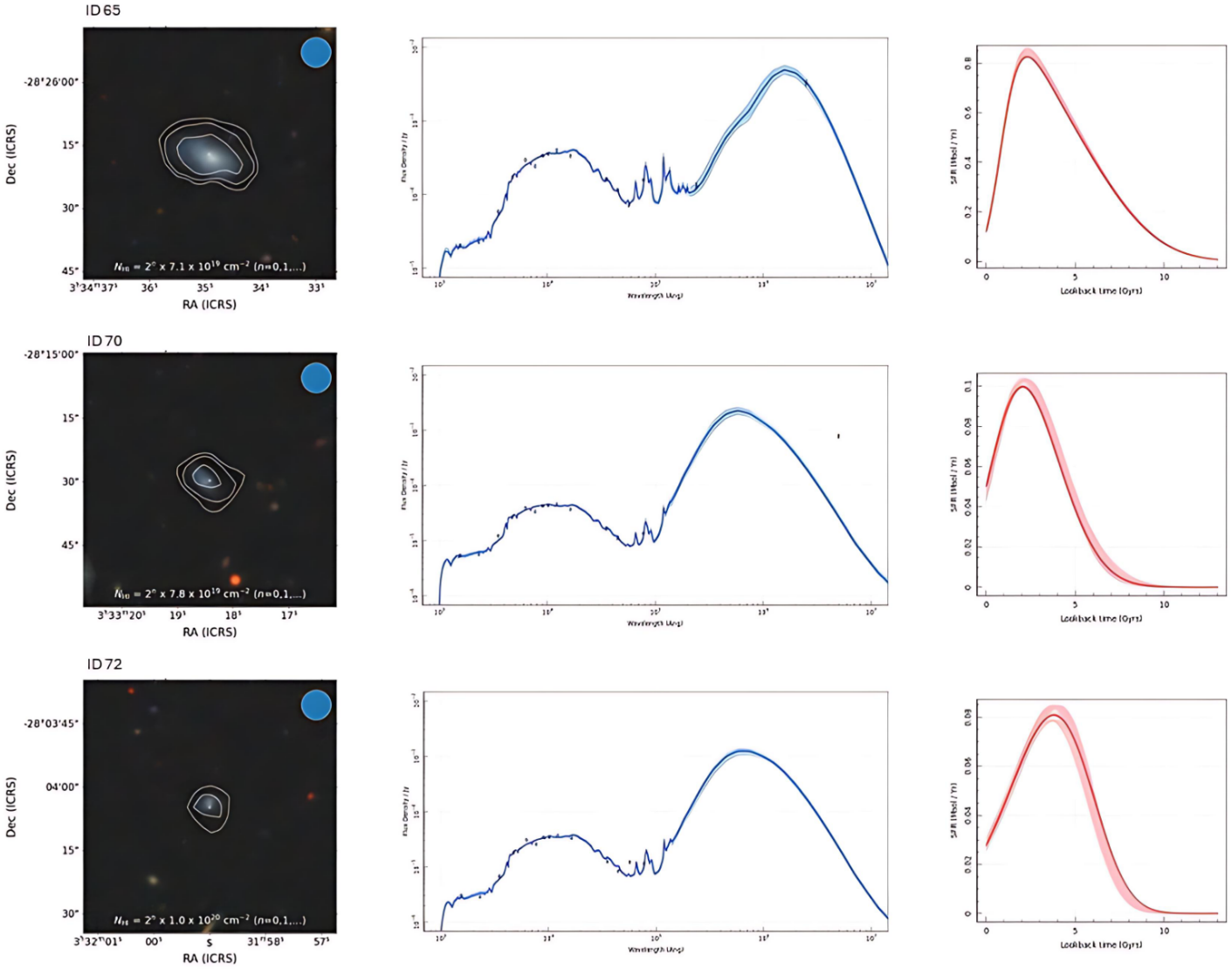
Continued on next page...



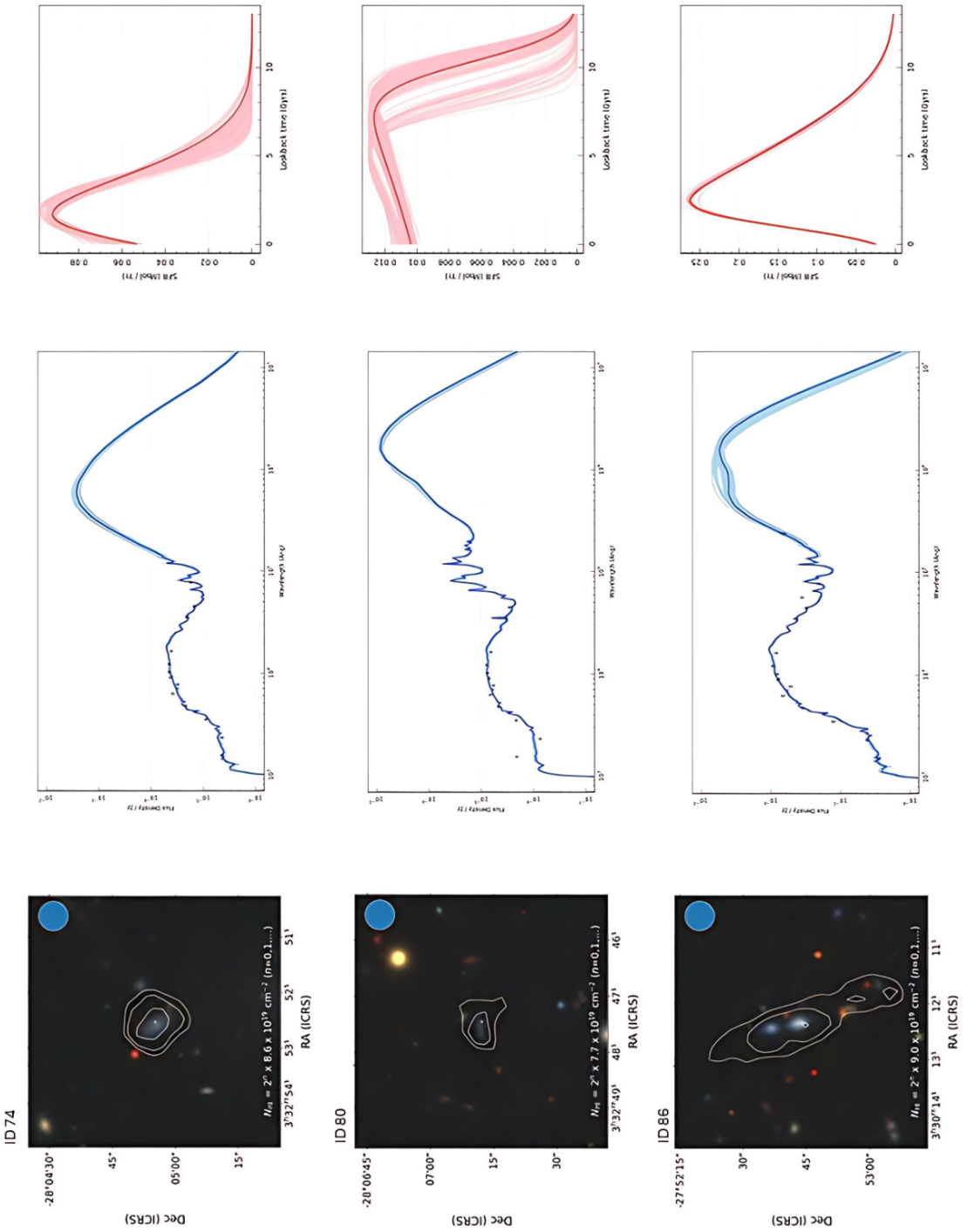
Continued on next page...



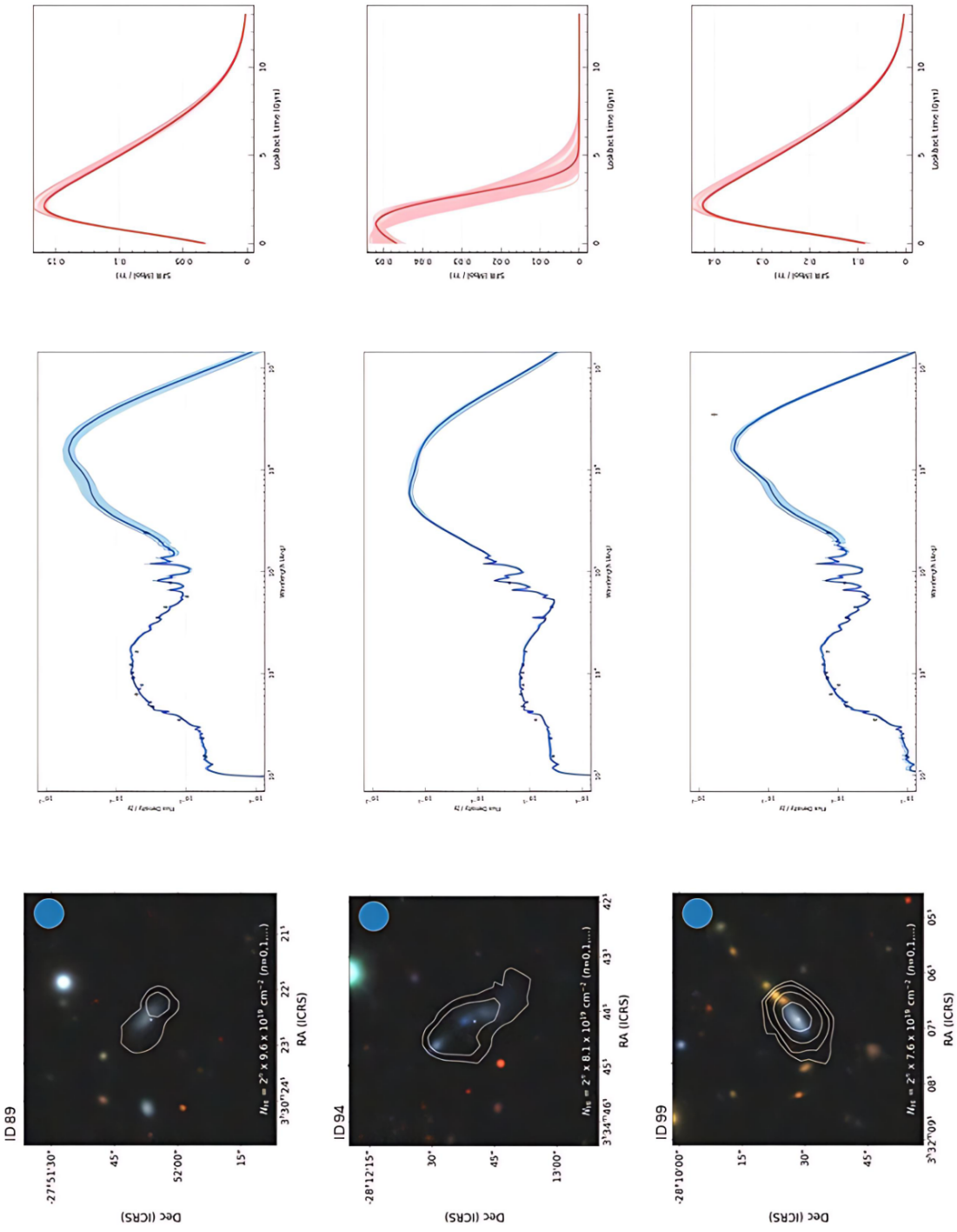
Continued on next page...



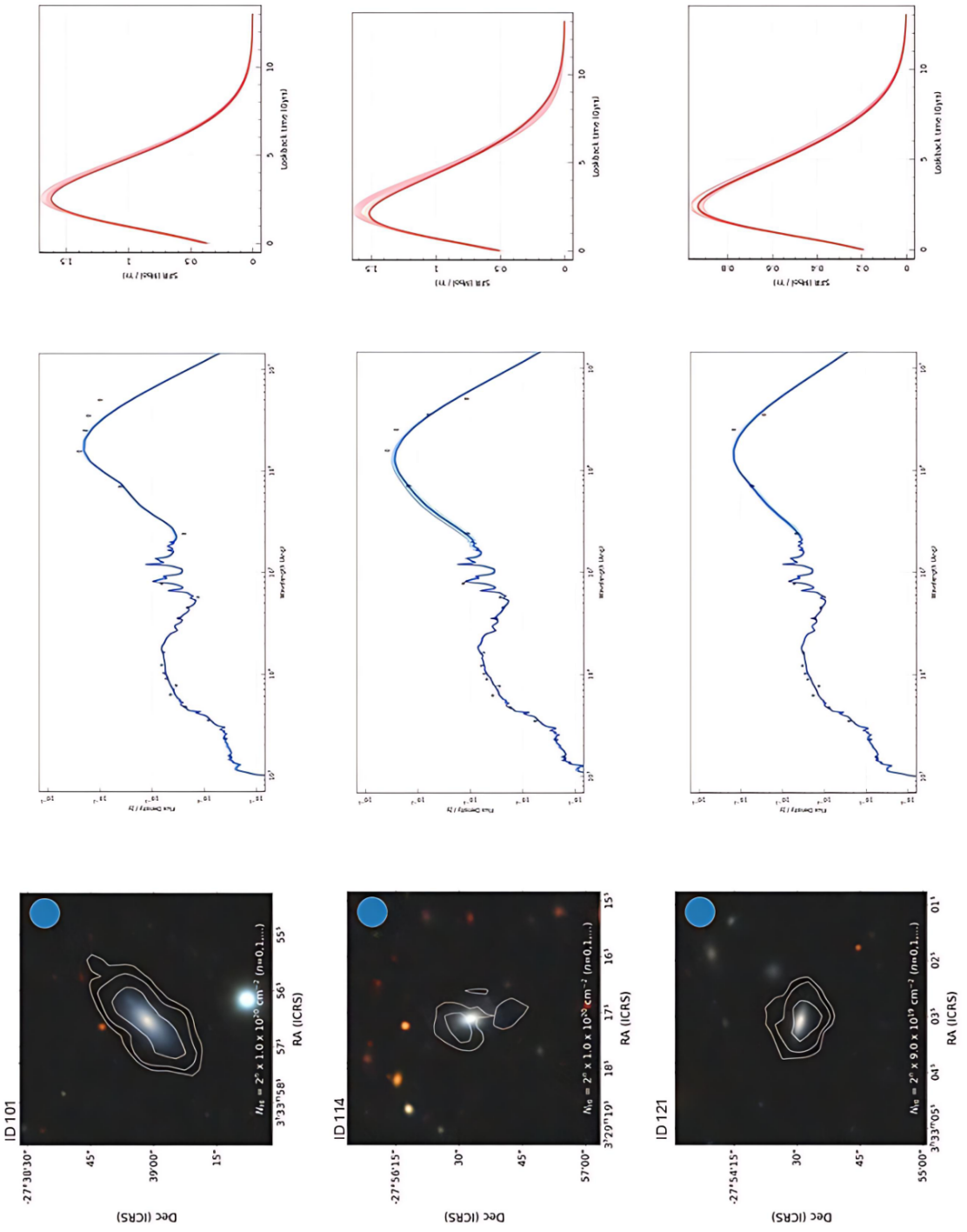
Continued on next page...



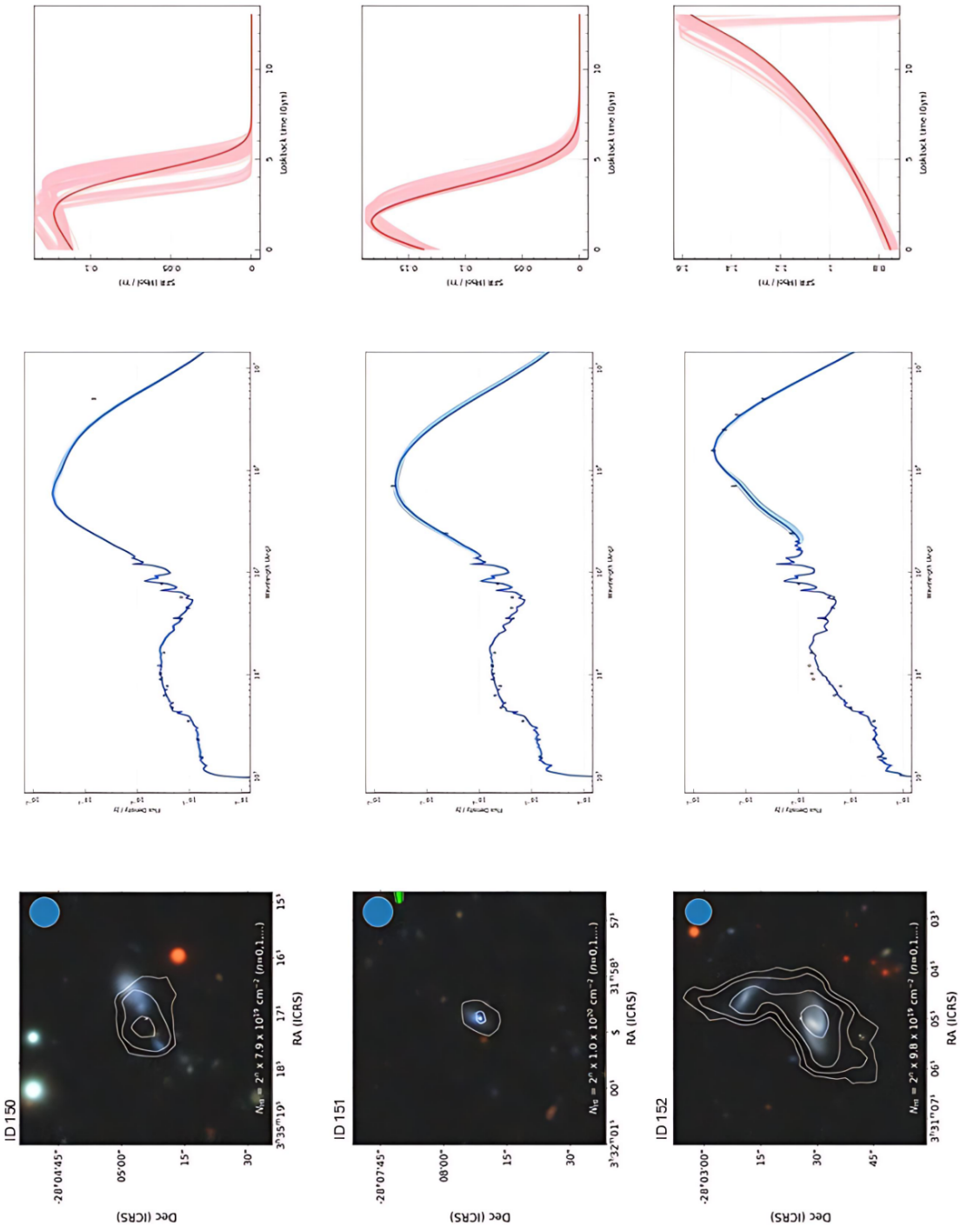
Continued on next page...



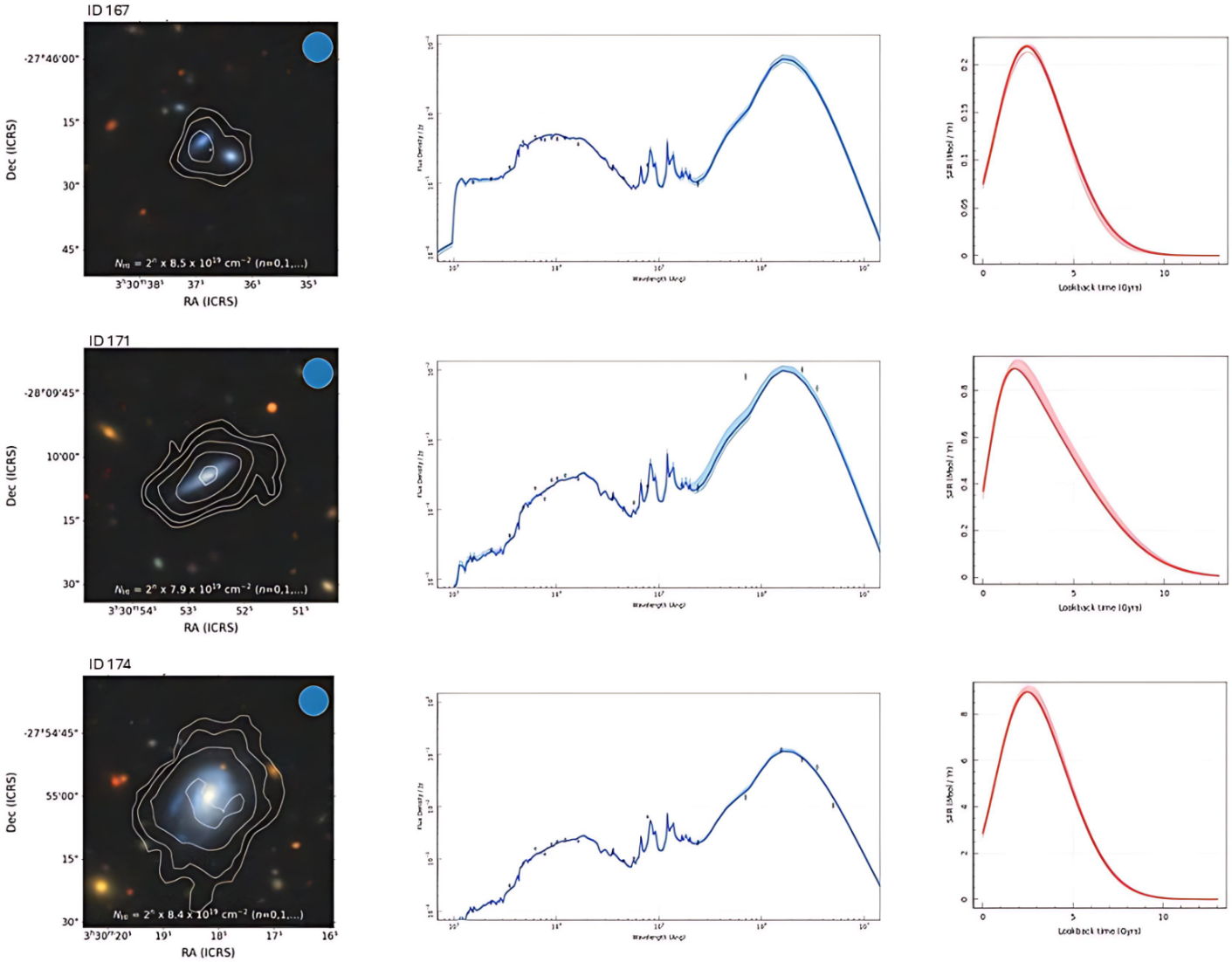
Continued on next page...



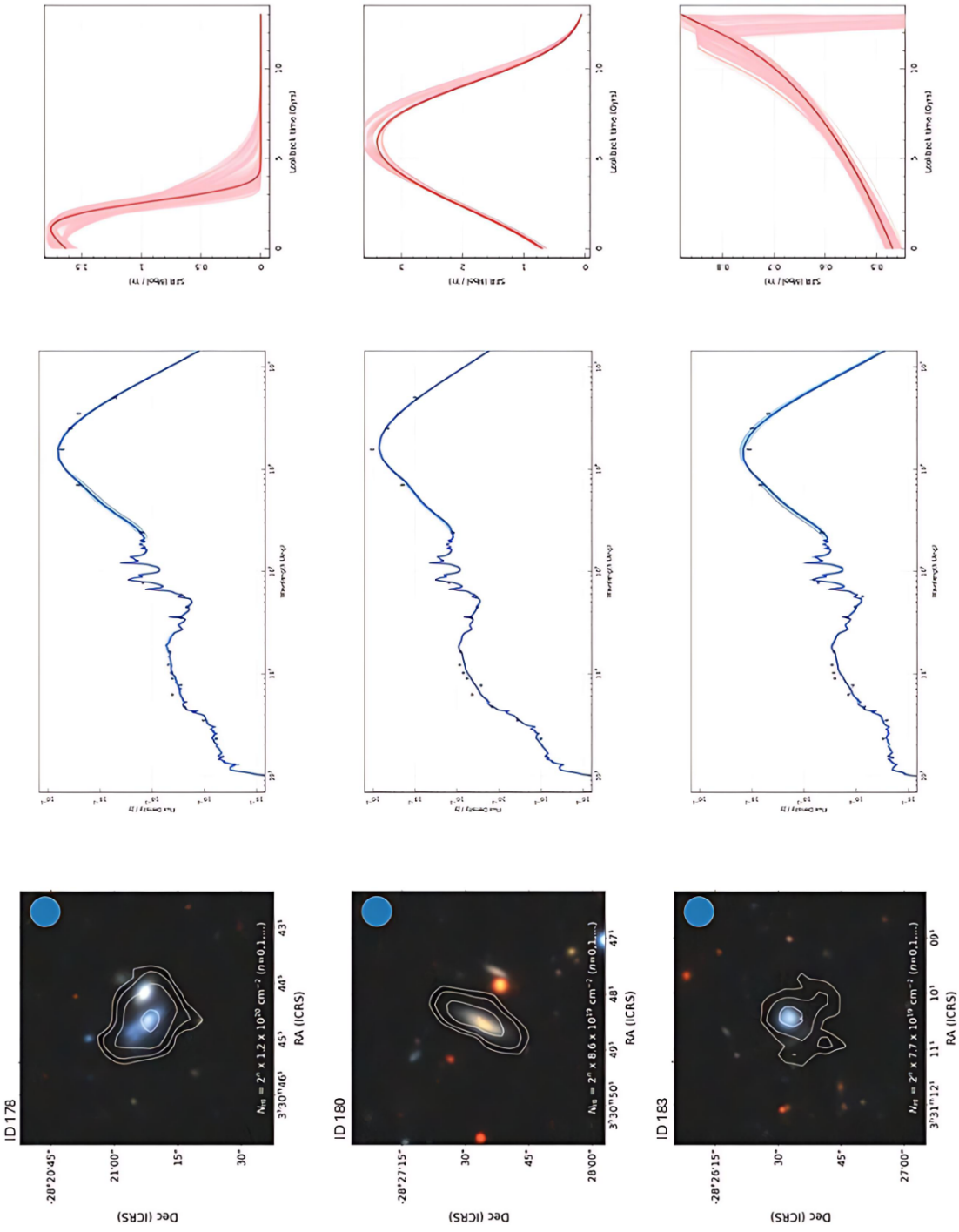
Continued on next page...



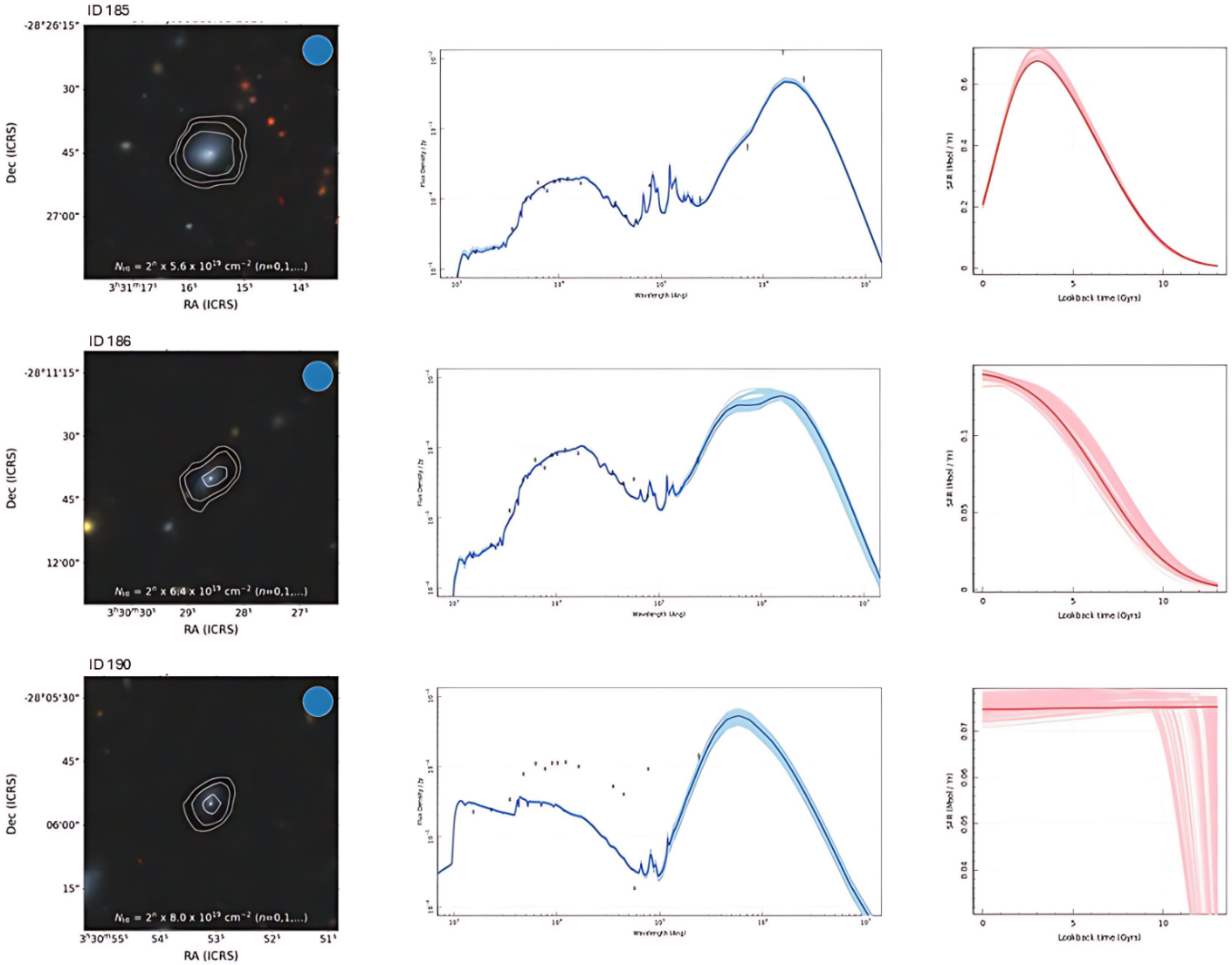
Continued on next page...



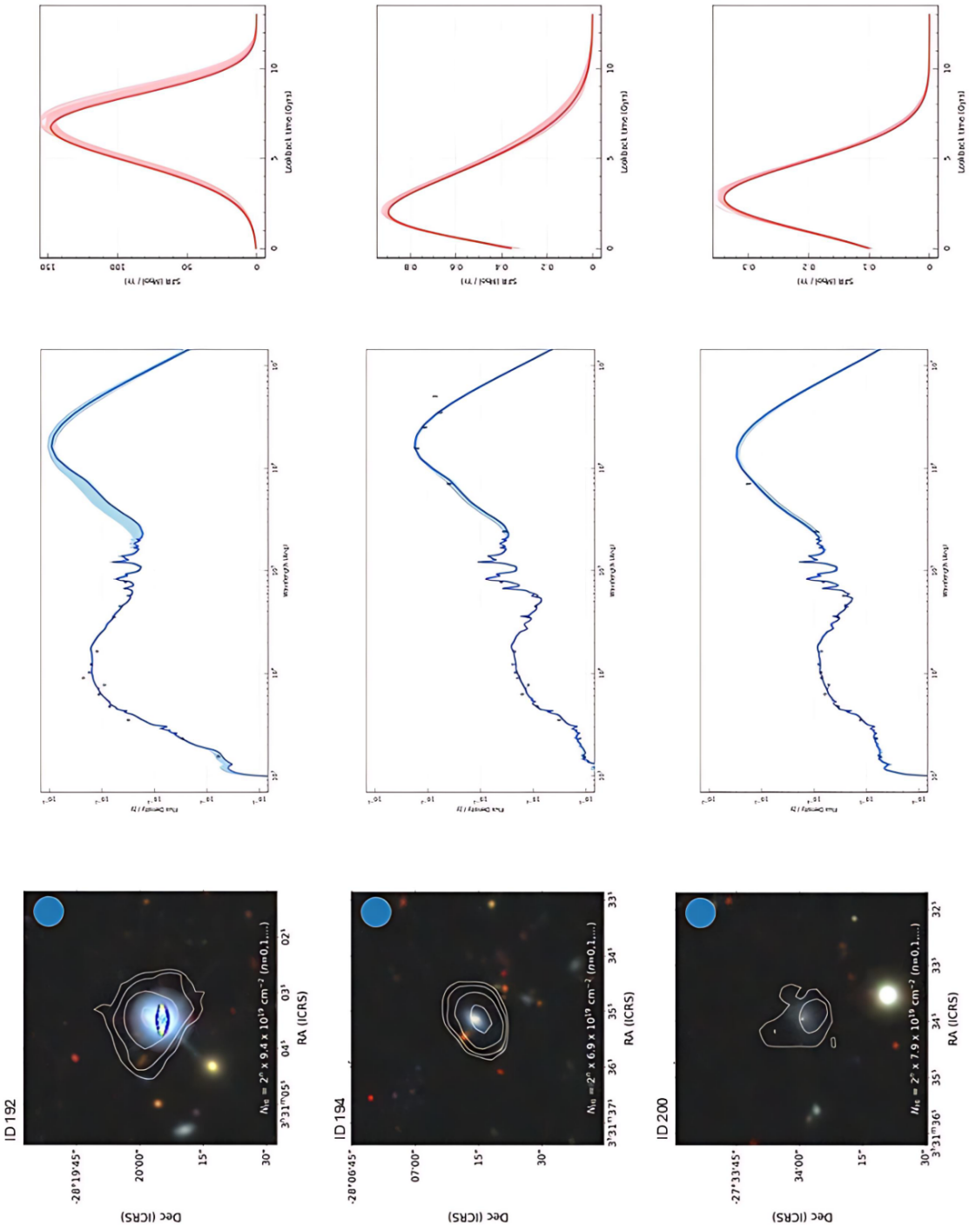
Continued on next page...



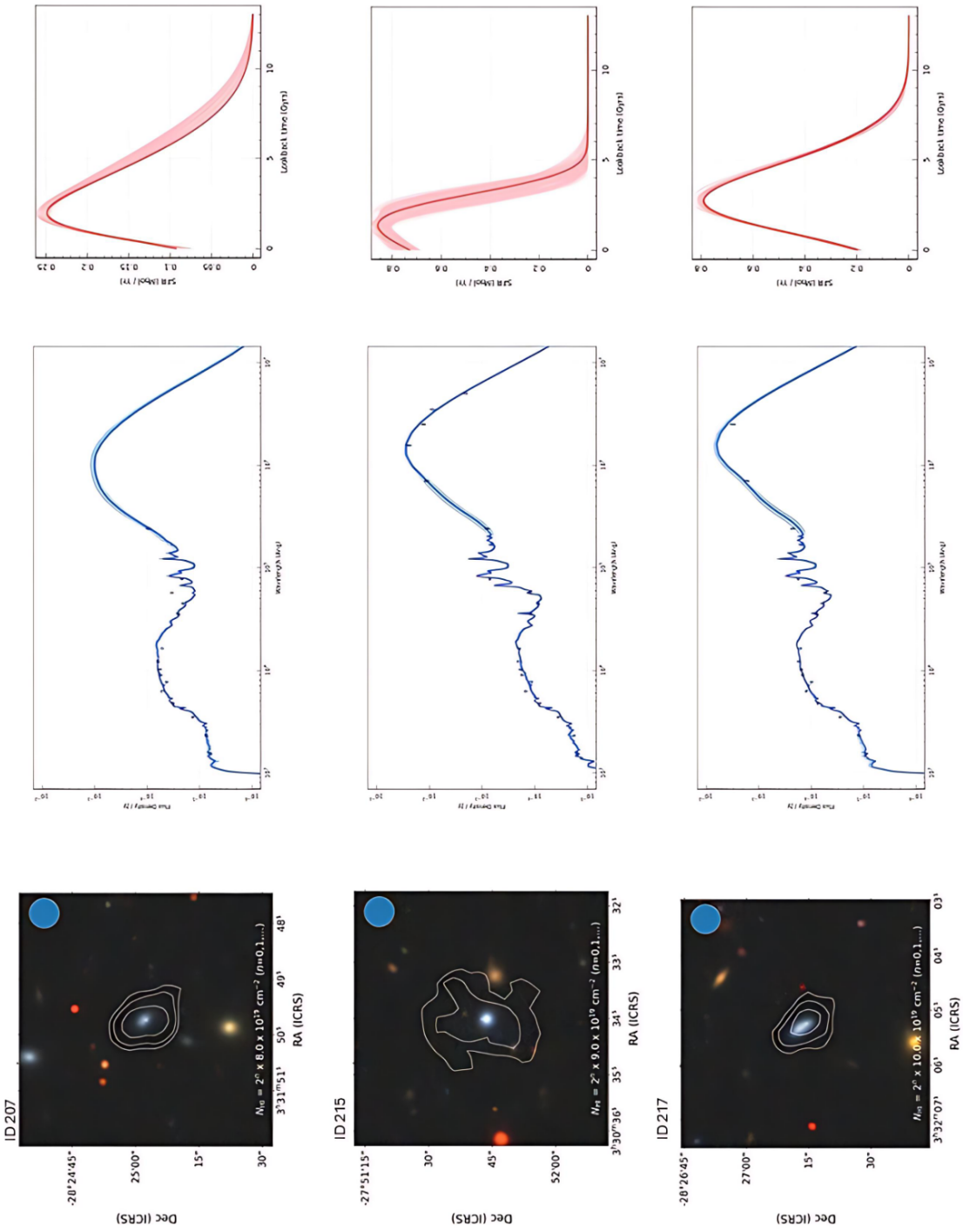
Continued on next page...



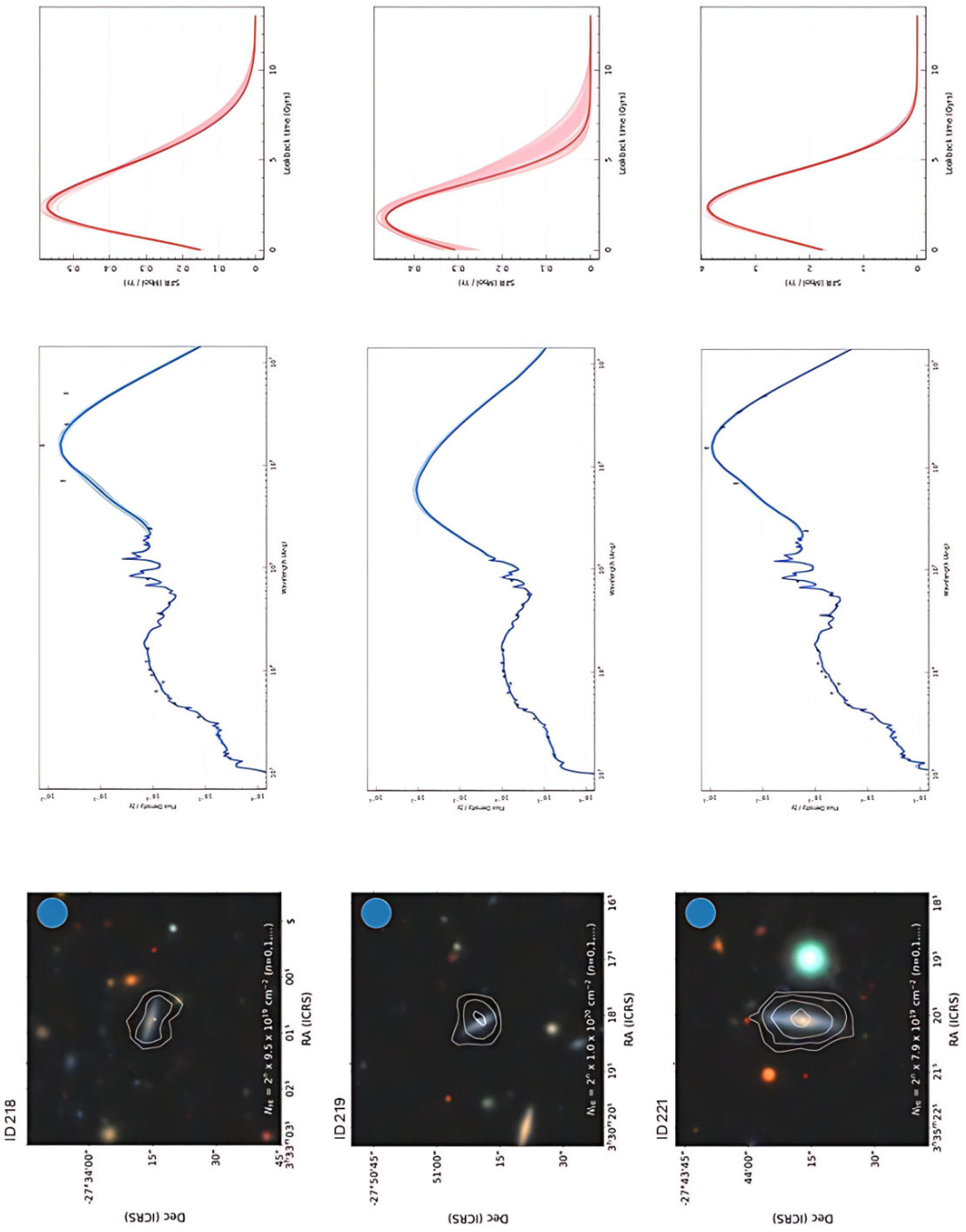
Continued on next page...



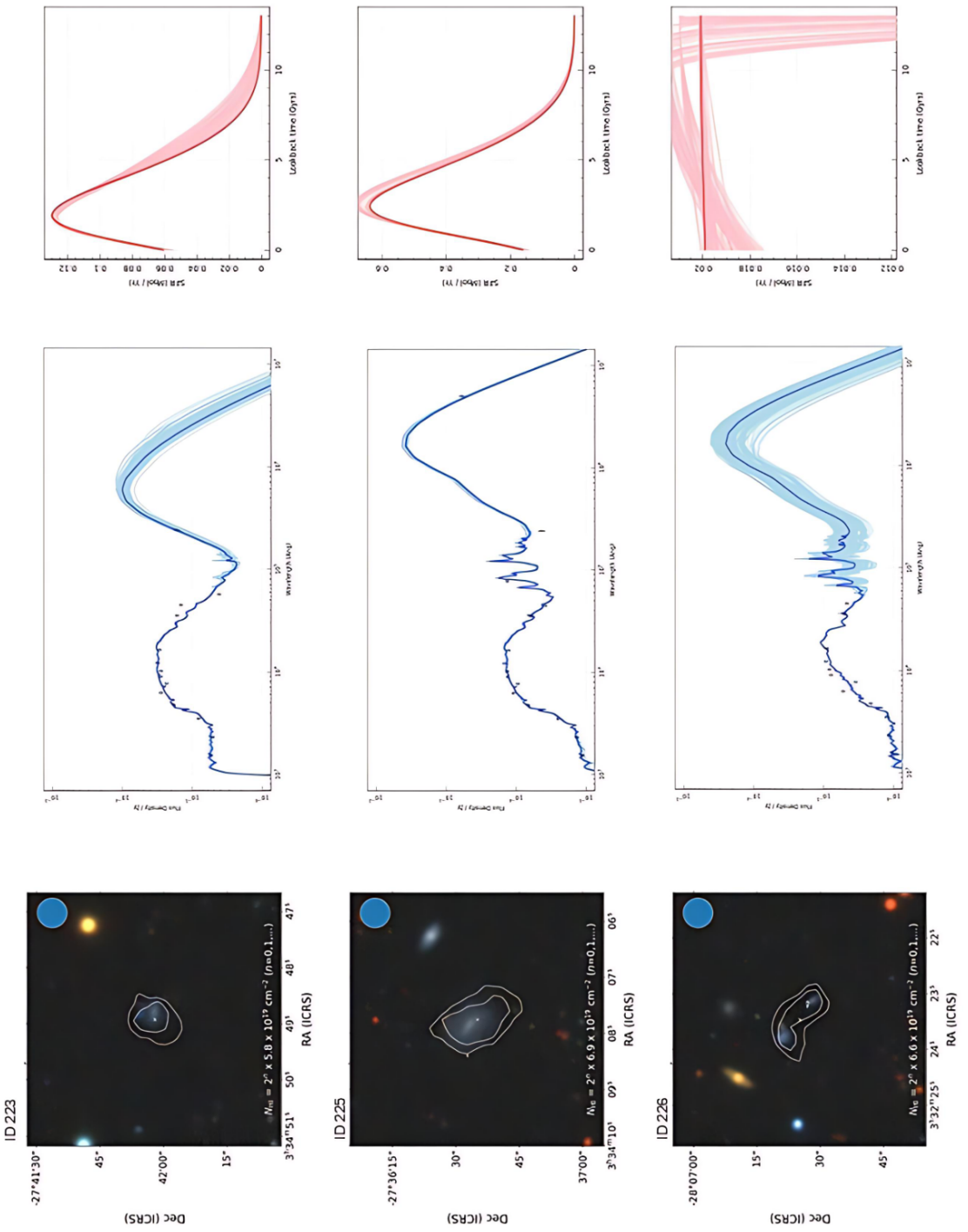
Continued on next page...



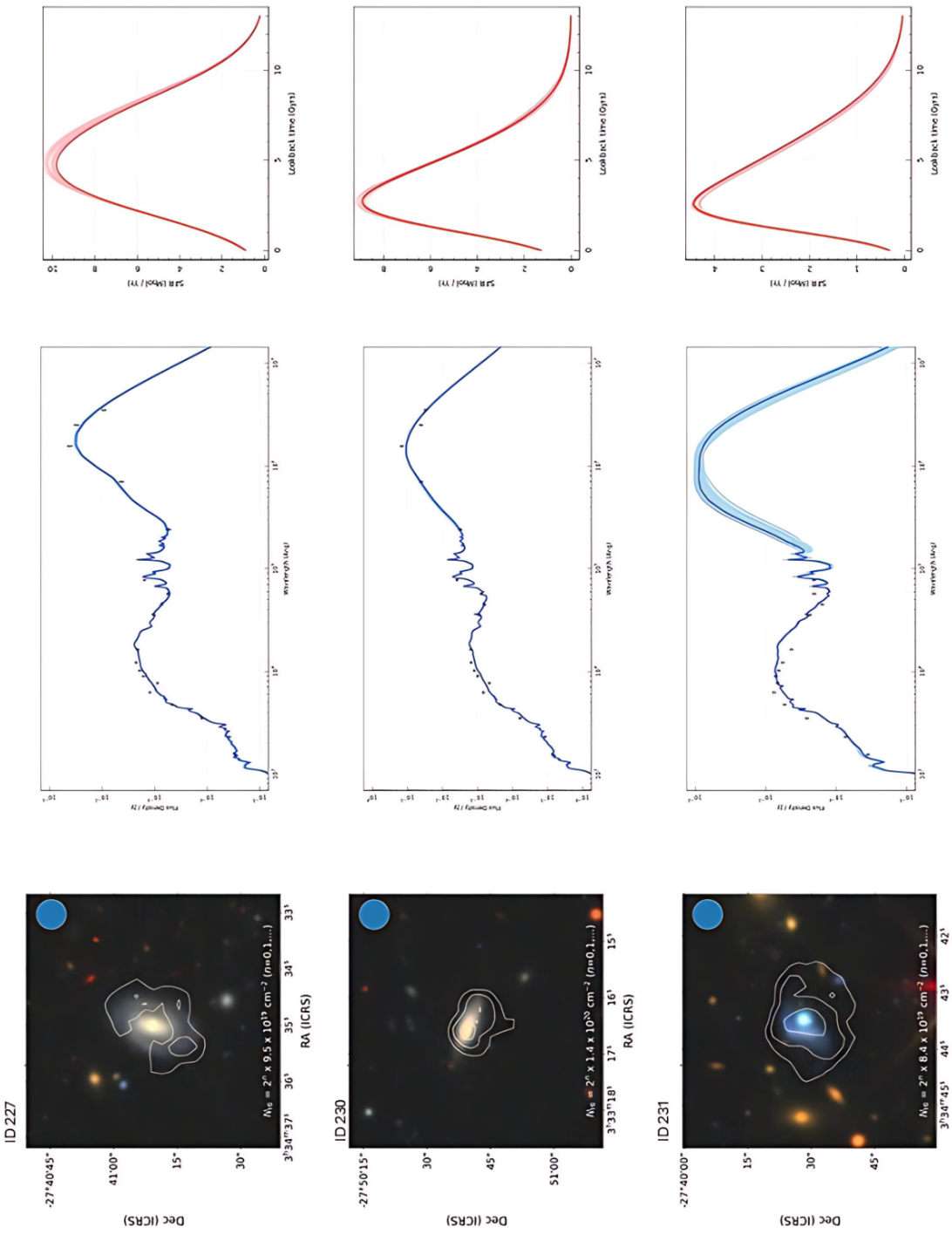
Continued on next page...



Continued on next page...



Continued on next page...



Continued on next page...

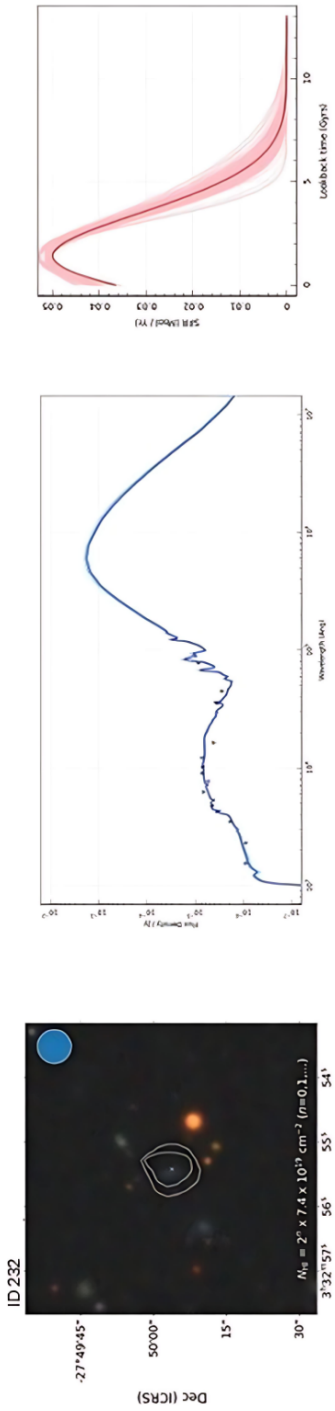
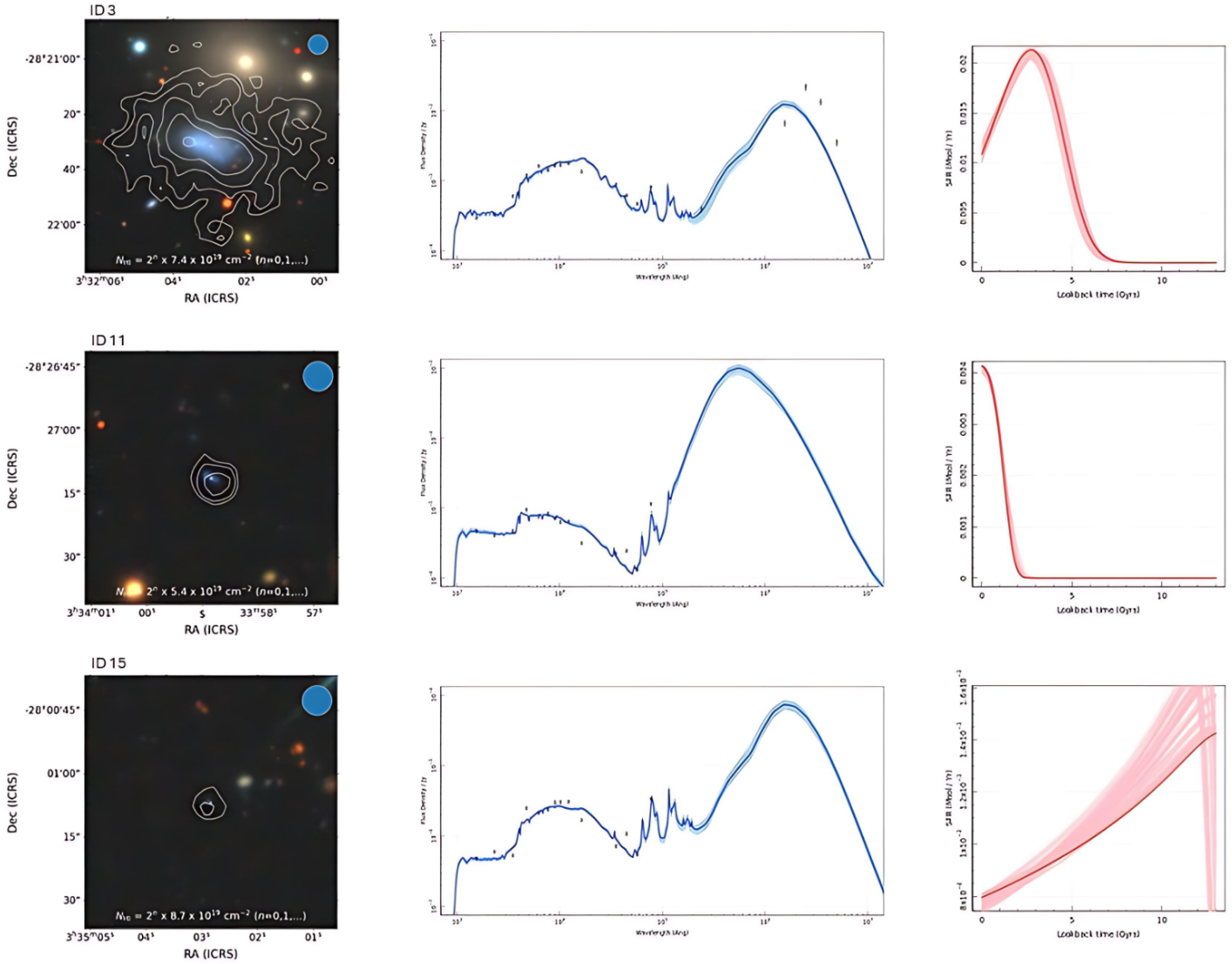
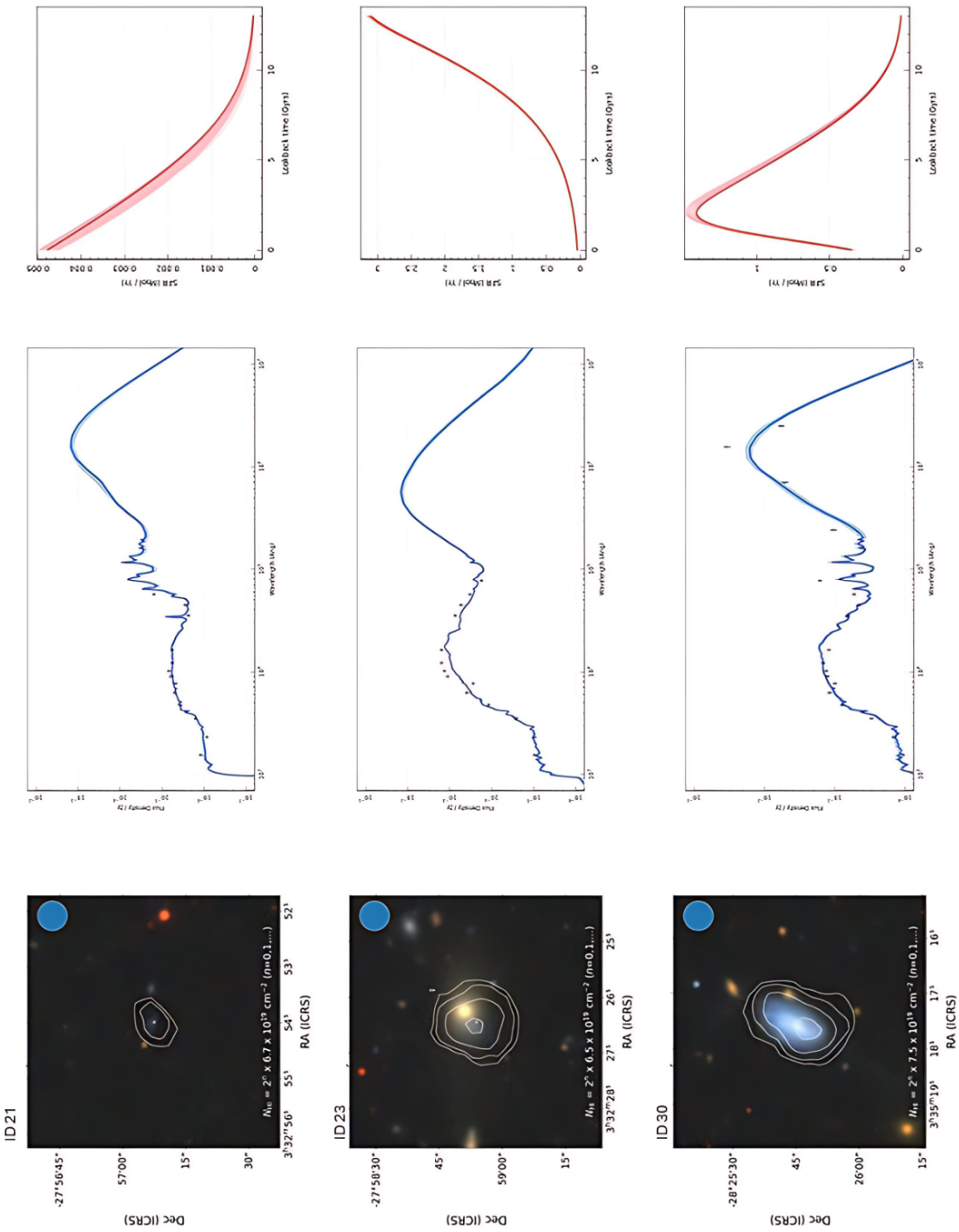

ID 232
Dec (ICRS)
RA (ICRS)
-27°49'45"
50'00"
15"
30"
3h32m57s
56s
55s
54s
0
0.01
0.02
0.03
0.04
0.05
5
10
Lookback time (Gyrs)

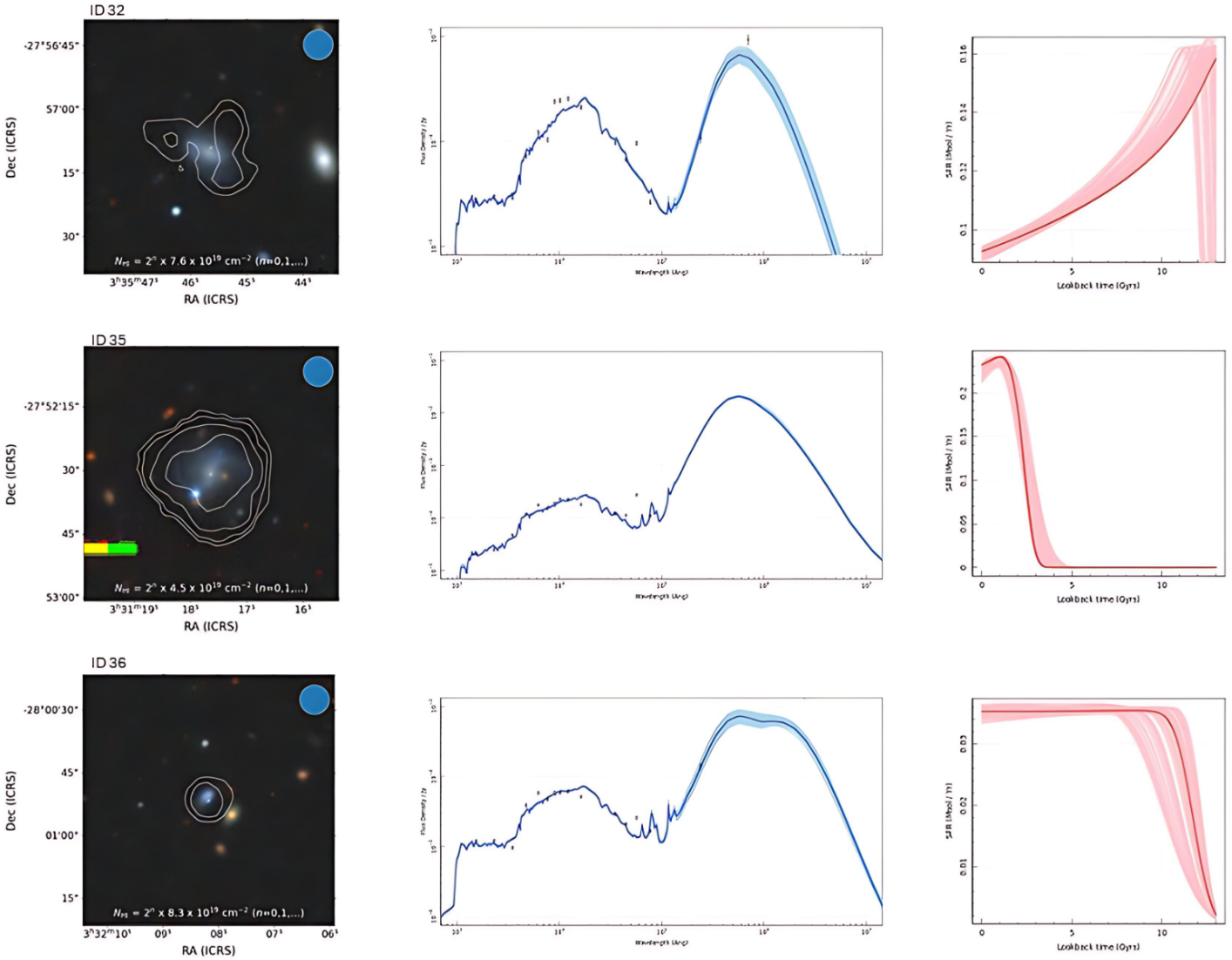
Figure A.2: DECaLS image overlaid with HI contours (left panel), SED (middle panel) and SFH (right panel) of galaxies where $5 < \chi^2 < 10$.



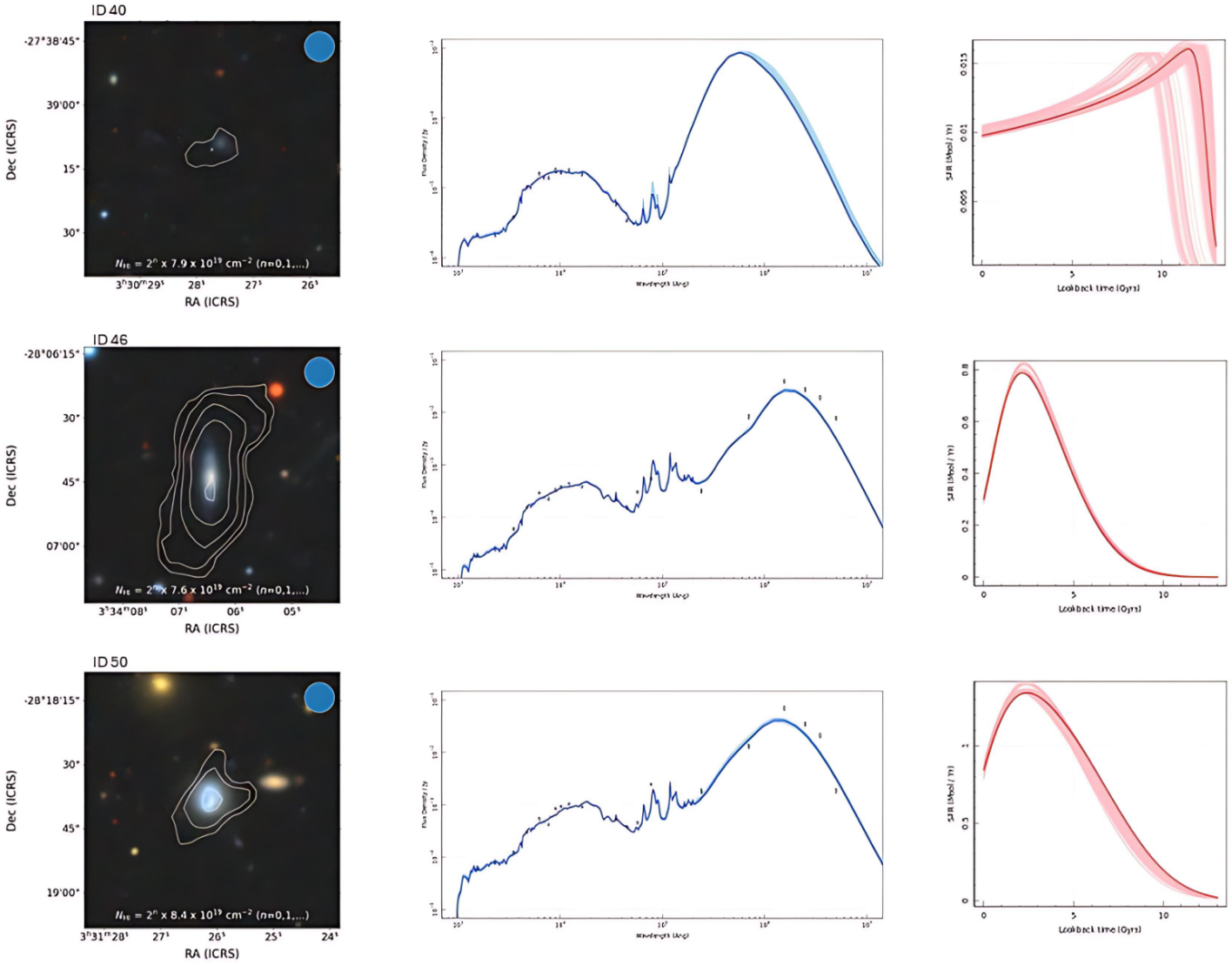
Continued on next page...



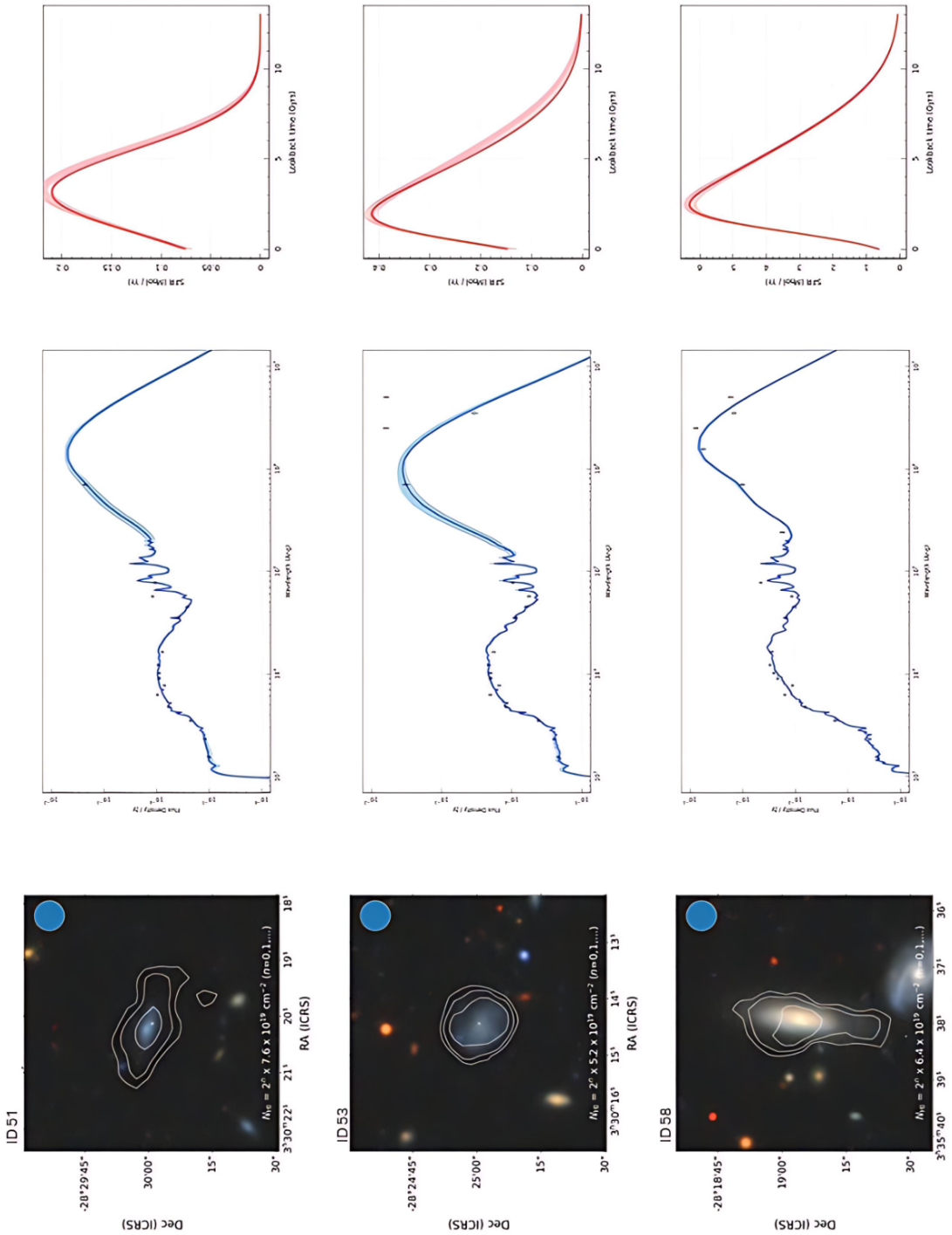
Continued on next page...



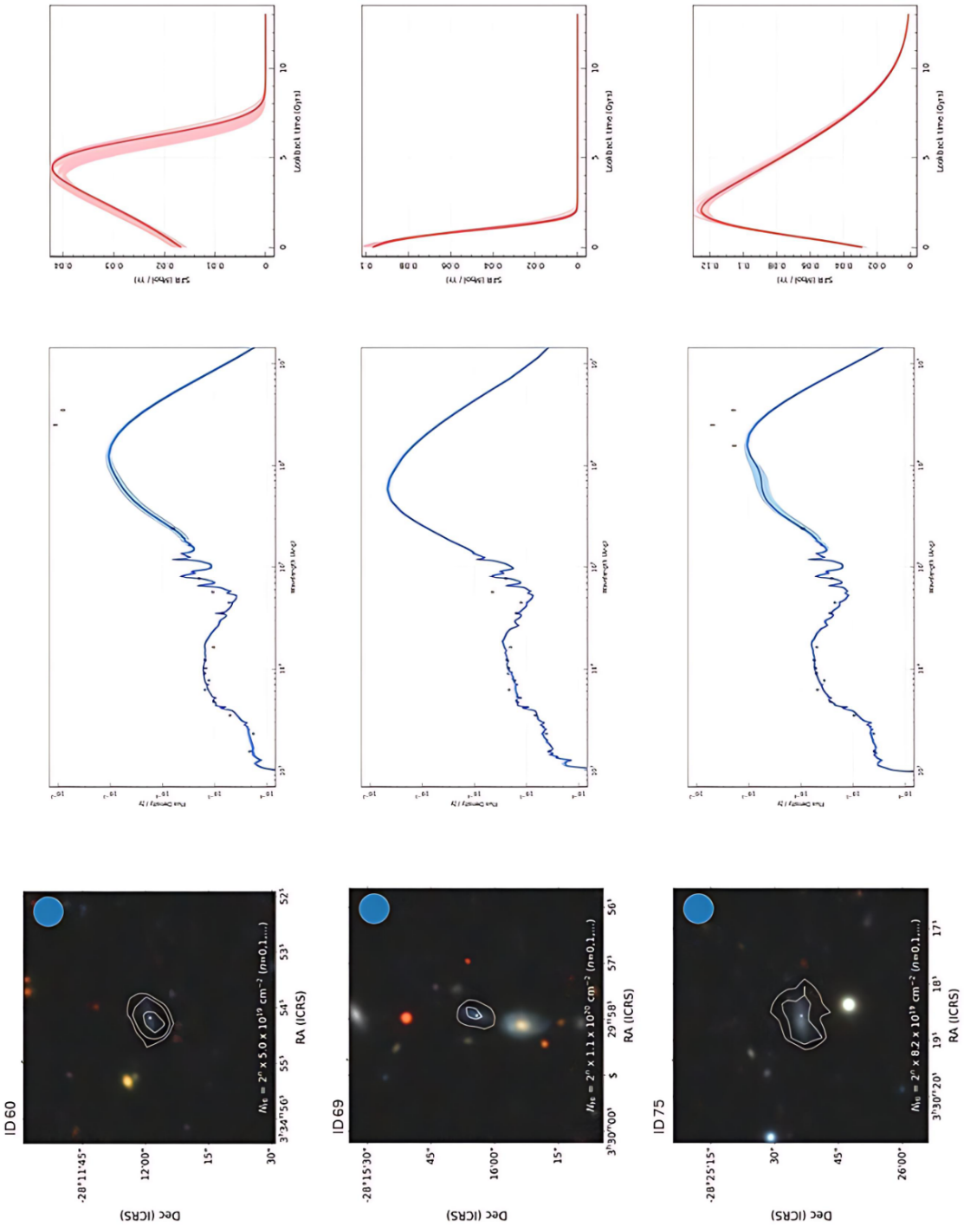
Continued on next page...



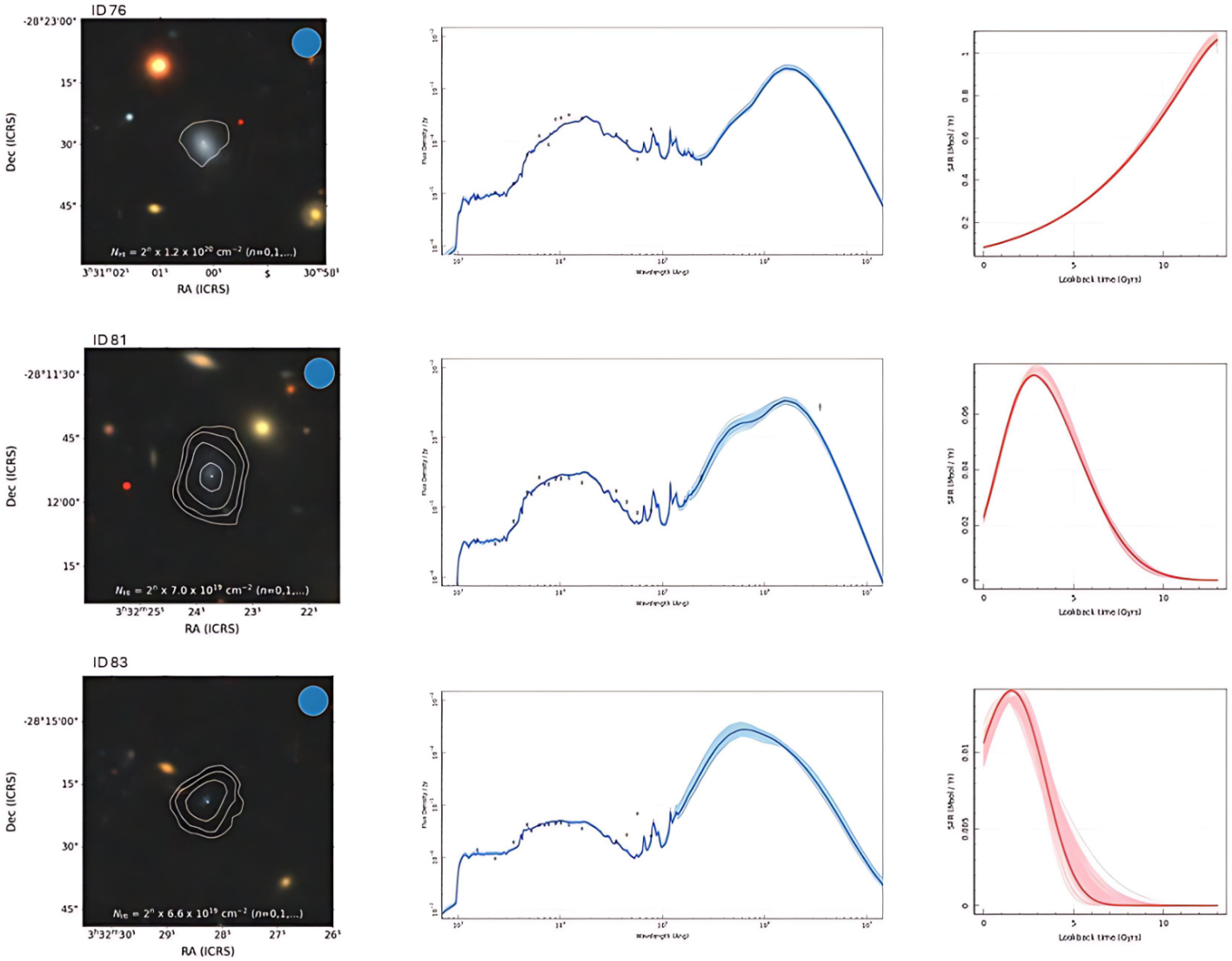
Continued on next page...



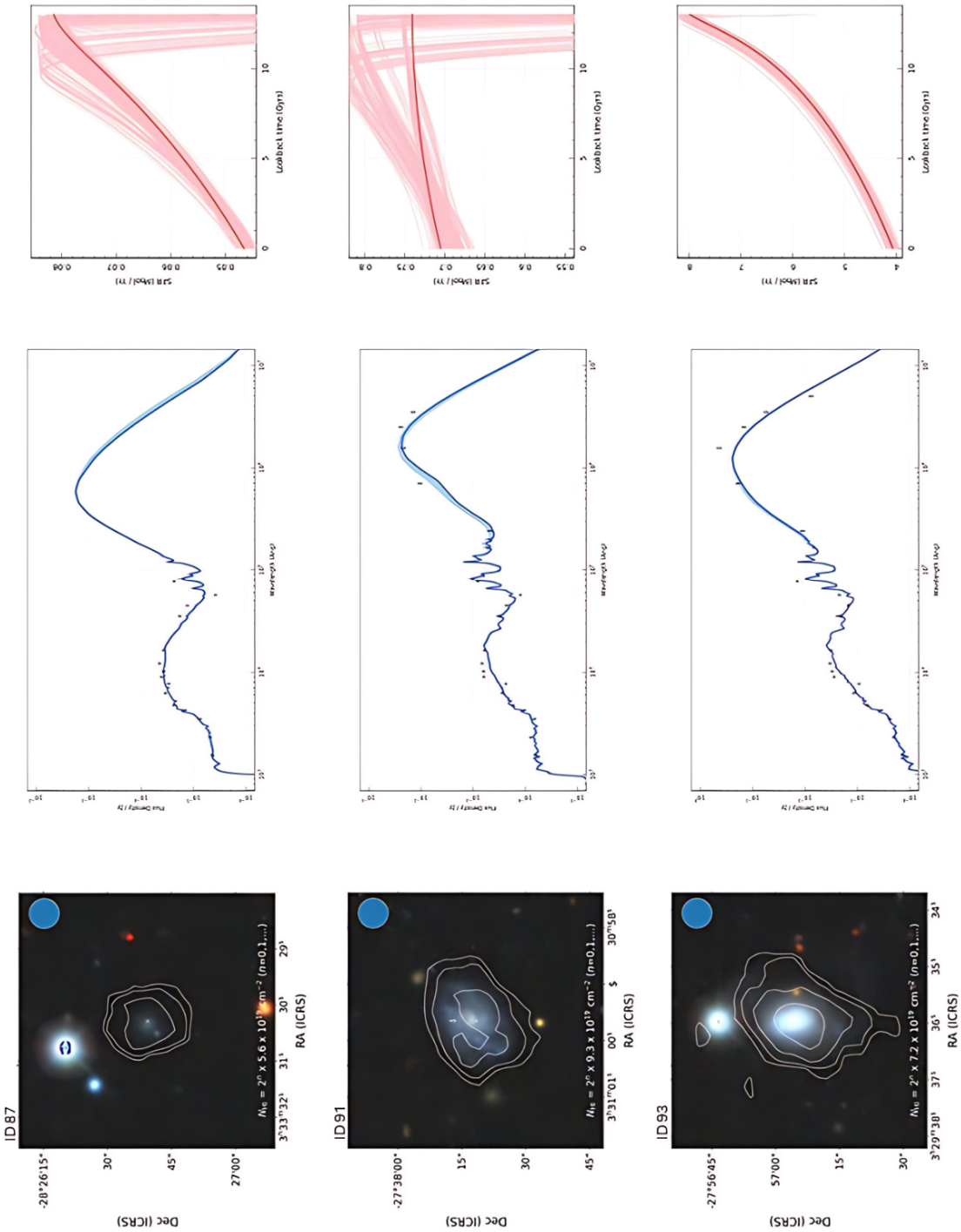
Continued on next page...



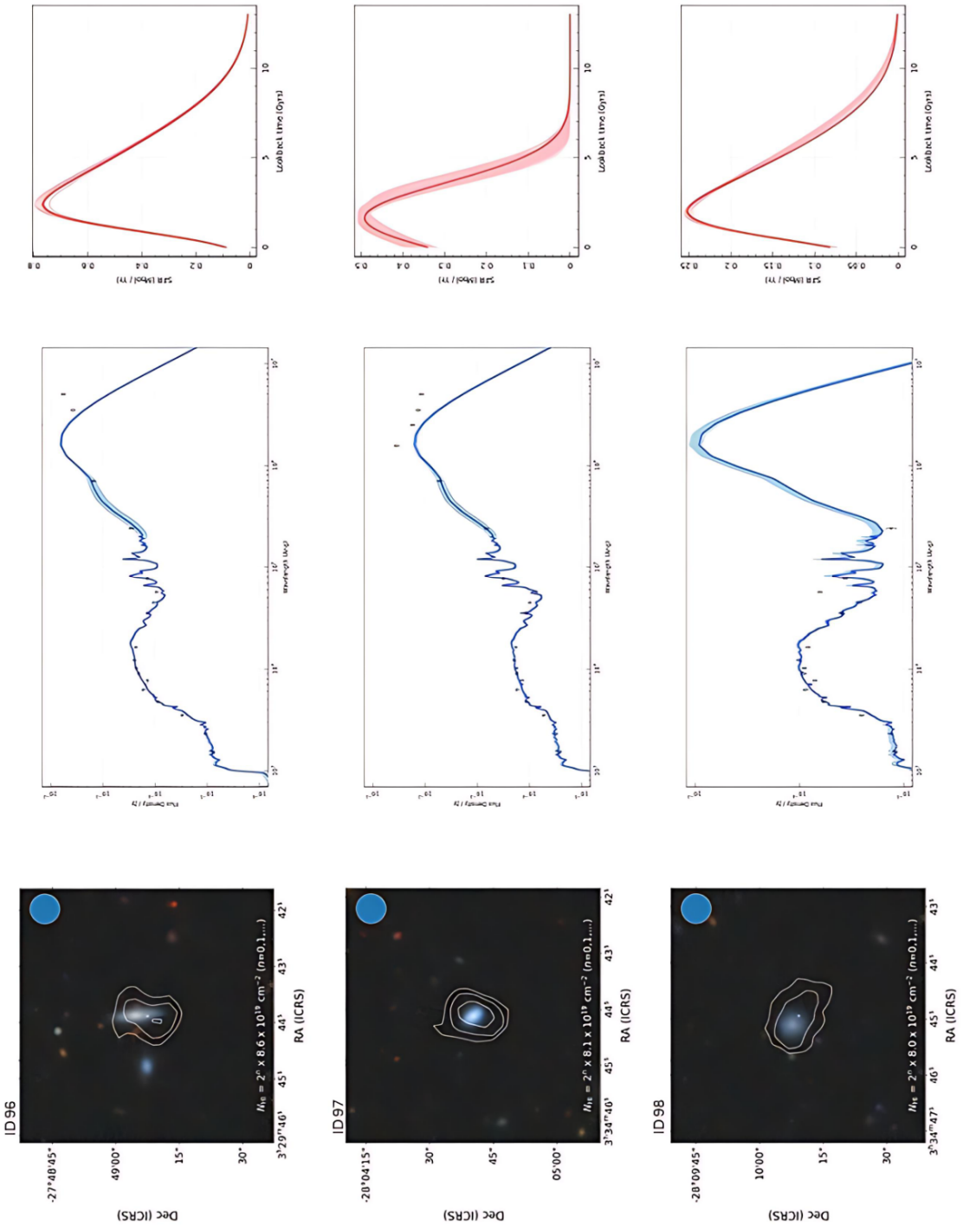
Continued on next page...



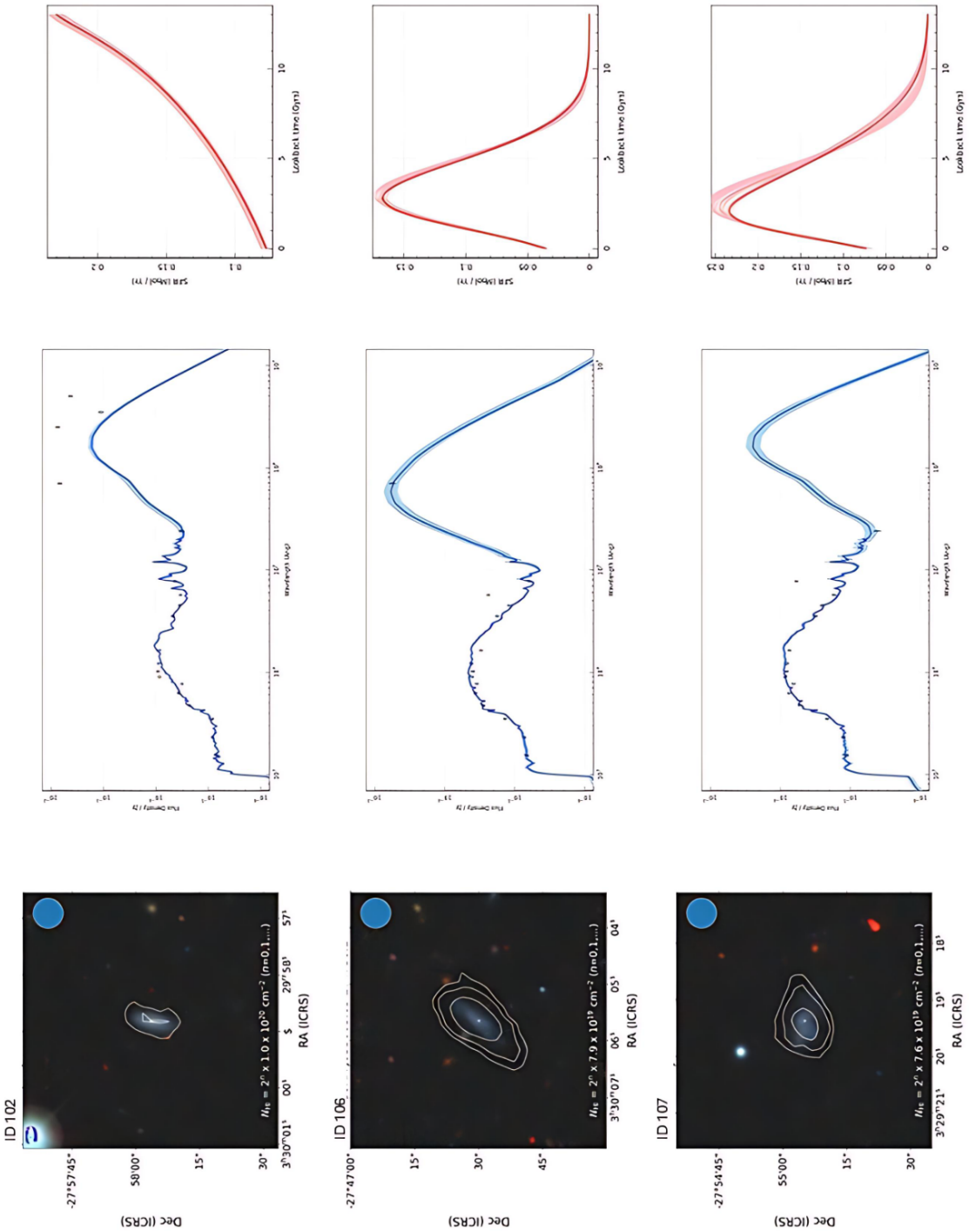
Continued on next page...



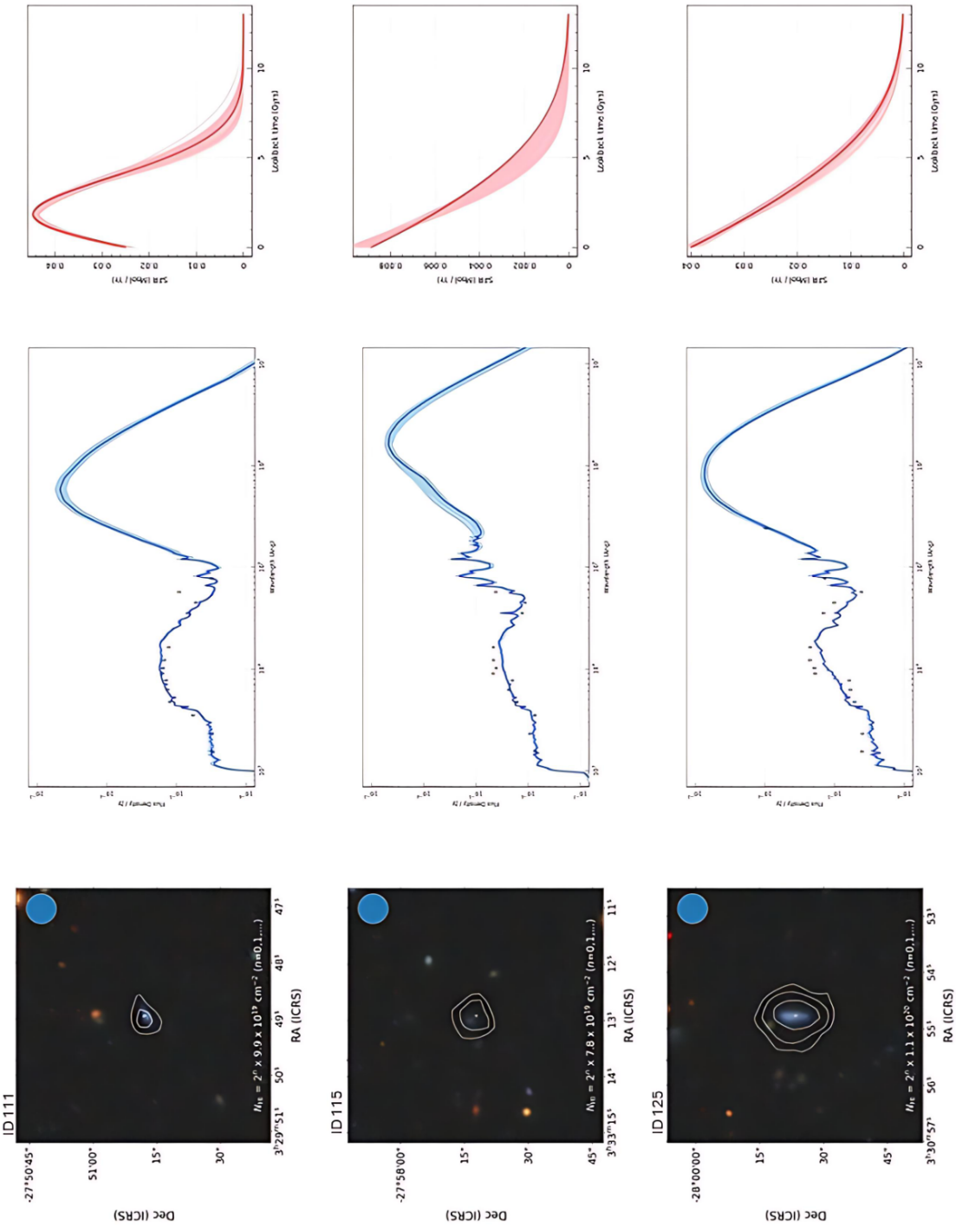
Continued on next page...



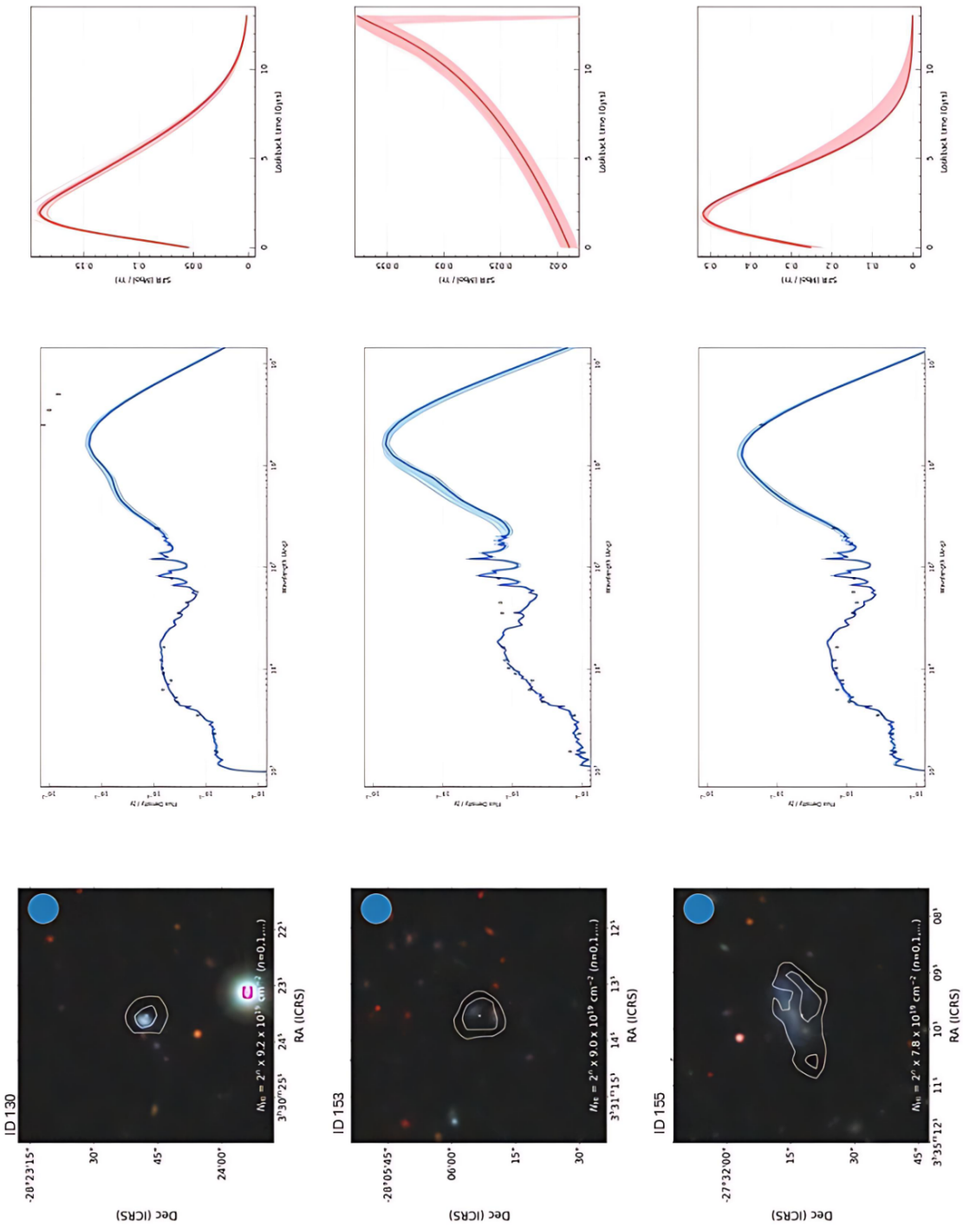
Continued on next page...



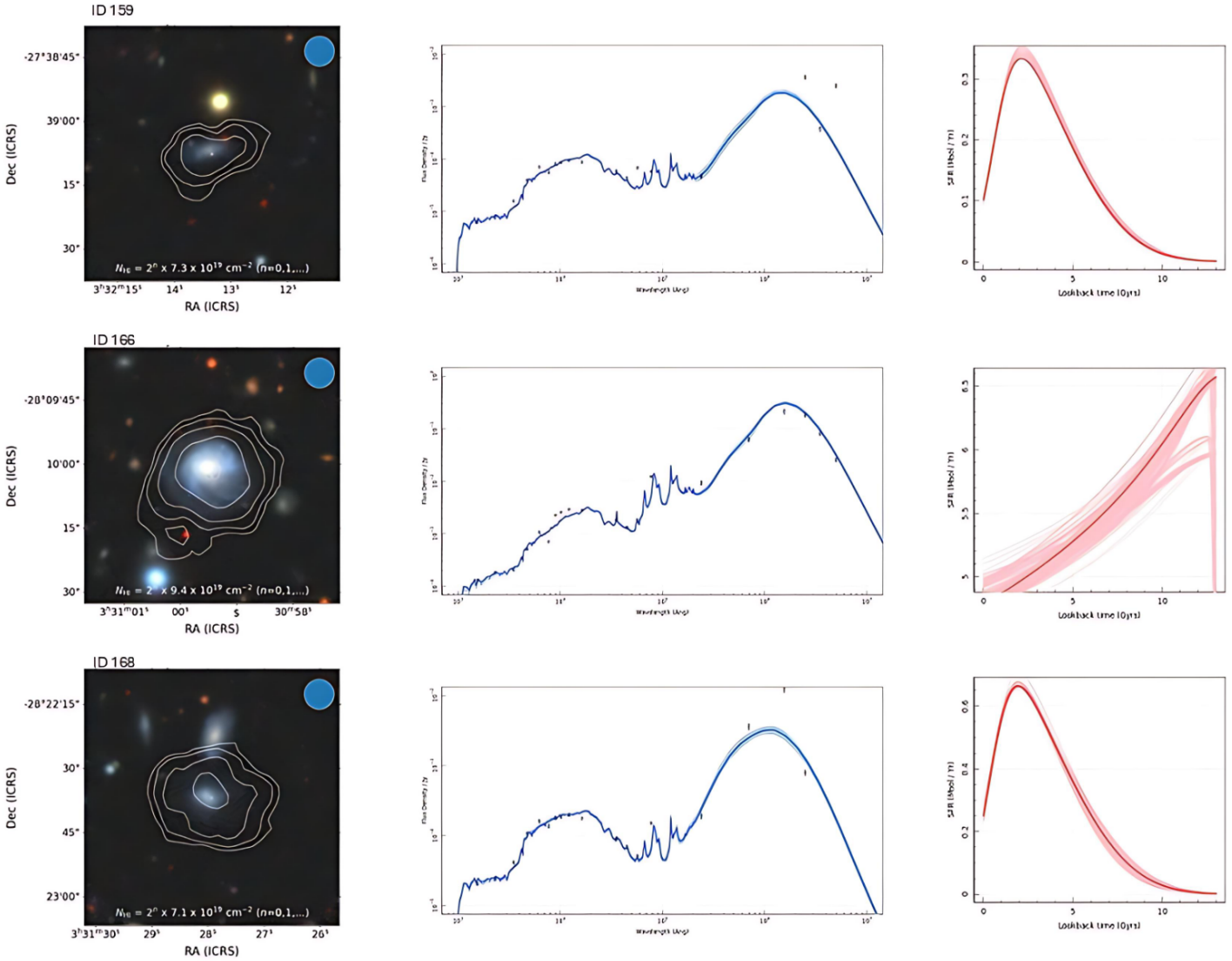
Continued on next page...



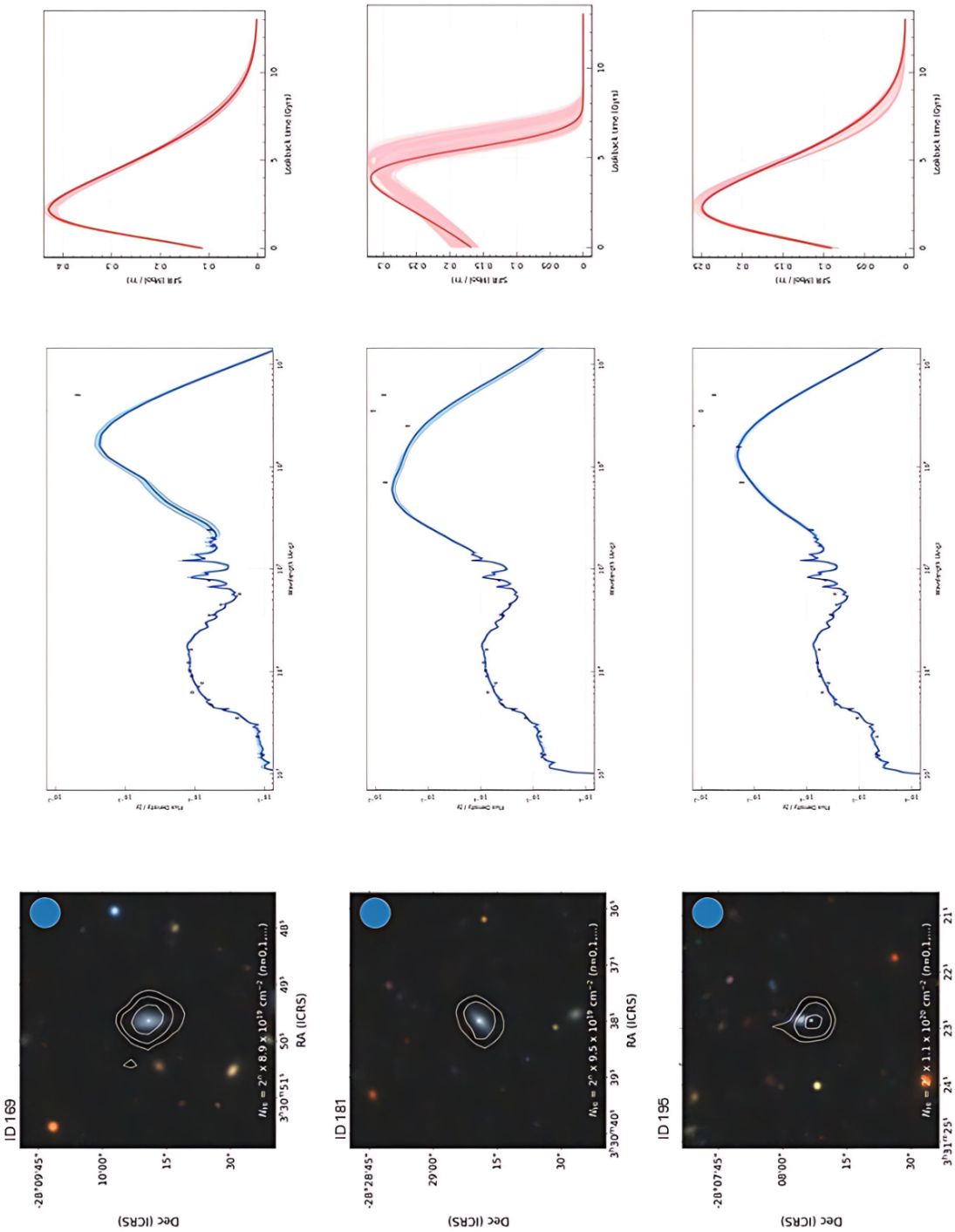
Continued on next page...



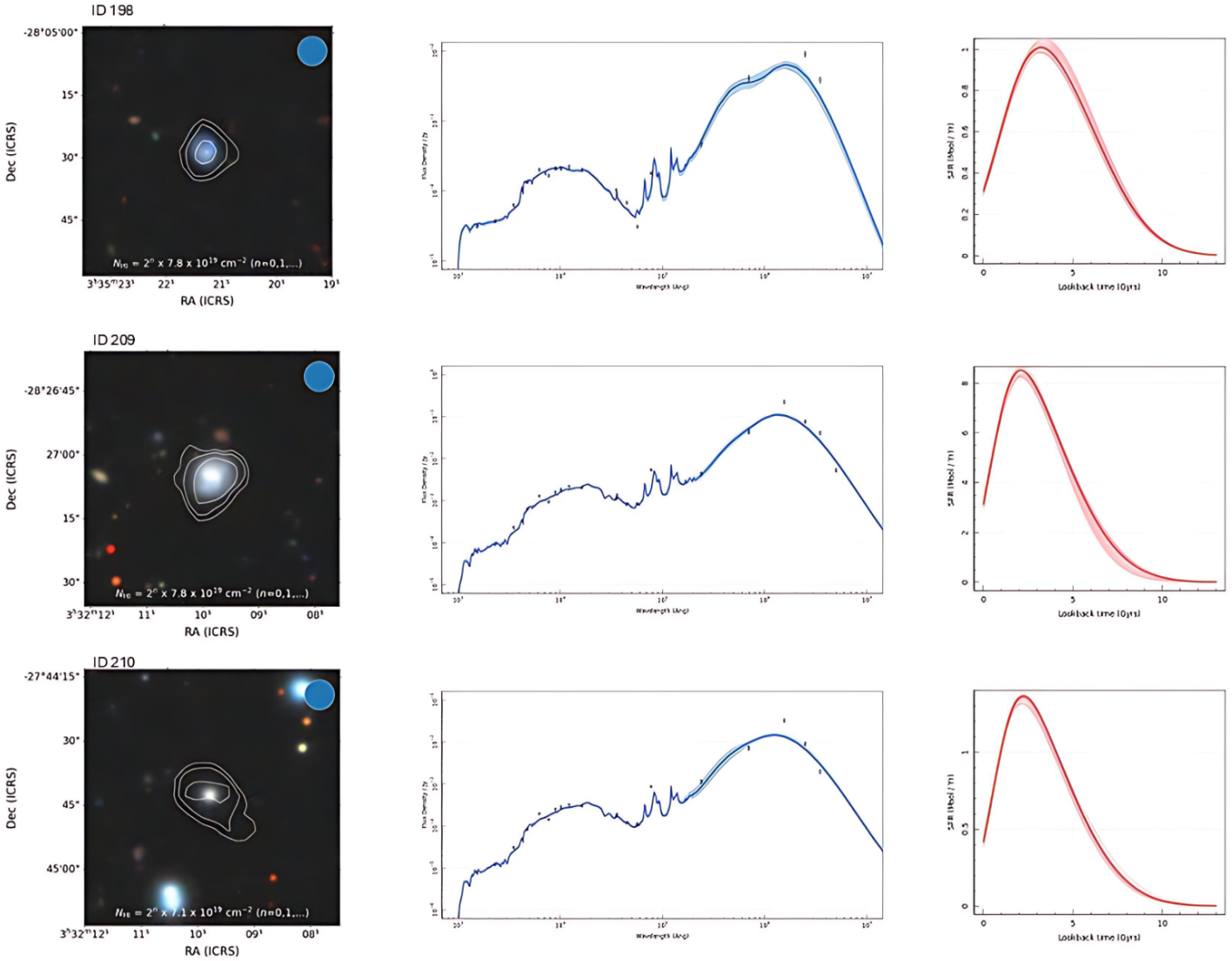
Continued on next page...



Continued on next page...



Continued on next page...



Continued on next page...

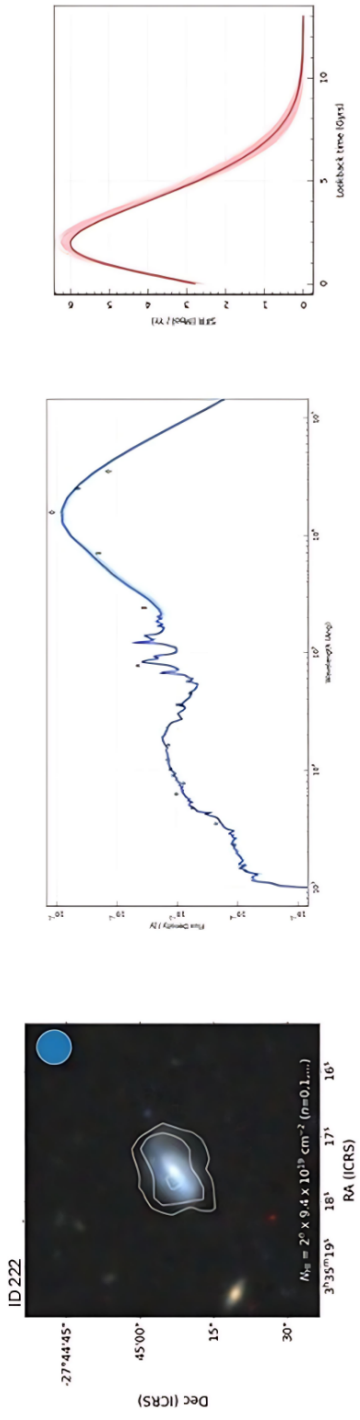
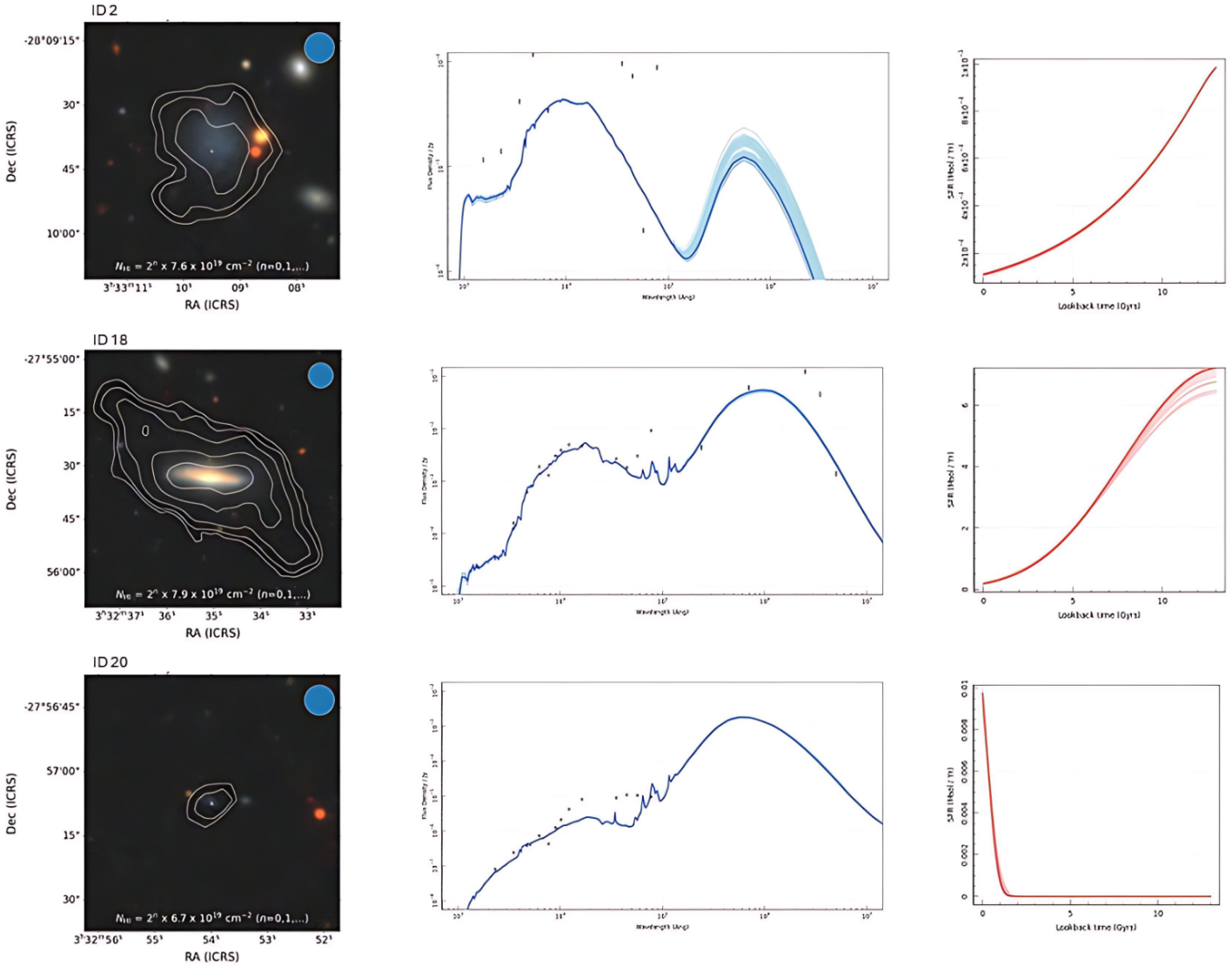
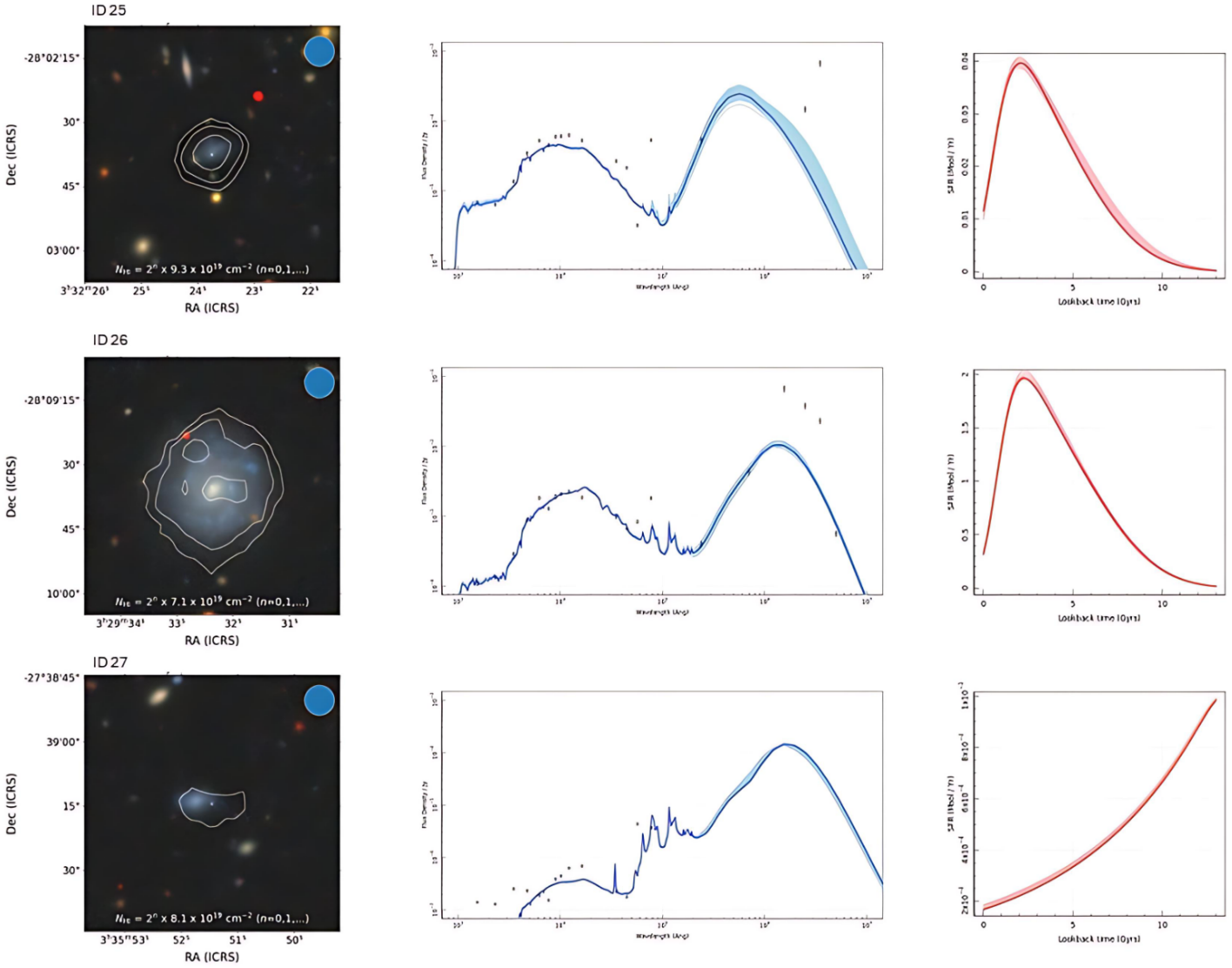

ID 222
Dec (ICRS)
-27°44'45"
45'00"
15"
30"
RA (ICRS)
3h35m19s
18s
17s
16s
Lookback time (Gyrs)

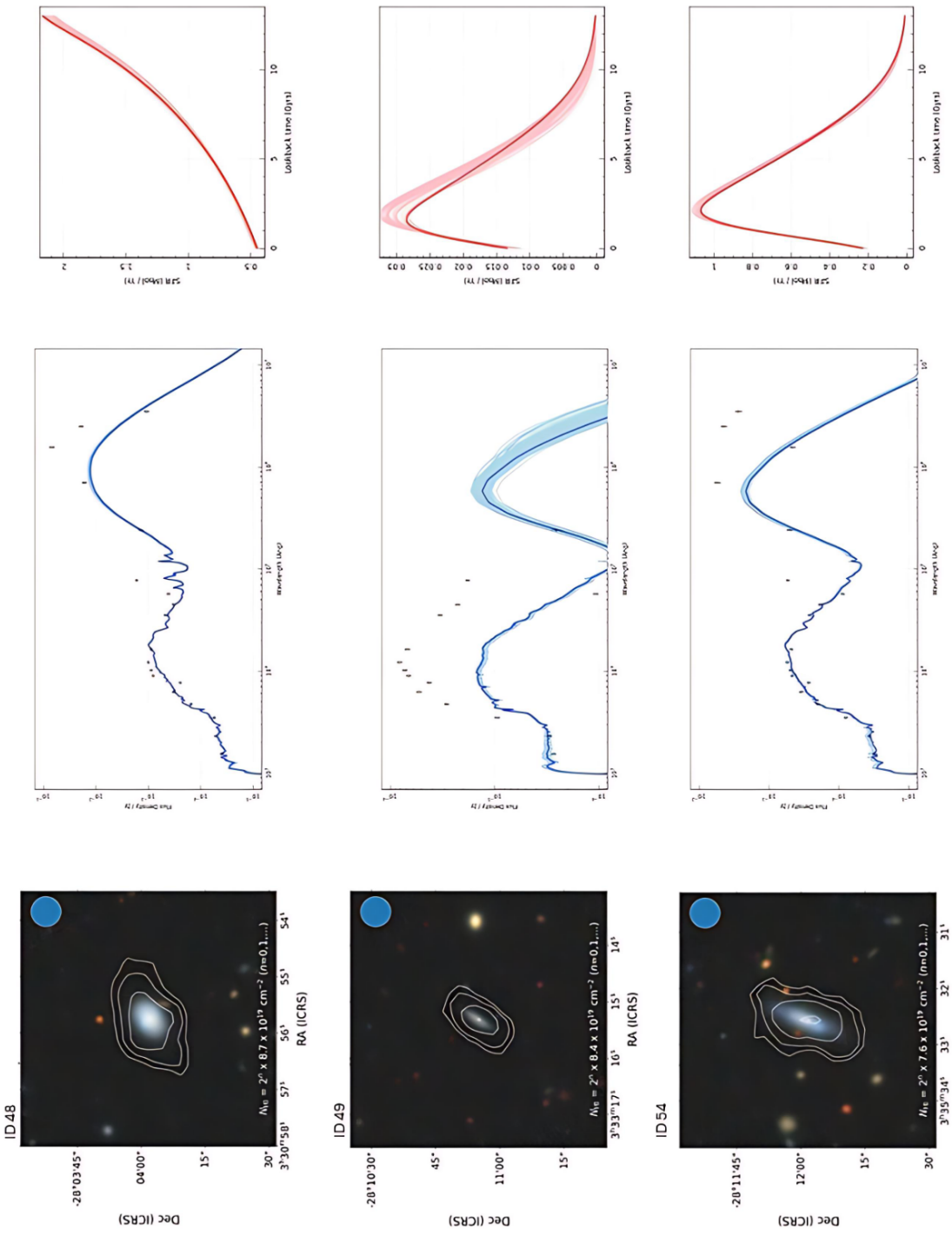
Figure A.3: DECaLS image overlaid with HI contours (left panel), SED (middle panel) and SFH (right panel) of galaxies where $\chi^2 > 10$.



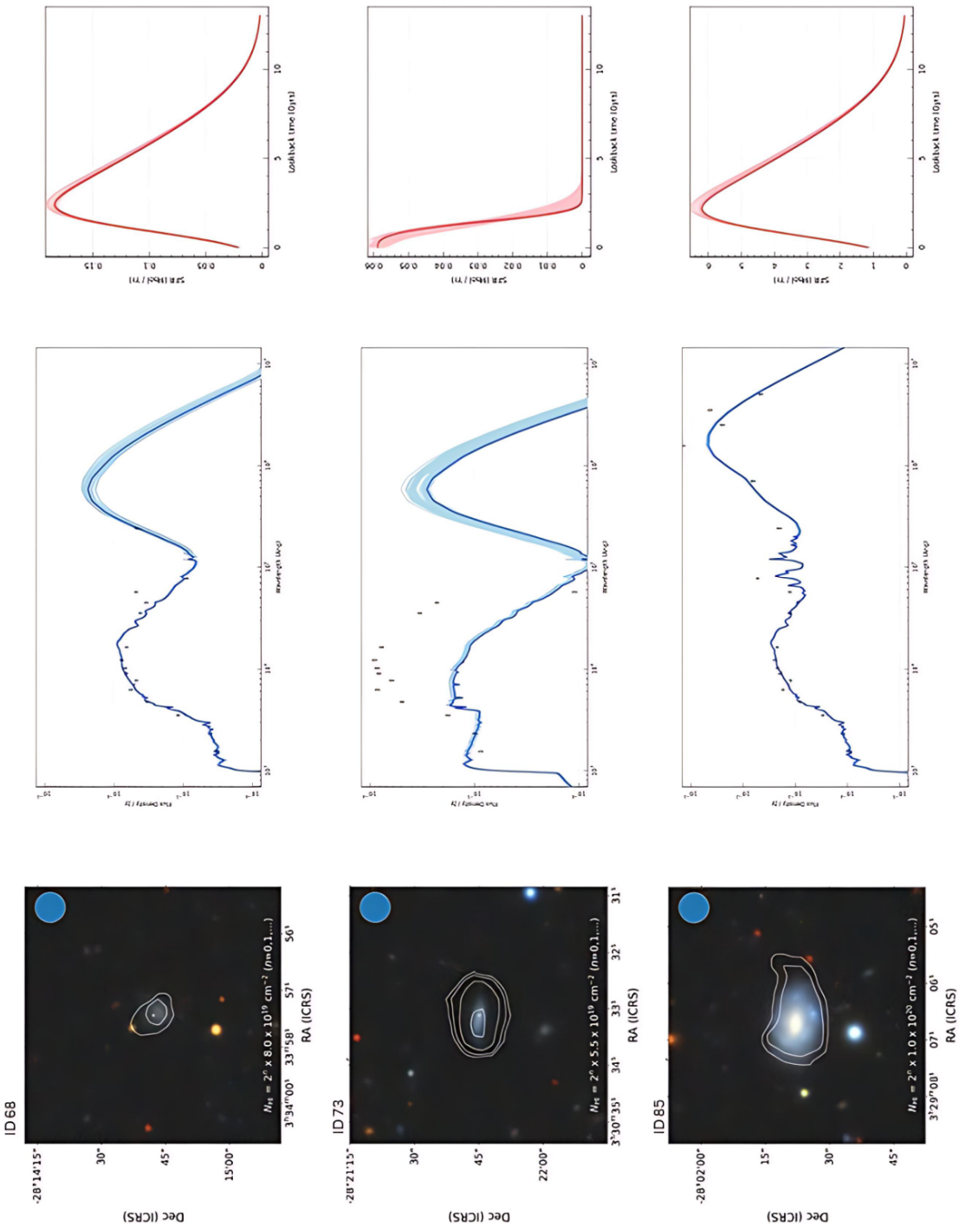
Continued on next page...



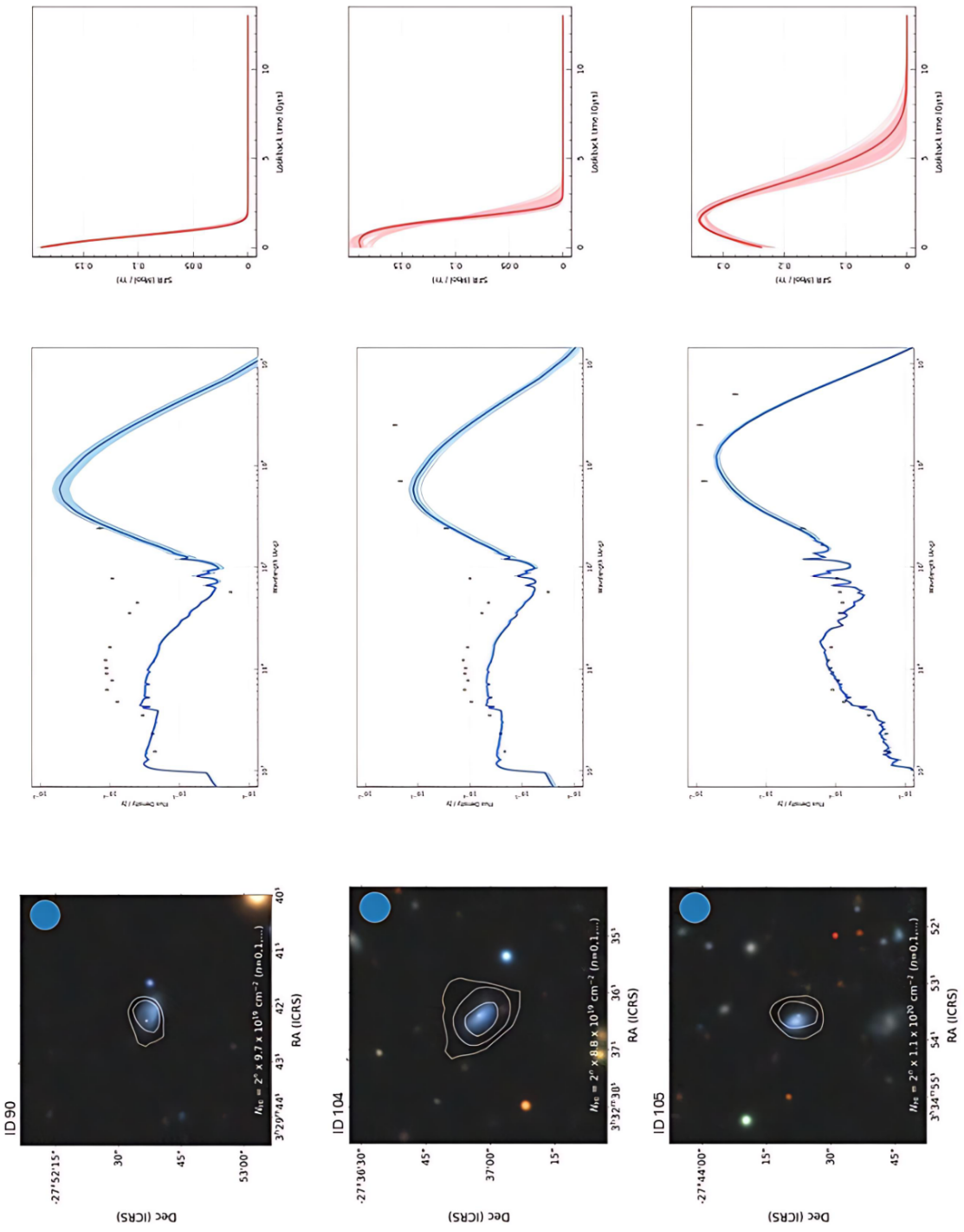
Continued on next page...



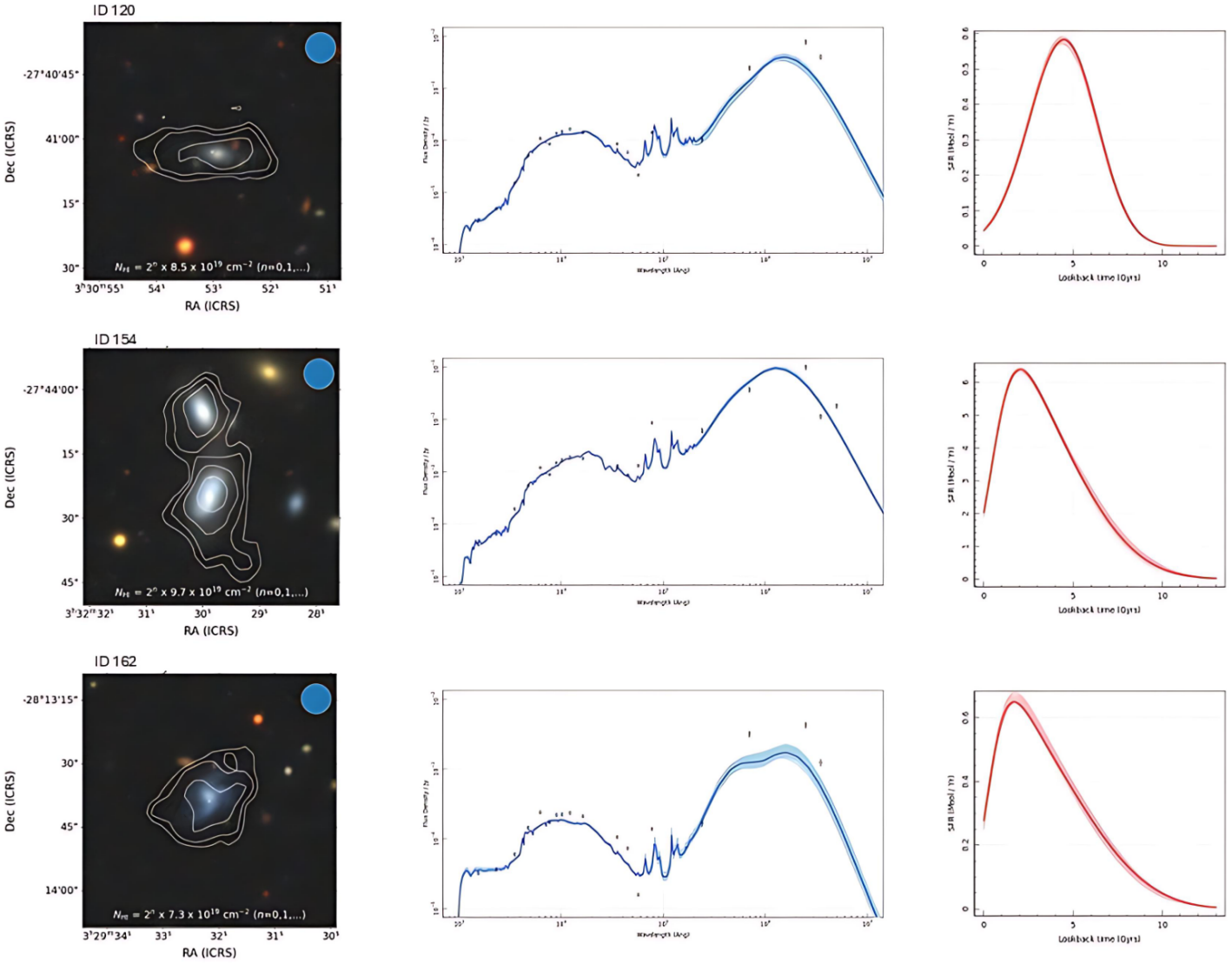
Continued on next page...



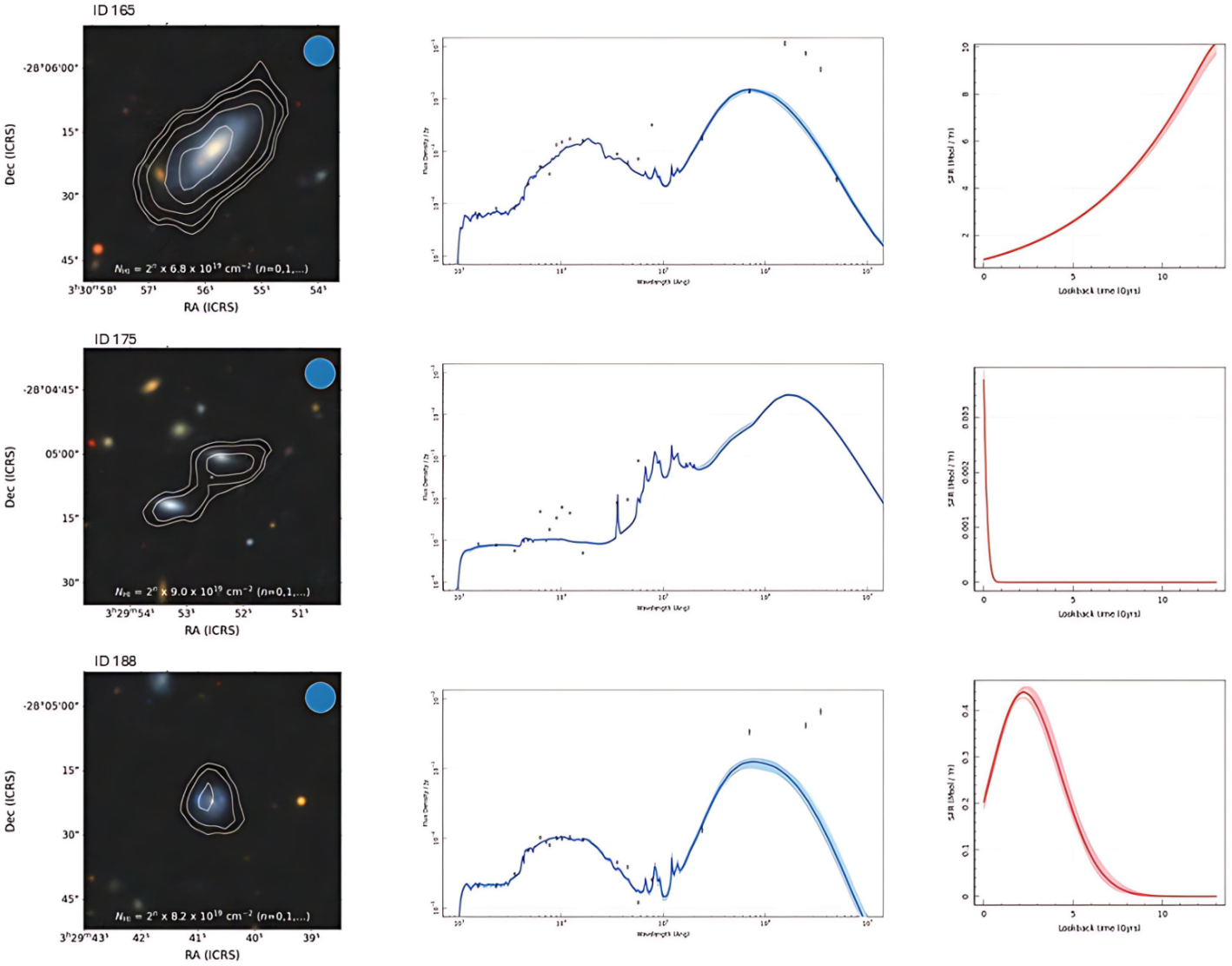
Continued on next page...



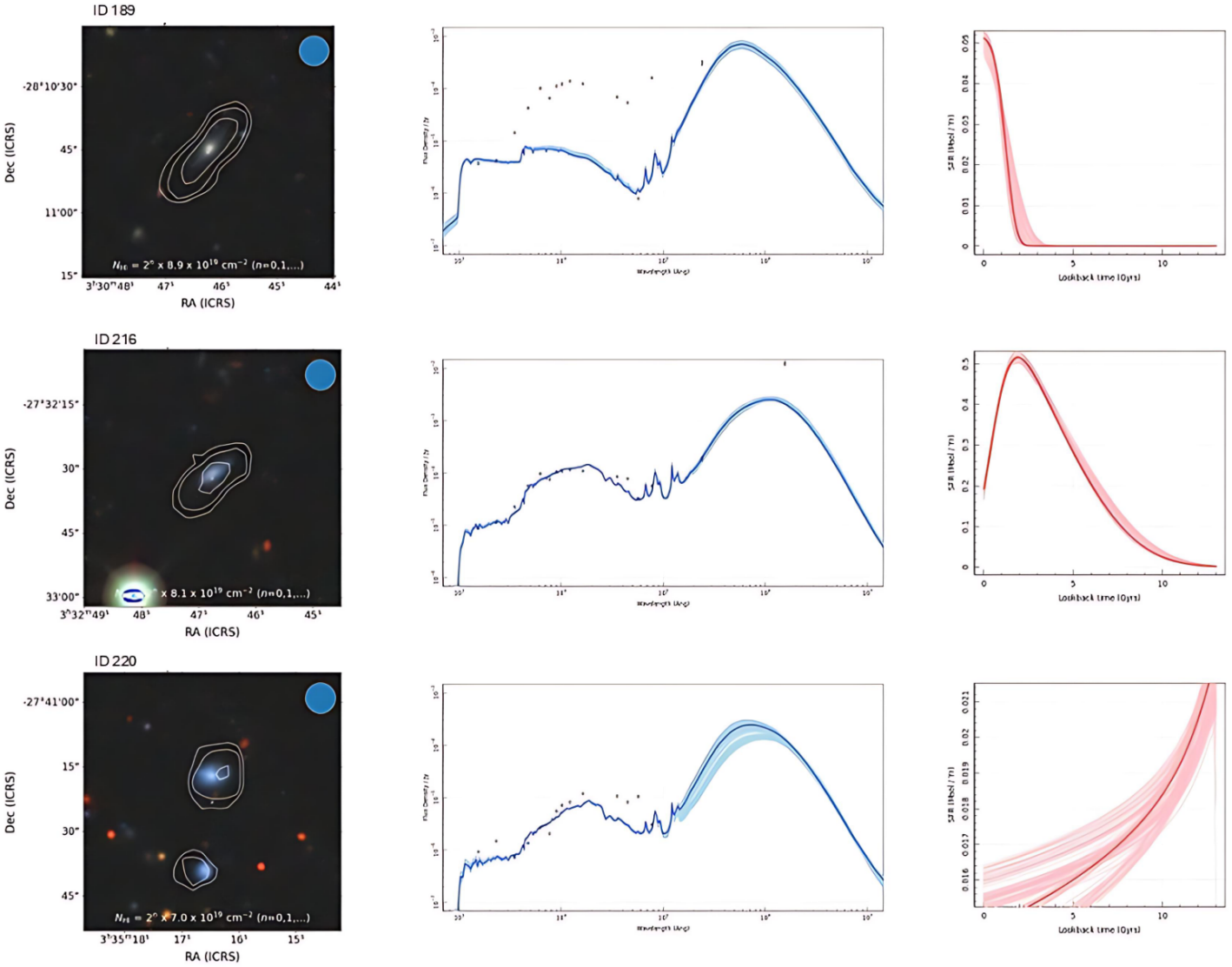
Continued on next page...



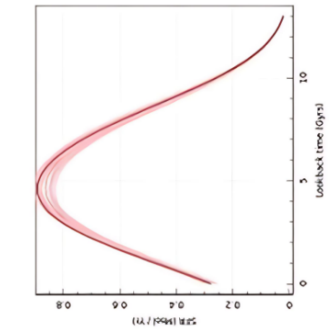
Continued on next page...

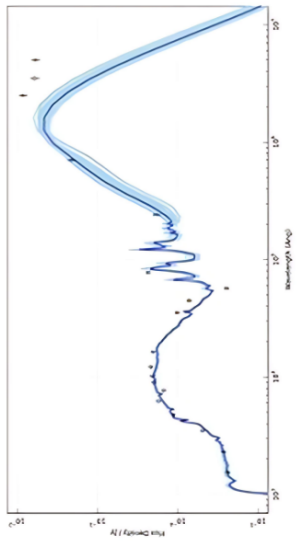


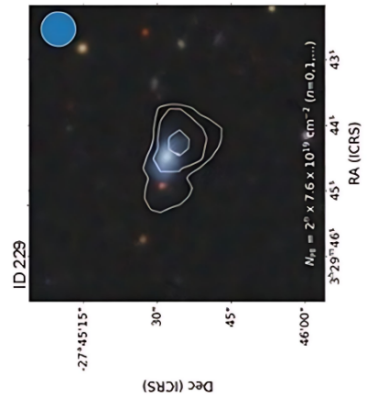
Continued on next page...



Continued on next page...






ID 229
$N_H = 2^0 \times 7.6 \times 10^{19}$ cm^{-2} (n=0,1,...)
Dec (ICRS)
RA (ICRS)
-27°45'15"
30"
45"
46'00"
3°29'46"
45s
44s
43s

Bibliography

Aihara, Hiroaki et al. (Jan. 2018). "The Hyper Suprime-Cam SSP Survey: Overview and survey design". In: 70, S4, S4. DOI: 10.1093/pasj/psx066. arXiv: 1704.05858 [astro-ph.IM].

Arnouts, S. et al. (Dec. 1999). "Measuring and modelling the redshift evolution of clustering: the Hubble Deep Field North". In: 310.2, pp. 540–556. DOI: 10.1046/j.1365-8711.1999.02978.x. arXiv: astro-ph/9902290 [astro-ph].

Bauer, A. E. et al. (Oct. 2011). "Star formation in a stellar mass-selected sample of galaxies to z= 3 from the GOODS-NICMOS Survey". In: 417.1, pp. 289–303. DOI: 10.1111/j.1365-2966.2011.19240.x. arXiv: 1106.2656 [astro-ph.CO].

Bellstedt, Sabine et al. (Nov. 2020a). "Galaxy And Mass Assembly (GAMA): a forensic SED reconstruction of the cosmic star formation history and metallicity evolution by galaxy type". In: 498.4, pp. 5581–5603. DOI: 10.1093/mnras/staa2620. arXiv: 2005.11917 [astro-ph.GA].

— (Aug. 2020b). "Galaxy And Mass Assembly (GAMA): assimilation of KiDS into the GAMA database". In: 496.3, pp. 3235–3256. DOI: 10.1093/mnras/staa1466. arXiv: 2005.11215 [astro-ph.GA].

Bertin, Emmanuel et al. (Jan. 2002). "The TERAPIX Pipeline". In: *Astronomical Data Analysis Software and Systems XI*. Ed. by David A. Bohlender, Daniel Durand, and Thomas H. Handley. Vol. 281. Astronomical Society of the Pacific Conference Series, p. 228.

Bigiel, F. et al. (Dec. 2008). "The Star Formation Law in Nearby Galaxies on Sub-Kpc Scales". In: 136.6, pp. 2846–2871. DOI: 10.1088/0004-6256/136/6/2846. arXiv: 0810.2541 [astro-ph].

Blum, Robert D. et al. (June 2016). "The DECam Legacy Survey". In: *American Astronomical Society Meeting Abstracts #228*. Vol. 228. American Astronomical Society Meeting Abstracts, 317.01, p. 317.01.

Blyth, S. et al. (Jan. 2016). "LADUMA: Looking at the Distant Universe with the MeerKAT Array". In: *MeerKAT Science: On the Pathway to the SKA*, 4, p. 4. DOI: 10.22323/1.277.0004.

Bordiu, C. et al. (Mar. 2025). "The SARAO MeerKAT Galactic Plane Survey extended source catalogue". In: 695, A144, A144. DOI: 10.1051/0004-6361/202450356. arXiv: 2502.15640 [astro-ph.GA].

Bowler, R. A. A. et al. (Apr. 2020). "A lack of evolution in the very bright end of the galaxy luminosity function from z 8 to 10". In: 493.2, pp. 2059–2084. DOI: 10.1093/mnras/staa313. arXiv: 1911.12832 [astro-ph.GA].

Bruzual, G. and S. Charlot (Oct. 2003). "Stellar population synthesis at the resolution of 2003". In: 344.4, pp. 1000–1028. DOI: 10.1046/j.1365-8711.2003.06897.x. arXiv: astro-ph/0309134 [astro-ph].

Bundy, Kevin et al. (Jan. 2015). "Overview of the SDSS-IV MaNGA Survey: Mapping nearby Galaxies at Apache Point Observatory". In: 798.1, 7, p. 7. DOI: 10.1088/0004-637X/798/1/7. arXiv: 1412.1482 [astro-ph.GA].

Carnall, A C et al. (Aug. 2018). "Inferring the star formation histories of massive quiescent galaxies with bagpipes: evidence for multiple quenching mechanisms". In: *Monthly Notices of the Royal Astronomical Society* 480.4, pp. 4379–4401. ISSN: 0035-8711. DOI:

10.1093/mnras/sty2169. eprint: https://academic.oup.com/mnras/article-pdf/480/4/4379/25539546/sty2169.pdf. URL: https://doi.org/10.1093/mnras/sty2169.

Carnall, A. C. (Nov. 2019). "Galaxy physical parameters from spectroscopy with Bagpipes". In: *The Art of Measuring Galaxy Physical Properties*, 44, p. 44. DOI: 10.5281/zenodo.3592320.

Carniani, Stefano et al. (Apr. 2025). "The eventful life of a luminous galaxy at z=14: metal enrichment, feedback, and low gas fraction?" In: *Astronomy amp; Astrophysics* 696, A87. ISSN: 1432-0746. DOI: 10.1051/0004-6361/202452451. URL: http://dx.doi.org/10.1051/0004-6361/202452451.

Chabrier, Gilles (July 2003). "Galactic Stellar and Substellar Initial Mass Function". In: 115.809, pp. 763–795. DOI: 10.1086/376392. arXiv: astro-ph/0304382 [astro-ph].

Charlot, Stéphane and Fall (Aug. 2000). "A Simple Model for the Absorption of Starlight by Dust in Galaxies". In: 539.2, pp. 718–731. DOI: 10.1086/309250. arXiv: astro-ph/0003128 [astro-ph].

Chevallard, Jacopo and Stéphane Charlot (Oct. 2016). "Modelling and interpreting spectral energy distributions of galaxies with BEAGLE". In: 462.2, pp. 1415–1443. DOI: 10.1093/mnras/stw1756. arXiv: 1603.03037 [astro-ph.GA].

Conselice, Christopher J. (Aug. 2014). "The Evolution of Galaxy Structure Over Cosmic Time". In: 52, pp. 291–337. DOI: 10.1146/annurev-astro-081913-040037. arXiv: 1403.2783 [astro-ph.GA].

Croom, Scott M. et al. (Mar. 2012). "The Sydney-AAO Multi-object Integral field spectrograph". In: 421.1, pp. 872–893. DOI: 10.1111/j.1365-2966.2011.20365.x. arXiv: 1112.3367 [astro-ph.CO].

Da Cunha, Elisabete et al. (Aug. 2008). "A simple model to interpret the ultraviolet, optical and infrared emission from galaxies". In: *Monthly Notices of the Royal Astronomical Society* 388.4, pp. 1595–1617. ISSN: 0035-8711. DOI: 10.1111/j.1365-2966.2008.13535.x. eprint: https://academic.oup.com/mnras/article-pdf/388/4/1595/3036508/mnras0388-1595.pdf. URL: https://doi.org/10.1111/j.1365-2966.2008.13535.x.

Dale, Daniel A. et al. (Jan. 2014). "A Two-Parameter Model for the Infrared/Submillimeter/Radio Spectral Energy Distributions of Galaxies and AGN". In: *American Astronomical Society Meeting Abstracts #223*. Vol. 223. American Astronomical Society Meeting Abstracts, 453.01, p. 453.01.

Davé, Romeel et al. (Mar. 2012). "An analytic model for the evolution of the stellar, gas and metal content of galaxies". In: 421.1, pp. 98–107. DOI: 10.1111/j.1365-2966.2011.20148.x. arXiv: 1108.0426 [astro-ph.CO].

Davies, L. J. M. et al. (Sept. 2021). "Deep Extragalactic VIsible Legacy Survey (DEVILS): consistent multiwavelength photometry for the DEVILS regions (COSMOS, XMMLSS, and ECDFS)". In: 506.1, pp. 256–287. DOI: 10.1093/mnras/stab1601. arXiv: 2106.06241 [astro-ph.GA].

— (Oct. 2018). "Deep Extragalactic VIsible Legacy Survey (DEVILS): motivation,design, and target catalogue". In: 480.1, pp. 768–799. DOI: 10.1093/mnras/sty1553. arXiv: 1806.05808 [astro-ph.GA].

Dey, Subhrata et al. (Feb. 2024). "Radio-only and radio-FUV SED modeling of 14 ULIRGs: insights into global properties of infrared bright galaxies". In: *arXiv e-prints*, arXiv:2402.10786, arXiv:2402.10786. DOI: 10.48550/arXiv.2402.10786. arXiv: 2402.10786 [astro-ph.GA].

Dickinson, Mark et al. (Apr. 2003). "The Evolution of the Global Stellar Mass Density at 0<z<3". In: 587.1, pp. 25–40. DOI: 10.1086/368111. arXiv: astro-ph/0212242 [astro-ph].

Duffy, Alan R. et al. (Nov. 2012). "Predictions for ASKAP neutral hydrogen surveys". In: 426.4, pp. 3385–3402. DOI: 10.1111/j.1365-2966.2012.21987.x. arXiv: 1208.5592 [astro-ph.CO].

Erfanianfar, G. et al. (Jan. 2016). "Non-linearity and environmental dependence of the star-forming galaxies main sequence". In: 455.3, pp. 2839–2851. DOI: 10.1093/mnras/stv2485. arXiv: 1511.01899 [astro-ph.GA].

Fernández, Ximena et al. (June 2016). "Highest Redshift Image of Neutral Hydrogen in Emission: A CHILES Detection of a Starbursting Galaxy at z = 0.376". In: 824.1, L1, p. L1. DOI: 10.3847/2041-8205/824/1/L1. arXiv: 1606.00013 [astro-ph.GA].

Fraser-McKelvie, A. et al. (May 2021). "A SAMI and MaNGA view on the stellar kinematics of galaxies on the star-forming main sequence". In: 503.4, pp. 4992–5005. DOI: 10.1093/mnras/stab573. arXiv: 2102.13342 [astro-ph.GA].

Freudling, W. et al. (Aug. 2008). "The ALFA Ultra Deep Survey (AUDS)". In: *The Evolution of Galaxies Through the Neutral Hydrogen Window*. Ed. by Robert Minchin and Emmanuel Momjian. Vol. 1035. American Institute of Physics Conference Series. AIP, pp. 242–245. DOI: 10.1063/1.2973592.

Gavazzi, G. et al. (Aug. 2015). "Hα3: an Hα imaging survey of HI selected galaxies from ALFALFA. VI. The role of bars in quenching star formation from z = 3 to the present epoch". In: 580, A116, A116. DOI: 10.1051/0004-6361/201425351. arXiv: 1505.07836 [astro-ph.GA].

Giovanelli, Riccardo et al. (Dec. 2005). "The Arecibo Legacy Fast ALFA Survey. I. Science Goals, Survey Design, and Strategy". In: 130.6, pp. 2598–2612. DOI: 10.1086/497431. arXiv: astro-ph/0508301 [astro-ph].

Gordon, Karl D. et al. (July 2008). "The Behavior of the Aromatic Features in M101 H II Regions: Evidence for Dust Processing". In: 682.1, pp. 336–354. DOI: 10.1086/589567. arXiv: 0804.3223 [astro-ph].

Gould, Robert J. (Mar. 1994). "Radiative Hyperfine Transitions". In: 423, p. 522. DOI: 10.1086/173830.

Ilbert, O. et al. (Oct. 2006). "Accurate photometric redshifts for the CFHT legacy survey calibrated using the VIMOS VLT deep survey". In: 457.3, pp. 841–856. DOI: 10.1051/0004-6361:20065138. arXiv: astro-ph/0603217 [astro-ph].

Jaffé, Yara L. et al. (May 2013). "BUDHIES I: characterizing the environments in and around two clusters at z0.2". In: 431.3, pp. 2111–2125. DOI: 10.1093/mnras/stt250. arXiv: 1302.1876 [astro-ph.CO].

Jarvis, M. et al. (Jan. 2016). "The MeerKAT International GHz Tiered Extragalactic Exploration (MIGHTEE) Survey". In: *MeerKAT Science: On the Pathway to the SKA*, 6, p. 6. DOI: 10.22323/1.277.0006. arXiv: 1709.01901 [astro-ph.GA].

Jarvis, Matt J. et al. (Oct. 2012). "The VISTA Deep Extragalactic Observations (VIDEO) survey". In: *Monthly Notices of the Royal Astronomical Society* 428.2, pp. 1281–1295.

ISSN: 0035-8711. DOI: 10.1093/mnras/sts118. eprint: https://academic.oup.com/mnras/article-pdf/428/2/1281/3240562/sts118.pdf. URL: https://doi.org/10.1093/mnras/sts118.

Jonas, Justin (2018). "The MeerKAT Radio Telescope". In: *PoS* MeerKAT2016, p. 001. DOI: 10.22323/1.277.0001.

Kapahi, Vijay K. et al. (Oct. 1998). "The Molonglo Reference Catalog 1 Jy Radio Source Survey. II. Radio Structures of Galaxy Identifications". In: 118.2, pp. 275–326. DOI: 10.1086/313144.

Kauffmann, Guinevere (June 2015). "The outer stellar populations and environments of unusually H I-rich galaxies". In: 450.1, pp. 618–629. DOI: 10.1093/mnras/stv654. arXiv: 1501.01017 [astro-ph.GA].

Kazemi-Moridani, Amir (Feb. 2024). "Early Results from the LADUMA (Looking At the Distant Universe with the MeerKAT Array) Survey". In: *American Astronomical Society Meeting Abstracts*. Vol. 243. American Astronomical Society Meeting Abstracts, 127.05D, p. 127.05D.

Kennicutt, Robert C. and Neal J. Evans (Sept. 2012). "Star Formation in the Milky Way and Nearby Galaxies". In: 50, pp. 531–608. DOI: 10.1146/annurev-astro-081811-125610. arXiv: 1204.3552 [astro-ph.GA].

Leitner, Samuel N. (Feb. 2012). "On the Last 10 Billion Years of Stellar Mass Growth in Star-forming Galaxies". In: 745.2, 149, p. 149. DOI: 10.1088/0004-637X/745/2/149. arXiv: 1108.0938 [astro-ph.CO].

Leja, Joel et al. (May 2019). "How to Measure Galaxy Star Formation Histories. II. Nonparametric Models". In: 876.1, 3, p. 3. DOI: 10.3847/1538-4357/ab133c. arXiv: 1811.03637 [astro-ph.GA].

Leroy, Adam K. et al. (Dec. 2008). "The Star Formation Efficiency in Nearby Galaxies: Measuring Where Gas Forms Stars Effectively". In: 136.6, pp. 2782–2845. DOI: 10.1088/0004-6256/136/6/2782. arXiv: 0810.2556 [astro-ph].

Lizarraga, Andrew et al. (Nov. 2024). "Understanding Galaxy Morphology Evolution Through Cosmic Time via Redshift Conditioned Diffusion Models". In: *arXiv e-prints*, arXiv:2411.18440, arXiv:2411.18440. DOI: 10.48550/arXiv.2411.18440. arXiv: 2411.18440 [astro-ph.GA].

Lonsdale, Carol J. et al. (Aug. 2003). "SWIRE: The SIRTF Wide-Area Infrared Extragalactic Survey". In: 115.810, pp. 897–927. DOI: 10.1086/376850. arXiv: astro-ph/0305375 [astro-ph].

Madau, Piero and Mark Dickinson (Aug. 2014). "Cosmic Star-Formation History". In: 52, pp. 415–486. DOI: 10.1146/annurev-astro-081811-125615. arXiv: 1403.0007 [astro-ph.CO].

Maddox, N. et al. (Feb. 2021). "MIGHTEE-HI: The H I emission project of the MeerKAT MIGHTEE survey". In: 646, A35, A35. DOI: 10.1051/0004-6361/202039655. arXiv: 2011.09470 [astro-ph.GA].

— (Feb. 2015). "Variation of galactic cold gas reservoirs with stellar mass". In: 447.2, pp. 1610–1617. DOI: 10.1093/mnras/stu2532. arXiv: 1412.0852 [astro-ph.GA].

Mancuso, C. et al. (Dec. 2016). "The Main Sequences of Star-forming Galaxies and Active Galactic Nuclei at High Redshift". In: 833.2, 152, p. 152. DOI: 10.3847/1538-4357/833/2/152. arXiv: 1610.05910 [astro-ph.GA].

Mauduit, J. -C. et al. (July 2012). "The Spitzer Extragalactic Representative Volume Survey (SERVS): Survey Definition and Goals". In: 124.917, p. 714. DOI: 10.1086/666945. arXiv: 1206.4060 [astro-ph.CO].

Meyer, M. (Jan. 2009). "Exploring the HI Universe with ASKAP". In: *Panoramic Radio Astronomy: Wide-field 1-2 GHz Research on Galaxy Evolution*, 15, p. 15. DOI: 10.22323/1.089.0015. arXiv: 0912.2167 [astro-ph.CO].

Meyer, M. et al. (Nov. 2017). "Tracing HI Beyond the Local Universe". In: 34, p. 52. DOI: 10.1017/pasa.2017.31. arXiv: 1705.04210 [astro-ph.CO].

Meyer, M. J. et al. (June 2004). "The HIPASS catalogue - I. Data presentation". In: 350.4, pp. 1195–1209. DOI: 10.1111/j.1365-2966.2004.07710.x. arXiv: astro-ph/0406384 [astro-ph].

Noeske, K. G. et al. (May 2007). "Star Formation in AEGIS Field Galaxies since z=1.1: The Dominance of Gradually Declining Star Formation, and the Main Sequence of Star-forming Galaxies". In: 660.1, pp. L43–L46. DOI: 10.1086/517926. arXiv: astro-ph/0701924 [astro-ph].

Noll, S. et al. (Dec. 2009). "Analysis of galaxy spectral energy distributions from far-UV to far-IR with CIGALE: studying a SINGS test sample". In: 507.3, pp. 1793–1813. DOI: 10.1051/0004-6361/200912497. arXiv: 0909.5439 [astro-ph.CO].

Oesch, P. A. et al. (Mar. 2016). "A Remarkably Luminous Galaxy at z=11.1 Measured with Hubble Space Telescope Grism Spectroscopy". In: 819.2, 129, p. 129. DOI: 10.3847/0004-637X/819/2/129. arXiv: 1603.00461 [astro-ph.GA].

Oliver, S. J. et al. (Aug. 2012a). "The Herschel Multi-tiered Extragalactic Survey: HerMES". In: 424.3, pp. 1614–1635. DOI: 10.1111/j.1365-2966.2012.20912.x. arXiv: 1203.2562 [astro-ph.CO].

— (Aug. 2012b). "The Herschel Multi-tiered Extragalactic Survey: HerMES". In: 424.3, pp. 1614–1635. DOI: 10.1111/j.1365-2966.2012.20912.x. arXiv: 1203.2562 [astro-ph.CO].

Pacifici, Camilla et al. (Feb. 2023). "The Art of Measuring Physical Parameters in Galaxies: A Critical Assessment of Spectral Energy Distribution Fitting Techniques". In: 944.2, 141, p. 141. DOI: 10.3847/1538-4357/acacff. arXiv: 2212.01915 [astro-ph.GA].

Parkash, Vaishali et al. (Sept. 2018). "Relationships between HI Gas Mass, Stellar Mass, and the Star Formation Rate of HICAT+WISE (H I-WISE) Galaxies". In: 864.1, 40, p. 40. DOI: 10.3847/1538-4357/aad3b9. arXiv: 1807.06246 [astro-ph.GA].

Planck Collaboration et al. (Sept. 2016). "Planck 2015 results. XIII. Cosmological parameters". In: 594, A13, A13. DOI: 10.1051/0004-6361/201525830. arXiv: 1502.01589 [astro-ph.CO].

Ponomareva, Anastasia A. et al. (Nov. 2021). "MIGHTEE-H I: the baryonic Tully-Fisher relation over the last billion years". In: 508.1, pp. 1195–1205. DOI: 10.1093/mnras/stab2654. arXiv: 2109.04992 [astro-ph.GA].

— (July 2023). "MIGHTEE-H I: the first MeerKAT H I mass function from an untargeted interferometric survey". In: 522.4, pp. 5308–5319. DOI: 10.1093/mnras/stad1249. arXiv: 2304.13051 [astro-ph.GA].

Popesso, P. et al. (Mar. 2019). "The main sequence of star-forming galaxies - I. The local relation and its bending". In: 483.3, pp. 3213–3226. DOI: 10.1093/mnras/sty3210. arXiv: 1812.07057 [astro-ph.GA].

Popesso, P. et al. (Feb. 2023). "The main sequence of star-forming galaxies across cosmic times". In: 519.1, pp. 1526–1544. DOI: 10.1093/mnras/stac3214. arXiv: 2203.10487 [astro-ph.GA].

Rajohnson, Sambatriniaina H. A. et al. (July 2024). "H I galaxy signatures in the SARAO MeerKAT galactic plane survey - III. Unveiling the obscured part of the Vela Supercluster". In: 531.3, pp. 3486–3508. DOI: 10.1093/mnras/stae1354. arXiv: 2405.15629 [astro-ph.GA].

— (May 2022). "MIGHTEE-H I: the H I size-mass relation over the last billion years". In: 512.2, pp. 2697–2706. DOI: 10.1093/mnras/stac693. arXiv: 2203.06149 [astro-ph.GA].

Rhee, Jonghwan et al. (Jan. 2023). "Deep investigation of neutral gas origins (DINGO): H I stacking experiments with early science data". In: 518.3, pp. 4646–4671. DOI: 10.1093/mnras/stac3065. arXiv: 2210.09697 [astro-ph.GA].

Robotham, A S G et al. (Feb. 2018). "ProFound: Source Extraction and Application to Modern Survey Data". In: *Monthly Notices of the Royal Astronomical Society* 476.3, pp. 3137–3159. ISSN: 0035-8711. DOI: 10.1093/mnras/sty440. eprint: https://academic.oup.com/mnras/article-pdf/476/3/3137/24458018/sty440.pdf. URL: https://doi.org/10.1093/mnras/sty440.

— (June 2020). "ProSpect: generating spectral energy distributions with complex star formation and metallicity histories". In: 495.1, pp. 905–931. DOI: 10.1093/mnras/staa1116. arXiv: 2002.06980 [astro-ph.GA].

Salim, Samir et al. (Dec. 2007). "UV Star Formation Rates in the Local Universe". In: 173.2, pp. 267–292. DOI: 10.1086/519218. arXiv: 0704.3611 [astro-ph].

Schreiber, C. et al. (Oct. 2018). "Near infrared spectroscopy and star-formation histories of $3 \leq z \leq 4$ quiescent galaxies". In: 618, A85, A85. DOI: 10.1051/0004-6361/201833070. arXiv: 1807.02523 [astro-ph.GA].

Serra, Paolo et al. (Apr. 2015). "SOFIA: a flexible source finder for 3D spectral line data". In: 448.2, pp. 1922–1929. DOI: 10.1093/mnras/stv079. arXiv: 1501.03906 [astro-ph.IM].

Sinigaglia, Francesco et al. (Aug. 2022). "MIGHTEE-HI: Evolution of HI Scaling Relations of Star-forming Galaxies at $z < 0.5$". In: 935.1, L13, p. L13. DOI: 10.3847/2041-8213/ac85ae. arXiv: 2208.01121 [astro-ph.GA].

Speagle, J. S. et al. (Oct. 2014). "A Highly Consistent Framework for the Evolution of the Star-Forming "Main Sequence" from z ~0-6". In: 214.2, 15, p. 15. DOI: 10.1088/0067-0049/214/2/15. arXiv: 1405.2041 [astro-ph.GA].

Thorne, Jessica E. et al. (July 2021). "Deep Extragalactic VIsible Legacy Survey (DEVILS): SED fitting in the D10-COSMOS field and the evolution of the stellar mass function and SFR-$M_\star$ relation". In: 505.1, pp. 540–567. DOI: 10.1093/mnras/stab1294. arXiv: 2011.13605 [astro-ph.GA].

Tinsley, B. M. (Jan. 1980). "Evolution of the Stars and Gas in Galaxies". In: 5, pp. 287–388. DOI: 10.48550/arXiv.2203.02041. arXiv: 2203.02041 [astro-ph.GA].

Tudorache, Madalina N. et al. (Nov. 2024). "MIGHTEE-HI: The star-forming properties of HI selected galaxies". In: *arXiv e-prints*, arXiv:2411.14940, arXiv:2411.14940. DOI: 10.48550/arXiv.2411.14940. arXiv: 2411.14940 [astro-ph.GA].

Vaccari, M. et al. (Jan. 2016). "The VOICE Survey : VST Optical Imaging of the CDFS and ES1 Fields". In: *The 4th Annual Conference on High Energy Astrophysics in*

Southern Africa (HEASA 2016), 26, p. 26. DOI: 10.22323/1.275.0026. arXiv: 1704.01495 [astro-ph.GA].

Varadaraj, R. G. et al. (Sept. 2023). "The bright end of the galaxy luminosity function at z 7 from the VISTA VIDEO survey". In: 524.3, pp. 4586–4613. DOI: 10.1093/mnras/stad2081. arXiv: 2304.02494 [astro-ph.GA].

Walter, Fabian et al. (Oct. 2020a). "The Evolution of the Baryons Associated with Galaxies Averaged over Cosmic Time and Space". In: 902.2, 111, p. 111. DOI: 10.3847/1538-4357/abb82e. arXiv: 2009.11126 [astro-ph.GA].

— (Oct. 2020b). "The Evolution of the Baryons Associated with Galaxies Averaged over Cosmic Time and Space". In: 902.2, 111, p. 111. DOI: 10.3847/1538-4357/abb82e. arXiv: 2009.11126 [astro-ph.GA].

Westmeier, T. et al. (Nov. 2022). "WALLABY pilot survey: Public release of H I data for almost 600 galaxies from phase 1 of ASKAP pilot observations". In: 39, e058, e058. DOI: 10.1017/pasa.2022.50. arXiv: 2211.07094 [astro-ph.GA].

Westmeier, Tobias et al. (Sept. 2021). *SoFiA 2: An automated, parallel HI source finding pipeline*. Astrophysics Source Code Library, record ascl:2109.005. ascl: 2109.005.

Worthey, Guy (Nov. 1994). "Comprehensive Stellar Population Models and the Disentanglement of Age and Metallicity Effects". In: 95, p. 107. DOI: 10.1086/192096.

Xi, Hongwei, Bo Peng, et al. (Sept. 2024). "FAST Ultra-Deep Survey: Data Release for FUDS0". In: 274.1, 18, p. 18. DOI: 10.3847/1538-4365/ad67d5. arXiv: 2409.06153 [astro-ph.GA].

Zhou, Zhimin et al. (Sept. 2018). "The Relation between H I Gas and Star Formation Properties in Nearby Galaxies". In: 130.991, p. 094101. DOI: 10.1088/1538-3873/aad407. arXiv: 1807.09955 [astro-ph.GA].