

Central Lancashire Online Knowledge (CLoK)

Title	Using Machine Learning to Estimate Near-ultraviolet Magnitudes and Probe Quenching Mechanisms of $z = 0$ Nuggets in the RESOLVE and ECO Surveys
Type	Article
URL	https://clock.uclan.ac.uk/id/eprint/55539/
DOI	https://doi.org/10.3847/1538-4357/adc71b
Date	2025
Citation	Carr, Derrick S., Kannappan, Sheila J., Hutchens, Zackary L., Polimera, Mugdha S., Norris, Mark orcid iconORCID: 0000-0002-7001-805X, Eckert, Kathleen D. and Moffett, Amanda J. (2025) Using Machine Learning to Estimate Near-ultraviolet Magnitudes and Probe Quenching Mechanisms of $z = 0$ Nuggets in the RESOLVE and ECO Surveys. <i>Astrophysical Journal</i> , 985 (1). ISSN 0004-637X
Creators	Carr, Derrick S., Kannappan, Sheila J., Hutchens, Zackary L., Polimera, Mugdha S., Norris, Mark, Eckert, Kathleen D. and Moffett, Amanda J.

It is advisable to refer to the publisher's version if you intend to cite from the work.
<https://doi.org/10.3847/1538-4357/adc71b>

For information about Research at UCLan please go to <http://www.uclan.ac.uk/research/>

All outputs in CLoK are protected by Intellectual Property Rights law, including Copyright law. Copyright, IPR and Moral Rights for the works on this site are retained by the individual authors and/or other copyright owners. Terms and conditions for use of this material are defined in the <http://clock.uclan.ac.uk/policies/>

Central Lancashire Online Knowledge (CLoK)

Title	Using Machine Learning to Estimate Near-ultraviolet Magnitudes and Probe Quenching Mechanisms of $z = 0$ Nuggets in the RESOLVE and ECO Surveys
Type	Article
URL	https://clock.uclan.ac.uk/55539/
DOI	https://doi.org/10.3847/1538-4357/adc71b
Date	2025
Citation	Carr, Derrick S., Kannappan, Sheila J., Hutchens, Zackary L., Polimera, Mugdha S., Norris, Mark A., Eckert, Kathleen D. and Moffett, Amanda J. (2025) Using Machine Learning to Estimate Near-ultraviolet Magnitudes and Probe Quenching Mechanisms of $z = 0$ Nuggets in the RESOLVE and ECO Surveys. <i>The Astrophysical Journal</i> , 985 (1). ISSN 1538-4357
Creators	Carr, Derrick S., Kannappan, Sheila J., Hutchens, Zackary L., Polimera, Mugdha S., Norris, Mark A., Eckert, Kathleen D. and Moffett, Amanda J.

It is advisable to refer to the publisher's version if you intend to cite from the work.
<https://doi.org/10.3847/1538-4357/adc71b>

For information about Research at UCLan please go to <http://www.uclan.ac.uk/research/>

All outputs in CLoK are protected by Intellectual Property Rights law, including Copyright law. Copyright, IPR and Moral Rights for the works on this site are retained by the individual authors and/or other copyright owners. Terms and conditions for use of this material are defined in the <http://clock.uclan.ac.uk/policies/>



Using Machine Learning to Estimate Near-ultraviolet Magnitudes and Probe Quenching Mechanisms of $z = 0$ Nuggets in the RESOLVE and ECO Surveys

Derrick S. Carr¹, Sheila J. Kannappan¹, Zackary L. Hutchens^{1,2}, Mugdha S. Polimera^{1,3}, Mark A. Norris⁴, Kathleen D. Eckert¹, and Amanda J. Moffett⁵

¹ Department of Physics & Astronomy, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA

² Department of Physics & Astronomy, University of North Carolina Asheville, 1 University Heights, Asheville, NC 28804, USA

³ Center for Astrophysics | Harvard & Smithsonian, 60 Garden Street, Cambridge, MA 02138, USA

⁴ Jeremiah Horrocks Institute, University of Central Lancashire, Preston, PR1 2HE, UK

⁵ Department of Physics and Astronomy, University of North Georgia, 3820 Mundy Mill Rd., Oakwood, GA 30566, USA

Received 2024 June 14; revised 2025 March 17; accepted 2025 March 27; published 2025 May 16

Abstract

We present a $z = 0$ census of nuggets—compact galaxies that form via gas-rich violent disk instability—within the luminosity- and volume-limited RESOLVED Spectroscopy Of a Local Volume (RESOLVE) and Environmental COntext (ECO) surveys. We use random forest (RF) models to predict near-ultraviolet (NUV) magnitudes for ECO galaxies that lack high-quality NUV magnitudes, thereby doubling the number of ECO galaxies with reliable extinction-corrected star formation rates (SFRs) and red/green/blue classifications based on specific SFRs (sSFRs). The resulting RF-enhanced RESOLVE+ECO nugget sample allows us to analyze rare subpopulations—green nuggets and nuggets with active galactic nuclei (AGN)—likely associated with quenching. Green nuggets are more similar to red nuggets than to blue nuggets in halo mass (M_{halo}) distribution, with both red and green nuggets being found mainly at $M_{\text{halo}} \geq 10^{11.4} M_{\odot}$, where permanent halo quenching is predicted. At these masses, the AGN frequency for green nuggets is higher ($48.2\%_{-5.3\%}^{+5.3\%}$) than for either blue ($39.2\%_{-2.8\%}^{+2.9\%}$) or red ($29.3\%_{-2.8\%}^{+3.0\%}$) nuggets. Between $M_{\text{halo}} = 10^{11.4} - 10^{12} M_{\odot}$, at the onset of permanent quenching, the AGN frequency for green nuggets is nearly double the frequency for blue or red nuggets, implying AGN are associated with this transition. At $M_{\text{halo}} < 10^{11.4} M_{\odot}$, where temporary cyclic quenching is expected, the AGN frequency for blue nuggets ($7.5\%_{-1.2\%}^{+1.4\%}$) is lower than for either green ($31.3\%_{-7.5\%}^{+8.7\%}$) or red ($18.8\%_{-7.8\%}^{+11.5\%}$) nuggets. At all masses, nuggets with AGN have reduced sSFRs and likely also atomic gas content compared to nuggets without AGN, but the quenching is more extreme below $M_{\text{halo}} = 10^{11.4} M_{\odot}$.

Unified Astronomy Thesaurus concepts: Galaxy bulges (578); Galaxy evolution (594); Galaxy quenching (2040); Compact dwarf galaxies (281); Blue compact dwarf galaxies (165); Random Forests (1935); Early-type galaxies (429); Active galactic nuclei (16)

1. Introduction

Dense galaxies formed by gas-rich compaction events, also known as nuggets, represent key evolutionary phases that have helped shape the local bulged galaxy population. They were first discovered as quiescent objects, known as red nuggets, at redshift $z > 1.6$ (A. Cimatti et al. 2004; I. Trujillo et al. 2006; P. G. van Dokkum et al. 2008). Subsequently, G. Barro et al. (2013) was able to identify a high- z population of star-forming nuggets, known as blue nuggets, as their likely progenitors. Shortly after, a toy model by A. Dekel & A. Burkert (2014) suggested how nuggets fit into the galaxy evolution picture. In this model, blue nuggets form via compaction events, which are defined by gas-rich violent disk instabilities driven by wet mergers or colliding gas streams. The blue nuggets then quench into red nuggets through a combination of halo quenching and internal quenching mechanisms. After this transition, A. Dekel & A. Burkert (2014) suggest that red nuggets become massive ellipticals through mostly minor mergers and satellite accretion. Building on this picture, I. G. de la Rosa et al. (2016) studied the $z \sim 0.1$ bulged galaxy population and the $z \sim 1.5$ nugget population and concluded that nuggets are primordial seeds not

only for massive ellipticals but also for lenticular galaxies and classical bulged spirals (see also Y. Gao & L.-L. Fan 2020). Thus, understanding present-day spheroids requires understanding nuggets.

Nuggets are rare in the local Universe, which makes them challenging to study robustly. Observations of the evolution of massive ($M_{*} > 10^{10} M_{\odot}$) nuggets in the high- z Universe show that nuggets of similar mass and density become less common over cosmic time (G. Barro et al. 2013; I. Damjanov et al. 2014; A. Charbonnier et al. 2017). At $z = 0$, C. Saulder et al. (2015) searched for massive red nuggets with low effective radii and elevated velocity dispersions, and identified only 76 candidates within a parent Sloan Digital Sky Survey (SDSS) sample of $>230,000$ (0.03%). That said, dwarf nuggets may still be forming today. M. L. Palumbo et al. (2020) performed a dedicated search for compact dwarf starburst galaxies with properties consistent with blue nuggets, and found that $\sim 5\%$ of all dwarf galaxies are candidate blue nuggets, roughly matching theoretical expectations from the toy model of A. Dekel & A. Burkert (2014).

At low z , nuggets are expected to quench permanently over a range in halo mass from $M_{\text{halo}} \sim 10^{11.4} M_{\odot}$ to $M_{\text{halo}} \sim 10^{12.1} M_{\odot}$ due to shock heating of cosmic gas (e.g., A. Zolotov et al. 2015). Virial shocks that can significantly suppress cold-gas accretion are theorized to form in the inner halo above $M_{\text{halo}} \sim 10^{11.4} M_{\odot}$ and to become pervasive out to the virial



Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](https://creativecommons.org/licenses/by/4.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

radius above $M_{\text{halo}} \sim 10^{12.1} M_{\odot}$ (A. Dekel & Y. Birnboim 2006; A. Zolotov et al. 2015). These two halo masses are herein labeled the *threshold* scale and *bimodality* scale, respectively (S. J. Kannappan et al. 2013). Via the stellar mass–halo mass relation (e.g., K. D. Eckert et al. 2016), the corresponding central galaxies have stellar masses $M_* \sim 10^{9.6}$ and $10^{10.5} M_{\odot}$, respectively. We note that the cosmic web geometry that permits cold streams in higher-mass halos at high z does not persist at low z (A. Dekel & Y. Birnboim 2006).

Below $M_{\text{halo}} \sim 10^{11.4} M_{\odot}$, in the dwarf galaxy regime, A. Zolotov et al. (2015) and S. Tacchella et al. (2016) have found that simulated nuggets cyclically quench due to temporary gas depletion and outflows, resulting in ~ 0.3 dex oscillations on the star-forming main sequence. Using the complete volume- and luminosity-limited RESolved Spectroscopy Of a Local Volume (RESOLVE) survey in D. S. Carr et al. (2024, hereafter C24), we reported preliminary $z \sim 0$ evidence of cyclic quenching below the threshold scale through an analysis of star formation activity in dwarf nuggets. We also reported evidence of halo quenching above the threshold scale based on the halo mass distributions of blue and red nuggets. However, we could not confirm that the halo masses of nuggets plausibly caught in the blue-to-red transition (also known as green nuggets) closely match the halo masses of red nuggets, with only 10 green nuggets in our sample. Additionally, we would like to test whether green nuggets have intermediate atomic gas-to-stellar mass (G/S) ratios suggestive of quenching from blue to red.

Some studies have suggested that active galactic nuclei (AGN) play a role in the blue-to-red nugget transition, although their exact role is unclear. Most simulations that focus on nuggets cannot assess AGN feedback as they do not incorporate AGN physics (e.g., A. Zolotov et al. 2015; S. Tacchella et al. 2016; J. Primack et al. 2018); they only speculate that AGN feedback plays a role in permanently quenching nuggets. In one of the only studies to use simulations to assess the role of AGN in quenching nuggets, J. Nogueira-Cavalcante et al. (2019) compared green nuggets from hydrosimulations with green nuggets from $z \sim 0.8$ observations and concluded that green nuggets quench faster than typical green valley galaxies due to being in an efficient mode of AGN feedback.

Focusing on AGN statistics, G. Barro et al. (2013) found a higher frequency of X-ray-detected AGN in $z \sim 2$ blue nuggets ($\sim 30\%$) compared to extended blue galaxies ($\sim 1\%$) where both had stellar masses $M_* > 10^{10} M_{\odot}$. The authors proposed that AGN may help remove gas from nuggets after virial shock heating shuts down cold accretion. Similarly, D. D. Kocevski et al. (2017) found that massive ($M_* > 10^{10} M_{\odot}$) blue nuggets at $z \sim 2$ are $4.7\times$ and $7.6\times$ times more likely to host an X-ray-detected AGN than red nuggets and extended blue galaxies, respectively. They argued that AGN likely contribute to the feedback energy required to produce the short quenching timescale observed for nuggets (~ 0.5 Gyr; G. Barro et al. 2013). While these studies find a role for AGN in (massive) blue and red nuggets, they do not probe the AGN frequency in green nuggets relative to red and blue nuggets, which would clarify whether AGN actually coincide with the moment of transition. The $z=0$ nugget sample of C24 included green nuggets, but only eight above the threshold scale, yielding a green nugget AGN frequency consistent with both the red and the blue nugget AGN frequencies within the uncertainties,

despite the blue nugget AGN frequency being $\sim 2\times$ higher than the red nugget AGN frequency.

Below the threshold scale, the role of AGN in nugget quenching has received little attention. S. Tacchella et al. (2016) found that cyclic quenching in this regime is driven by compaction cycles where first gas inflow exceeds depletion/outflows, then the reverse occurs. However, they could only track outflows from star formation feedback. Despite their simulations not incorporating AGN, A. Zolotov et al. (2015) speculated that AGN feedback is likely associated with blue nuggets at all masses and may help boost internal quenching in nuggets. We studied nuggets in the cyclic quenching regime in C24, but our sample had only four nuggets with AGN below the threshold scale (one blue, one green, two red), too few to probe the role of AGN in temporary quenching. Evidence of AGN feedback in non-nugget dwarf galaxies has been found in several studies (e.g., G. Dashyan et al. 2018; S. J. Penny et al. 2018; C. M. Manzano-King et al. 2019), so an exploration of AGN quenching in dwarf nuggets is warranted. However, with AGN detections being uncommon in dwarf galaxies ($\sim 5\%–20\%$ of dwarfs host AGN; M. S. Polimera et al. 2022; M. Mezcuca & H. D. Sánchez 2024), a larger nugget census than in C24 is needed to probe the dwarf nugget AGN regime.

The present study aims to develop a larger parent sample capable of answering the above open questions. The Environmental Context (ECO) survey, a survey $\sim 8\times$ larger than the RESOLVE survey we used for C24 yet similarly designed, is a logical starting point. However, the analysis of RESOLVE nuggets was contingent on classifying nuggets as blue/green/red based on extinction-corrected star formation rates (SFRs) derived from custom-processed deep (exposure time > 1000 s) Galaxy Evolution Explorer (GALEX) near-ultraviolet (NUV) imaging (K. D. Eckert et al. 2016, tabulated in Z. L. Hutchens et al. 2023). Unfortunately, deep GALEX NUV imaging does not exist for roughly half of the ECO footprint, and high-quality NUV magnitudes are necessary for generating accurate SFRs (see Section 3.1). That said, nearly all of ECO does have shallower GALEX NUV photometry, typically from the All-Sky Imaging Survey (AIS; P. Morrissey et al. 2007). If we can use machine learning that has been calibrated using the half of ECO *with* high-quality custom-processed NUV photometry to predict NUV magnitudes for the other portion of ECO *without* such data, then we can use all of the ECO survey for our nugget analysis.

In this work, we train random forest (RF) regression models to fill in the missing NUV data, enabling us to create a merged RESOLVE+ECO census of 1082 nuggets within a parent sample of 10,018 galaxies to probe AGN and halo quenching in nuggets. We use this enhanced data set to answer three questions:

1. Are the halo mass and G/S distributions of green nuggets consistent with their being nuggets caught in the process of halo quenching?
2. Above the threshold scale ($M_{\text{halo}} > 10^{11.4} M_{\odot}$), can AGN be linked to the permanent blue-to-red nugget transition?
3. Below the threshold scale, can AGN be linked to temporary quenching in nuggets?

In this study, we use a standard Λ CDM cosmology with $\Omega_m = 0.3$, $\Omega_{\Lambda} = 0.7$ and $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$ except where noted otherwise.

2. Data and Methods

2.1. RESOLVE and ECO

The RESOLVE survey is a $\sim 53,000 \text{ Mpc}^3$ volume- and luminosity-limited census of galaxies that is highly complete down to the dwarf regime just above $M_* \sim 10^9 M_\odot$ (S. J. Kannappan & L. H. Wei 2008; K. D. Eckert et al. 2016). RESOLVE spans two equatorial strips that create two subvolumes, RESOLVE-A and RESOLVE-B, both limited by group redshift $4500\text{--}7000 \text{ km s}^{-1}$. RESOLVE-A spans R.A. $8.75\text{--}15.75 \text{ hr}$ and decl. $0^\circ\text{--}5^\circ$, while RESOLVE-B spans R.A. $22\text{--}3 \text{ hr}$ and decl. $-1.25 \text{ to } +1.25$. For nugget analysis, we refer to the same version of the RESOLVE survey as used by C24: RESOLVE-A was cut off at the luminosity completeness limit of $M_r = -17.33$ matching ECO, while RESOLVE-B was cut off at $M_r = -17.00$ (see K. D. Eckert et al. 2015). The full volume- and luminosity-limited RESOLVE parent survey contains 1453 galaxies. For machine learning development and calibration, we only use RESOLVE-A because RESOLVE-A is a subset of ECO. RESOLVE-A contains 959 galaxies.

The ECO survey is a $>400,000 \text{ Mpc}^3$ volume- and luminosity-limited survey that is similar in design to RESOLVE, except it is constructed with mostly archival data (A. J. Moffett et al. 2015; K. D. Eckert et al. 2016; Z. L. Hutchens et al. 2023). ECO naturally overlaps with the A-semester footprint of RESOLVE and adopts RESOLVE parameters for galaxies in both surveys, but it is much larger on-sky (R.A. 8.7 hr to 15.82 hr and decl. $-1^\circ\text{--}50^\circ$) and somewhat larger in group redshift range ($3000\text{--}7000 \text{ km s}^{-1}$). ECO is also complete down to $M_r = -17.33$. The luminosity-limited parent ECO survey contains 9640 galaxies.

2.2. Custom Photometry

RESOLVE and ECO both use custom-processed multiwavelength photometry from K. D. Eckert et al. (2015), A. J. Moffett et al. (2015), and K. D. Eckert et al. (2016), tabulated and updated in Z. L. Hutchens et al. (2023). The magnitude extrapolation estimated total galaxy magnitude by combining the flux in fixed annuli that were created using *gri* coadded SDSS images, with estimated flux from methods such as an outer-disk fit, curve of growth (J. Munoz-Mateos et al. 2009), and an outer-disk color correction (see K. D. Eckert et al. 2015 for more details). The methods of K. D. Eckert et al. (2015) return brighter magnitudes and bluer colors compared to SDSS catalog magnitudes. The NUV imaging comes from the GALEX mission, so it did not require background subtraction but still underwent magnitude extrapolation via *gri* annuli. Only galaxies with a total NUV exposure time $>1000 \text{ s}$ have custom-processed NUV photometry from K. D. Eckert et al. (2015, 2016), so we have high-quality NUV data for $\sim 97\%$ of RESOLVE-A galaxies, $\sim 96\%$ of RESOLVE-B galaxies, and $\sim 45\%$ of ECO galaxies. While RESOLVE and ECO have very similar photometric processing methods, RESOLVE has slightly better photometry due to by-eye inspection of the data for each galaxy. More details on the custom-processed photometry pipeline can be found in K. D. Eckert et al. (2015). We refer to various NUV magnitude estimates throughout this study, so we provide a glossary for each NUV term in Table 1. Conversion between m and M uses the equation

$$m = M - 5 + 5 \log(D_L), \quad (1)$$

where the luminosity distance, D_L , is the line-of-sight comoving distance in megaparsecs, loscmvdist , times $(1 + cz/c)$ based on Z. L. Hutchens et al. (2023).

2.3. Structural Parameters

Our approach uses seeing-deconvolved effective radii, R_e , to identify intrinsically small objects. C24 used PyProFit, a profile-fitting algorithm, to obtain R_e for RESOLVE galaxies. Roughly 90% of RESOLVE galaxies have successful PyProFit models, while the other $\sim 10\%$ have R_e values from nondeconvolved fitting (K. D. Eckert et al. 2015). We use effective radii from the Dark Energy Camera Legacy Survey (DECaLS) Data Release 9 (A. Dey et al. 2019) for ECO. DECaLS provides seeing-deconvolved R_e values derived from light profile modeling via The Tractor algorithm (D. Lang et al. 2016). C24 showed that DECaLS R_e values agree well with PyProFit R_e values for RESOLVE galaxies that have both R_e estimates. Roughly 92% of ECO galaxies have nonnull R_e from DECaLS, and in the null cases we used R_e values from nondeconvolved fitting from the ECO database (Z. L. Hutchens et al. 2023). Visual inspection of galaxies with failed light profile fitting showed that they were typically either early-stage mergers or faint irregular dwarfs, both of which are difficult to fit with standard light profiles and fortunately unlikely to overlap with nuggets. In these instances, we prefer to use the photometric structural parameters of K. D. Eckert et al. (2015; cataloged in Z. L. Hutchens et al. 2023) rather than enforcing standard light profile models for galaxies with nonstandard profiles. We have previously shown that the nondeconvolved R_e values agree well with the deconvolved values except at very small R_e , and that PyProFit and Tractor R_e values are consistent (see C24, Section 2.5 and Figures 2 and 3). All radii are converted to units of kiloparsecs based on their Hubble distance using recession velocity cz as tabulated in Z. L. Hutchens et al. (2023). We also use the μ_Δ parameter from S. J. Kannappan et al. (2013), which acts as a quantitative morphological metric that has been shown to discern between quasi-bulgeless, bulge+disk, and bulge-dominated galaxies.

2.4. Environment Metrics

For this study, the environment metrics come mainly from Z. L. Hutchens et al. (2023), which used a four-step group-finding algorithm that results in higher completeness compared to typical friends-of-friends algorithms, where completeness is defined as the percentage of galaxies in the true group dark matter halo that the algorithm finds in the best-matched identified group, as measured using mock catalogs from simulations in which the true groups/dark matter halos are known. Z. L. Hutchens et al. (2023) provided group assignments, halo masses derived from halo abundance matching, and central/satellite flags. We have also used a “flyby” flag from C24 that identifies flyby galaxies as galaxies inside of a halo with $M_{\text{halo}} < 10^{12} M_\odot$ and within $1.5 \times R_{\text{vir}}$ of a halo with $M_{\text{halo}} \geq 10^{12} M_\odot$.

2.5. Stellar Masses, Atomic Gas Masses, Star Formation Rates, and Active Galactic Nuclei

For RESOLVE, stellar masses and extinction- and k -corrected magnitudes (including NUV) are derived from spectral energy distribution (SED) modeling (S. J. Kannappan & E. Gawiser 2007; S. J. Kannappan et al. 2009, 2013). The SED model grid is built using a Chabrier IMF (G. Chabrier 2003) and stellar population models from G. Bruzual & S. Charlot (2003). We used a variant of the S. J. Kannappan et al. (2013) modeling code adopted by K. D. Eckert et al. (2015), which differs from S. J. Kannappan et al. (2013) only in that it rejects middle-aged young populations with ages $\geq 1.4 \text{ Gyr}$ (K. D. Eckert et al. 2017).

Table 1
NUV Magnitude Glossary

Terminology	Description
m_{band}	Apparent magnitude of band. When referring to magnitudes directly taken from the ECO/GALEX pipeline or magnitudes that are used as inputs to SED fitting, we typically use m_{band} . The SED-fitting code applies foreground extinction corrections before fitting.
M_{band}	Absolute magnitude of band, calculated from m_{band} using luminosity distance. In this study, absolute magnitudes are k -corrected and corrected for foreground Milky Way extinction. When referring to magnitudes used as features in our random forest models, we use M_{band} . We may also convert m_{band} either directly from the GALEX/ECO database or from SED-fitting outputs to M_{band} for comparison purposes.
Custom-processed magnitudes	Magnitudes from photometry as described in K. D. Eckert et al. (2015, 2016) and Z. L. Hutchens et al. (2023). These estimates are used as an input to the SED-fitting algorithm, as described in Section 2.5.
Corrected magnitudes	Magnitudes that are internal extinction corrected in addition to the usual foreground extinction and k -corrections. These estimates are derived from our SED modeling and, for NUV, are directly used to compute corrected SFRs, as described in Section 2.5.
Shallow GALEX	NUV magnitude m_{NUV} from GALEX database with the highest exposure time <1000 s. These magnitudes are converted to M_{NUV} and used as inputs to the shallow and shallow+deep random forest models.
Deep GALEX	NUV magnitude m_{NUV} from GALEX database with the highest exposure time ≥ 1000 s. These magnitudes are converted to M_{NUV} and used as inputs to the deep and shallow+deep random forest models.
GALEX pipeline magnitudes	Umbrella term for either shallow or deep GALEX m_{NUV} or M_{NUV} taken from the GALEX database without custom processing.
Shallow-predicted M_{NUV}	M_{NUV} predicted using the shallow random forest model.
Deep-predicted M_{NUV}	M_{NUV} predicted using the deep random forest model.
Shallow+deep-predicted M_{NUV}	M_{NUV} predicted using the shallow+deep random forest model.
No-GALEX-predicted M_{NUV}	M_{NUV} predicted using the no-GALEX random forest model.
RF M_{NUV}	An umbrella term for M_{NUV} predicted by any of our random forest (RF) models, as described in Section 3. These values come from RF models trained to predict custom-processed M_{NUV} and may, depending on the model, use GALEX pipeline M_{NUV} as an input feature.
RF-corrected M_{NUV} or m_{NUV}	Similar to corrected M_{NUV} but where RF M_{NUV} is converted to m_{NUV} and used as an input for SED modeling in place of missing custom-processed M_{NUV} . Our analysis uses these corrected magnitudes to create internal extinction-corrected SFRs for galaxies that lack corrected M_{NUV} from custom-processed data.
SED-output M_{NUV} or m_{NUV}	NUV magnitudes derived from SED modeling that do not reflect internal extinction corrections but do include foreground extinction and k -corrections. In other words, this magnitude is the SED model's refinement of the raw custom-processed or RF m_{NUV} .
SED-inferred M_{NUV} or m_{NUV}	Similar to SED-output M_{NUV} or m_{NUV} but where no input m_{NUV} is provided to the SED-fitting algorithm.

ECO also has internal and foreground extinction- and k -corrected magnitudes and stellar masses derived from the same SED modeling as performed for RESOLVE. In this study, we reran the SED modeling using RF m_{NUV} wherever a custom-processed m_{NUV} was unavailable, with RF m_{NUV} coming from a RF machine trained on the available custom-processed NUV in Z. L. Hutchens et al. (2023). The RF algorithms are explained at length in Section 3. When rerunning the SED algorithm using RF-predicted magnitudes or GALEX pipeline magnitudes, we used a fixed NUV uncertainty of ~ 0.14 mag to allow for random and systematic uncertainties typical of GALEX NUV photometry. We note that for galaxies which do have GALEX NUV data, the NUV uncertainty does not depend strongly or monotonically on magnitude, suggesting that a fixed uncertainty is a reasonable approximation. This new SED modeling does return new stellar mass estimates and corrected magnitudes, a few of which we tabulate in Section 3.3 for future reference, but we only need the corrected NUV. Our analysis does employ stellar masses, but we have chosen to use the stellar masses provided in Z. L. Hutchens et al. (2023) for consistency with all other parameters within ECO Data Release 3 (DR3). We have

confirmed that stellar masses derived from SED modeling that use RF m_{NUV} are consistent with stellar masses for the same subset of galaxies as tabulated in ECO DR3 (Z. L. Hutchens et al. 2023): The median stellar mass offset between the two sets of stellar masses is 0.01 dex. Additionally, 97% of the stellar masses derived from SED modeling in this work are within 0.15 dex of their stellar mass in Z. L. Hutchens et al. (2023) (i.e., within the typical stellar mass uncertainty) and 99.5% are within 0.25 dex.

Following SED modeling, SFRs were calculated for ECO and RESOLVE using both the new internal extinction-corrected NUV magnitudes and mid-IR W3+W4 magnitudes from the Wide-field Infrared Survey Explorer (WISE) as described in M. S. Polimera et al. (2025, in preparation). The NUV and W3+W4 bands are used to compute a nondusty SFR (prescription from S. M. Wilkins et al. 2012) and dusty SFR (prescription from T. Jarrett et al. 2012), respectively. These SFRs were then combined to infer a total SFR following V. Buat et al. (2011).

We also use H I gas masses to compute G/S ratios. Gas masses for RESOLVE were previously derived from non-flux-limited, single-pointing 21 cm line data from the Arecibo and

Green Bank Telescopes as well as archival flux-limited data, primarily from ALFALFA (D. V. Stark et al. 2016; updated in Z. L. Hutchens et al. 2023; see M. P. Haynes et al. 2018 for ALFALFA). D. V. Stark et al. (2013) checked the relative flux calibrations of the ALFALFA data and the single-pointing Arecibo and GBT data, and found them to be consistent (D. Stark 2025, private communication). The non-flux-limited Arecibo and Green Bank observations were performed only for RESOLVE. For ECO galaxies outside RESOLVE-A, gas masses are based on ALFALFA-100 catalog data (M. P. Haynes et al. 2018) with value-added data products provided in Z. L. Hutchens et al. (2023). For both surveys, the best gas estimates for individual galaxies were derived from a combination of 21 cm detections, deconfused detections, upper limits, and the photometric gas fraction technique, as detailed in Z. L. Hutchens et al. (2023). As ALFALFA is a flux-limited survey, the H I masses for ECO galaxies outside RESOLVE-A are more often derived from the photometric gas fraction technique than the H I masses for RESOLVE galaxies are.

This work also uses an AGN catalog for RESOLVE and ECO as developed in M. S. Polimera et al. (2025, in preparation), which extends the work of M. S. Polimera et al. (2022). In short, this catalog includes AGN classification using traditional optical emission-line methods, e.g., BPT (Baldwin-Phillips-Terlevich, J. A. Baldwin et al. 1981) and VO (Veilleux-Osterbrock, S. Veilleux & D. E. Osterbrock 1987) diagrams, WISE photometry, and a new class of AGN (SF-AGN) that are mostly found in low-metallicity and starbursting dwarfs.

2.6. Star Formation Classification

We use the same specific star formation rate (sSFR) division as C24 to classify galaxies in the parent ECO survey into high, medium, and low star formation categories based on double-Gaussian fitting to identify a high star formation locus, a low star formation locus, and a medium star formation region between the two loci in sSFR versus stellar mass. For simplicity, we refer to these three categories as blue, red, and green, respectively.

2.7. Nugget Selection

We applied the selection criteria from C24 to the RF-enhanced parent ECO survey and added it to our previous RESOLVE selection to create a parent RESOLVE+ECO survey. As in C24, we required nugget candidates to be (1) offset to sizes below the R_e - M_* relation for red galaxies, (2) central galaxies in their groups, and (3) nonflybys as defined in Section 2.4. We used the RESOLVE R_e - M_* relation using R_e from PyProFit for all galaxies in RESOLVE. For ECO galaxies not in RESOLVE, we used the R_e - M_* relation based on ECO galaxies with R_e from DECaLS. Not double-counting galaxies that appear in both surveys and excluding ECO galaxies that have neither custom-processed m_{NUV} nor RF m_{NUV} from RF models trained to predict custom-processed NUV data, the size of the parent RESOLVE+ECO sample is 10,018, and the RESOLVE+ECO nugget sample is composed of 1082 nugget candidates (714 blue, 117 green, 251 red).⁶ This new nugget sample contains significantly more galaxies than the RESOLVE nugget sample from C24 (89 blue, 10 green, 42 red).

⁶ We rejected 116 galaxies from our analysis that did not have custom-processed m_{NUV} and for which we could not generate a successful RF m_{NUV} due to having u , g , r , i , or $z = 0$ or $\text{badphot} > 0$ (see Z. L. Hutchens et al. 2023).

3. Methods for Extending ECO with Machine Learning

In this section, we justify the need for RF m_{NUV} from machine learning trained to predict custom-processed NUV data and describe our RF algorithms, which use deep/shallow GALEX and existing parameters within ECO as features to predict RF m_{NUV} for galaxies that lack custom-processed m_{NUV} .

3.1. Justifying the Use of Machine Learning

Custom-processed NUV magnitudes derived from deep GALEX observations are vital for robust extinction corrections via SED modeling. We need extinction corrections to estimate accurate UV-derived SFRs for classifying nuggets, as dust can heavily obscure UV light. This UV sensitivity to dust makes SED-derived extinction corrections themselves sensitive to input NUV magnitudes. Figure 19 in Z. L. Hutchens et al. (2023) highlights the inconsistency of extinction corrections in ECO when custom-processed m_{NUV} data are included as inputs to stellar population synthesis modeling versus when no NUV data are included as inputs. Blue galaxies show a ~ 0.3 dex offset in extinction-corrected $u - r$ in the absence of custom-processed m_{NUV} inputs.

The importance of including high-quality UV data to minimize bias in SED-derived quantities has also been highlighted in other studies. Z. Fan & S. Wang (2017) performed UV/optical/IR SED fitting of stellar systems to probe the effect of WISE/GALEX data on SED modeling, and found that excluding GALEX UV data returns systematically older ages and higher metallicities than when including GALEX data. They also concluded that GALEX UV bands are more crucial for SED fitting than other bands (e.g., WISE W1/W2) and that high-quality UV data are required for robust fitting. S. Salim et al. (2016) derived sSFRs from SED models using UV/optical/IR for a low-redshift sample of $> 7 \times 10^5$ galaxies, and found that including UV data systematically drives sSFRs derived from SED fitting to higher values than when UV data are omitted from fitting, which is consistent with our findings. S. Salim et al. (2016) also found that having deeper GALEX data reduces the random uncertainty on SFR, which suggests that having deeper GALEX data improves SED fitting. Thus, high-quality UV data are necessary to minimize bias in SED outputs.

Using low-quality GALEX data, such as the default shallow imaging of the GALEX AIS survey (~ 150 s), does not yield robust m_{NUV} for our purposes. K. D. Eckert et al. (2016) found that shallow GALEX imaging was inadequate for recovering NUV flux, so they only performed NUV photometry for ECO galaxies with GALEX exposure time ≥ 1000 s. To verify their conclusions, we have compiled deep and shallow GALEX pipeline magnitudes for 4277 galaxies with both. Figure 1 shows the difference in quality between deep (> 1000 s) GALEX pipeline data and shallow (< 1000 s) GALEX pipeline data. The median offset between deep GALEX m_{NUV} and shallow GALEX m_{NUV} is -0.06 . The offset is best in the brightest bin (-0.02) and worst in the faintest bin (-0.17). We also compare SFRs calculated using internal extinction/ k -corrected M_{NUV} based on SED fitting that used shallow versus deep GALEX m_{NUV} data. The median fractional SFR error is 0.05, with the median fractional SFR errors in the highest and lowest SFR bins being 0.02 and 0.15, respectively. The scatter within the 68th percentile lines extends to ~ 0.2 – 0.4 dex at the faint end.

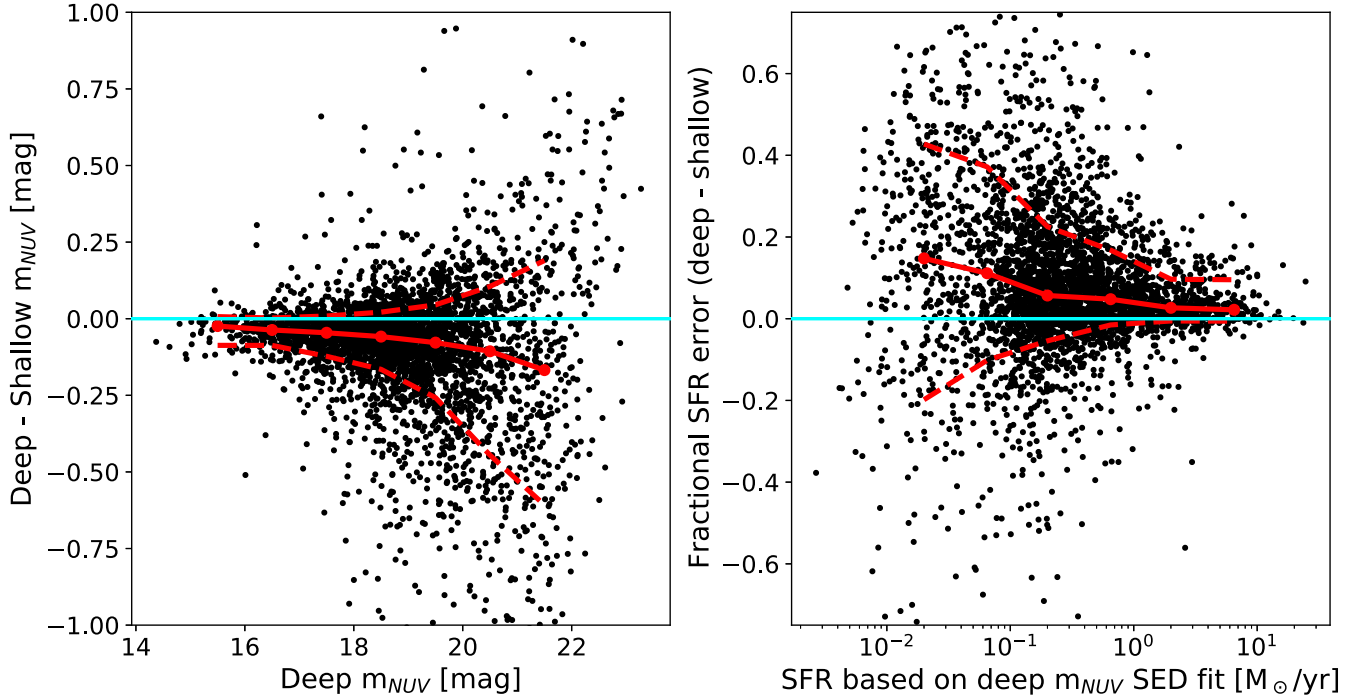


Figure 1. Comparison of pipeline GALEX data based on deep vs. shallow exposures (>1000 s or <1000 s, respectively). In both plots, black dots represent ECO galaxies with both deep and shallow GALEX data, and the cyan line marks $y = 0$. The solid red lines indicate the binned medians and the dashed red lines represent binned 84th percentiles and 16th percentiles. Left: deep m_{NUV} –shallow m_{NUV} vs. deep m_{NUV} . Deep m_{NUV} estimates are systematically brighter than shallow GALEX m_{NUV} estimates, especially for fainter galaxies. Right: fractional error in SFRs based on SED fits vs. SFRs based on SED fits using deep GALEX data. For this panel, we separately input pipeline GALEX deep and shallow m_{NUV} to the SED-fitting code and use the resulting corrected M_{NUV} values to calculate SFR_{deep} and $\text{SFR}_{\text{shallow}}$, respectively. The y-axis fractional error is $(\text{SFR}_{\text{deep}} - \text{SFR}_{\text{shallow}}) / \text{SFR}_{\text{deep}}$.

Applying a simple offset correction to shallow GALEX m_{NUV} data would not sufficiently correct for this discrepancy given the huge scatter. Furthermore, we aim to predict RF m_{NUV} not only for ECO galaxies without custom-processed m_{NUV} but also for those without shallow GALEX m_{NUV} . In what follows, we discuss RF models that yield RF m_{NUV} data that can be used for robust SFRs.

3.2. Random Forest Setup

We used machine learning algorithms that take physical parameters from the GALEX and ECO databases to predict RF M_{NUV} from random forest models trained to predict custom-processed M_{NUV} for ECO galaxies without custom-processed M_{NUV} . To start, we performed a cross-match between ECO DR3 and GALEX. We used a $5''$ cross-match radius and extracted the m_{NUV} and exposure time of all GALEX measurements for every ECO galaxy. For each galaxy, we used the deep GALEX m_{NUV} with the highest exposure time ≥ 1000 s as our deep input and the shallow GALEX m_{NUV} with the highest exposure time < 1000 s as our shallow input, if available.

We chose to use the random forest technique as RF algorithms can accept tabular data directly from the ECO database, utilize continuous or discrete features, and provide feature importances. We used the `RANDOMFORESTREGRESSOR` algorithm from `SCIKIT-LEARN`. We used the `SQUARED_ERROR` metric to optimize our RF models as it was found to be stronger than other metrics in J. E. Krick et al. (2020). We adopted a training/test split of 80/20 and the following parameters as features: M_{H} , M_{g} , M_{r} , M_{i} , M_{z} , stellar mass, halo mass, H I gas mass, 90% light radius, μ_{Δ} , and shallow/deep GALEX M_{NUV}

when available, where all non-GALEX properties are from Z. L. Hutchens et al. (2023).⁷ Near-IR magnitudes are not used as they are not uniformly available for all ECO galaxies. We chose to use absolute magnitudes to represent intrinsic, rather than distance-dependent, properties of galaxies. Stellar masses are also included to allow the RF to respond to the scatter in mass-to-light ratios. Halo mass is included to account for environment. We also provide the 90% light radius in kiloparsecs and μ_{Δ} as features to provide information on extent and morphology, respectively. While the H I gas mass is available for ECO galaxies, most of our RF models exclude it as a feature to keep M_{NUV} independent of H I data and thereby maintain the ability to compare SFRs and gas data. Details on which RF models include/exclude H I gas data can be found in Section 3.3 and Table 2.

Hyperparameter selection was performed in two steps. We first used a randomized search spanning a range for each of the following hyperparameters:

1. `N_ESTIMATORS`: 100 to 700 in steps of 10.
2. `MAX_FEATURES`: `SQRT`, `AUTO`.
3. `MAX_DEPTH`: 10 to 110 in steps of 10.
4. `MIN_SAMPLES_LEAF`: 2, 5, 7.
5. `MIN_SAMPLES_SPLIT`: 1, 2, 4.
6. `BOOTSTRAP`: `True`, `False`.

We focused on optimizing these specific hyperparameters as they were found to be the most important for

⁷ Though we do find differences in the predictive power of the deep and shallow RF models, our tests indicate that the numerical value of the exposure time has minimal feature importance ($< 0.02\%$) when used as an input, so we consider it only indirectly via use of deep versus shallow models.

Table 2
Feature Importance

Model	Feature Weights
Deep+shallow	Deep M_{NUV} (90%), shallow M_{NUV} (7%), 90% light radius (0.5%), rest <0.5%.
Deep	Deep M_{NUV} (98%), 90% light radius (0.6%), M_u (0.5%), rest <0.5%.
Shallow	Shallow M_{NUV} (96%), M_u (1.3%), 90% light radius (1%), rest <0.5%.
No-GALEX	M_u (58%), H I gas mass (23%), stellar mass (8%), rest <5%.

RANDOMFORESTREGRESSOR in S. Mucesh et al. (2021). The randomized search (`sklearn.model_selection.RANDOMIZEDSEARCHCV`) performed 80 cross-validated RF iterations using random combinations of the above hyperparameters and returned the set of hyperparameters that performed the best. From there, we performed a cross-validated grid search (`sklearn.model_selection.GRIDSEARCHCV`) centered on the set of best hyperparameters from the randomized search to determine our finalized hyperparameters. For example, if the randomized search returned `N_ESTIMATORS=200`, the grid search would evaluate `N_ESTIMATOR` between 100 and 300 in steps of 50. The finalized hyperparameters for the RF algorithms were as follows:

1. `N_ESTIMATORS`: 150.
2. `MAX_FEATURES`: AUTO.
3. `MAX_DEPTH`: 75.
4. `MIN_SAMPLES_LEAF`: 5.
5. `MIN_SAMPLES_SPLIT`: 4.
6. `BOOTSTRAP`: True.

We found that these hyperparameter values performed well for all the RF models we explored.

3.3. Random Forest Models

We created four RF models that differ based on their GALEX and H I gas data inputs. The features used for the four models are as follows:

1. Shallow model: includes M_u , M_g , M_r , M_i , M_z , stellar mass, halo mass, 90% light radius in kiloparsecs, μ_Δ , and shallow GALEX M_{NUV} .
2. Deep model: includes M_u , M_g , M_r , M_i , M_z , stellar mass, halo mass, 90% light radius in kiloparsecs, μ_Δ , and deep GALEX M_{NUV} .
3. Shallow+deep model: includes M_u , M_g , M_r , M_i , M_z , stellar mass, halo mass, 90% light radius in kiloparsecs, μ_Δ , deep GALEX M_{NUV} , and shallow GALEX M_{NUV} .
4. No-GALEX model: includes M_u , M_g , M_r , M_i , M_z , stellar mass, halo mass, 90% light radius kiloparsecs, μ_Δ , and H I gas mass.

While K. D. Eckert et al. (2016) derived custom-processed m_{NUV} for most ECO galaxies with deep GALEX data, we found ~ 300 galaxies with deep GALEX pipeline m_{NUV} that did not have custom-processed m_{NUV} , hence the use of deep and deep+shallow models. For the no-GALEX model, we only use galaxies with H I gas mass derived from clean 21 cm detections, strong upper limits, or deconfused observations as indicated in Z. L. Hutchens et al. (2023). Galaxies used to train the shallow, deep, and shallow+deep models were also used to train the no-GALEX model, assuming adequate H I gas estimates, by ignoring GALEX data. These four models were trained and tested on all galaxies that have acceptable values for every feature in their model—the total number of train+test galaxies for the shallow+deep, deep, shallow, and no-GALEX

models were 3457, 3811, 3567, and 2009, respectively. Table 2 lists each model’s top three feature importances. When available, the most important features are shallow and deep GALEX pipeline M_{NUV} . This result is expected as GALEX pipeline M_{NUV} and RESOLVE/ECO custom-processed M_{NUV} are measurements of the same intrinsic property, albeit at different levels of quality. For the no-GALEX model, the most important features were M_u , H I gas mass, and stellar mass. We note that the no-GALEX model provided very poor results when excluding H I mass, so we included H I mass.

We looked at the difference between custom-processed M_{NUV} and RF M_{NUV} from RF models trained to predict custom-processed M_{NUV} for test galaxies, and found that the RF models perform well overall, with higher accuracy correlating with the availability of better GALEX data. To quantify the variations that each of our four RF models can experience, we generated 100 RF regressions using 100 unique test/train splits for each model and evaluated the typical magnitude differences in test galaxies. The magnitude difference is defined as custom-processed M_{NUV} minus RF M_{NUV} and is denoted as ΔM , while the distribution of ΔM values for *all* test galaxies for a single RF run will be referred to as ΔM_{dist} . Figure 2 shows distributions composed of 100 means and 100 standard deviations, where each mean and standard deviation comes from one ΔM_{dist} , for each of the four models. The medians of the 100 ΔM_{dist} means for the shallow+deep, deep, shallow, and no-GALEX models are 0.001 mag, 0.0003 mag, 0.0005 mag, and 0.002 mag, respectively.⁸ The medians of the 100 ΔM_{dist} standard deviations for the shallow+deep, deep, shallow, and no-GALEX models are 0.271 mag, 0.280 mag, 0.292 mag, and 0.415 mag, respectively. Thus, the shallow, deep, and shallow+deep models yield fairly similar results with increasing accuracy, while the no-GALEX model displays an offset and increased scatter. Based on these results, we rank the performance of the models from best to worst as follows: (1) shallow+deep, (2) deep, (3) shallow, and (4) no-GALEX.

We find that our shallow-predicted M_{NUV} values match custom-processed M_{NUV} values better than shallow GALEX pipeline M_{NUV} values do. The vertical lines in Figure 2 show the mean and standard deviation of custom-processed M_{NUV} minus shallow GALEX M_{NUV} for ECO galaxies with both measurements. Comparing shallow GALEX pipeline M_{NUV} to our shallow-predicted model M_{NUV} illustrates the effectiveness of the RF technique. Shallow GALEX pipeline M_{NUV} values have a mean offset of -0.167 mag and a standard deviation of 0.347 mag from custom-processed M_{NUV} values, whereas shallow-predicted M_{NUV} values have a mean offset of -0.0014 mag and a median standard deviation of 0.311 mag based on the ΔM_{dist} statistics in Figure 2. Not only does the shallow model almost entirely correct the flux deficit of shallow GALEX pipeline data, but it also reduces the scatter relative to

⁸ We also calculated the medians of 100 ΔM_{dist} medians for the shallow+deep, deep, shallow, and no-GALEX models: they are 0.029 mag, 0.027 mag, 0.027 mag, and 0.006 mag, respectively.

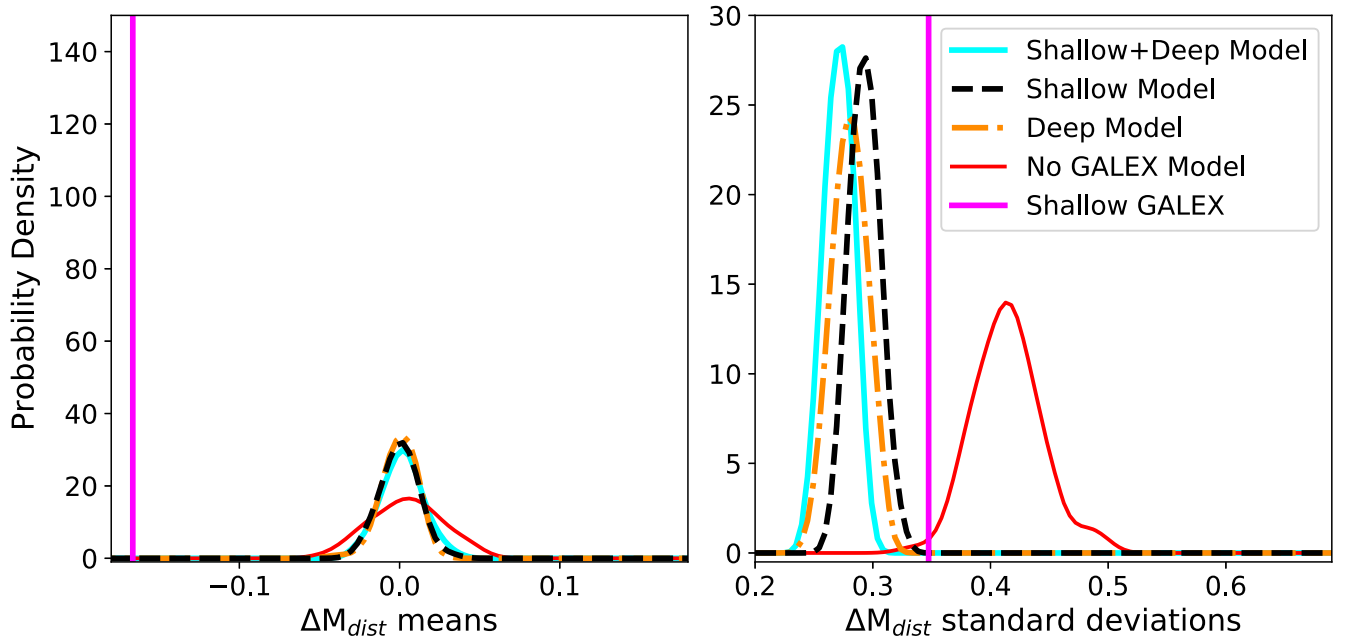


Figure 2. Performance of random forest (RF) model predictions and shallow GALEX pipeline magnitudes compared to custom-processed magnitudes. Curves show distributions of means and standard deviations of ΔM_{dist} (where each ΔM_{dist} is a distribution of the differences ΔM between custom-processed M_{NUV} and RF M_{NUV} for all test galaxies in a single RF regression), where 100 independent RF regressions were performed for each model. Each distribution in this figure is composed of 100 data points, where a data point is the mean (left panel) or the standard deviation (right panel) of ΔM_{dist} from one RF regression with a unique test/train split. For comparison, vertical magenta lines mark the mean (left panel) and standard deviation (right panel) of the differences between shallow GALEX M_{NUV} and custom-processed M_{NUV} for ECO galaxies with both. Models that use GALEX data return comparable mean results, with slightly higher standard deviations for shallower data. The no-GALEX model has a noticeably worse performance. The shallow model corrects the large offset seen in shallow GALEX pipeline data.

custom-processed NUV data. This reduced scatter means that our shallow model performs better than would a simple offset correction based on the difference between shallow GALEX M_{NUV} and custom-processed M_{NUV} data.

Table 3 lists RF-based photometry for ECO galaxies that lack custom-processed NUV data in Z. L. Hutchens et al. (2023). In `nuvrfshallow`, we provide RF M_{NUV} based on just the shallow model. In `nuvrfgalex`, we provide the best RF M_{NUV} estimate, where the order from best to worst of our models is deep+shallow, deep, and shallow. In `nuvrfbest`, we provide the same information but add the last-choice option of the no-GALEX model, which uses HI as an input feature. Finally, `nuvrfbestflag` identifies the source of the best M_{NUV} estimate in `nuvrfbest`. A flag of 0 indicates custom-processed M_{NUV} already existed for the galaxy, and -1 indicates that the galaxy has neither custom-processed M_{NUV} nor M_{NUV} from the RF models due to missing or unreliable input features (see footnote 6). We also provide RF-corrected M_{NUV} , M_u , and M_r based on SED fits using `nuvrfbest` as the NUV input.⁶

Unsurprisingly, our shallow-predicted m_{NUV} values improve on the SED-inferred m_{NUV} values in the current ECO database, which were inferred with no input NUV at all. Last modified in S. J. Kannappan et al. (2013), the SED-fitting algorithm uses available UV/optical/near-IR photometry and Bayesian stellar population modeling to infer intrinsic magnitudes and internal extinction and k -corrections. Thus, it outputs magnitudes in bands without input magnitudes. Currently, the best ECO database estimate for m_{NUV} for galaxies without custom-processed m_{NUV} is this SED-fit inferred value. Figure 3 shows $\text{NUV} - r$ versus stellar mass for ECO galaxies with custom-processed m_{NUV} (solid black contours) compared to ECO galaxies without custom-processed m_{NUV} (dashed contours in

both panels). We use the same stellar masses and m_r for all contours, so differences in the contours represent the method used to estimate m_{NUV} . The left panel uses SED-inferred m_{NUV} , and the right uses shallow-predicted RF m_{NUV} . As noted in Z. L. Hutchens et al. (2023), SED inference alone, with no NUV input, yields unusably inconsistent colors. In contrast, the contours derived from shallow-predicted RF m_{NUV} follow the contours derived from custom-processed m_{NUV} . Thus, we have successfully enhanced ECO by providing RF m_{NUV} estimates that are comparable in quality to custom-processed m_{NUV} data.

To interpret the possible influence of cosmic variance on our machine learning models, we have separated the parent ECO survey into 10 equal-sized subvolumes split by R.A., then recreated the four RF models 10 times where each time one subvolume is excluded from the training. Using a test set composed of a random selection of 20% of all galaxies and a training set composed of all nontest galaxies in the jackknifed volume (i.e., the full ECO volume with one subvolume removed), we found that all the models trained on different jackknifed volumes perform comparably to each other. For example, the means of the difference between the observed and predicted magnitudes for test galaxies in the jackknifed volumes for the shallow model ranged from -0.008 to -0.002 . The standard deviations of the differences between the observed and predicted magnitudes for test galaxies for the shallow model ranged from 0.302 to 0.309. The no-GALEX, deep, and shallow+deep models also show similar tightness in means and standard deviations for all jackknifed volumes. Given the strong agreement in RF predictions between all the jackknifed volumes, we conclude that cosmic variance does not strongly influence our machine learning models.

Table 3
RF-derived M_{NUV}

Column	Designation	Description
1	name	ECO galaxy name.
2	nuvrfshallow	RF M_{NUV} based on shallow model.
3	nuvrfgalex	RF M_{NUV} based on best model, excluding no-GALEX.
4	nuvrfbest	RF M_{NUV} based on best model, including no-GALEX.
5	nuvrfbestflag	Flag corresponding to best M_{NUV} , 0 if custom-processed M_{NUV} exists.
6	nuvcorr	Corrected M_{NUV} based on SED fitting where <i>nuvrfbest</i> is used as an input.
7	ucorr	Corrected M_u based on SED fitting where <i>nuvrfbest</i> is used as an input.
8	rcorr	Corrected M_r based on SED fitting where <i>nuvrfbest</i> is used as an input.

Notes. (1) For Column (3), best available model was used, going from best to worst: (1) shallow+deep model, (2) deep model, (3) shallow model. (2) Column (4) differs from Column (3) in allowing the no-GALEX model as a last choice. (3) For Column (5), the flag values correspond to the following: value of -1 : could not extract a successful RF M_{NUV} due to null feature(s), 0: custom-processed M_{NUV} already existed, 1: RF M_{NUV} predicted by the shallow+deep model, 2: RF M_{NUV} predicted by the deep model, 3: RF M_{NUV} predicted by the shallow model, 4: RF M_{NUV} predicted by the no-GALEX model. All RF M_{NUV} come from RF models trained to predict custom-processed M_{NUV} , not GALEX M_{NUV} . Absolute magnitudes may be converted to apparent magnitudes using Equation (1).

We fed these RF m_{NUV} estimates to the SED-fitting algorithm, and found that the resulting uncertainties on SED-derived output magnitudes were roughly consistent with the uncertainties for the input magnitudes to SED modeling. Input photometric uncertainties were derived as described in K. D. Eckert et al. (2015; and tabulated in Z. L. Hutchens et al. 2023). The SED model adds an additional uncertainty component, which is an estimate of the reproducibility between different surveys due to differences in background subtraction (M. R. Blanton et al. 2011) and other methodologies such as extrapolation. These added components are 0.1 mag in NUV, 0.046 mag in u , 0.022 mag in $griz$, and 0.1 mag in JHK . Figure 4 shows the difference of input magnitude minus SED-fit output magnitude (the latter without k -corrections or internal extinction corrections to match the input magnitude) for the $ugriz$ bands, with separate panels for galaxies with custom-processed m_{NUV} (left) versus with RF m_{NUV} (right) as the input data for SED fitting. The left panel reveals small systematic offsets for ECO galaxies with custom-processed m_{NUV} , where the median offsets of the $ugriz$ distributions are 0.072, 0.033, 0.033, 0.034, and 0.041, respectively. These values are only slightly larger than the median observational $ugriz$ uncertainties of 0.056, 0.024, 0.024, 0.026, and 0.034, respectively. For the right panel, we subtracted these measured systematic offsets from RF m_{NUV} , then scaled the remaining input–output differences by the inverse of the quadrature-summed input uncertainty. The resulting $ugriz$ distributions have standard deviations of 1.13, 1.09, 1.02, 1.04, and 0.96, respectively. This result suggests that the uncertainties associated with SED fitting increase the quadrature-summed input uncertainties by $\sim 10\%$ or less and can be neglected.

Following this analysis of uncertainties, we obtained k -corrected and internal+foreground extinction-corrected RF M_{NUV} values. These values were used to calculate extinction-corrected SFRs for galaxies in the parent ECO survey that lacked SFRs based on custom-processed NUV data. Finally, we generated the parent RESOLVE+ECO survey and nugget sample using all galaxies with extinction-corrected SFRs as described in Section 2.7.

4. Results

We have constructed our RESOLVE+ECO nugget sample to answer three key questions: (1) Are the halo mass and G/S distributions of green nuggets consistent with their being

nuggets caught in the process of halo quenching? (2) Above the threshold scale ($M_{\text{halo}} = 10^{11.4} M_{\odot}$), can AGN be linked to the blue-to-red nugget transition? (3) Below the threshold scale, can AGN be linked to temporary quenching in nuggets? Below, we devote a subsection to each question.

4.1. Halo Quenching for Green Nuggets

We find that green nuggets are likely nuggets caught in the process of halo quenching. While C24 showed that halo mass was the driver for the blue-to-red nugget transition, C24 had too few green nuggets to conclusively determine whether the halo mass distribution of green nuggets is more consistent with the halo mass distribution of blue nuggets or red nuggets. With RESOLVE+ECO, we have nearly $10\times$ the number of green nuggets, allowing us to draw stronger conclusions. Figure 5 (left panel) shows the halo mass distributions of blue, green, and red nuggets. The halo mass distributions of blue and red nuggets are extremely distinct, as a two-sample Kolmogorov–Smirnov (KS) test between them returns a p -value of 1.22×10^{-15} . For blue and green nuggets, a two-sample KS test returns a p -value of 7.77×10^{-15} , whereas for red nuggets and green nuggets, a two-sample KS test returns a p -value of 0.001. Green nuggets are far more similar to red nuggets than to blue nuggets, but consistent with being a transitional population on the way to quenching.

As shown in Figure 5 (right panel), the baryonic content of our nuggets tells a somewhat consistent story. A two-sample KS test of the G/S distributions of green and red nuggets returns $p = 0.002$, whereas the same test for green and blue nuggets returns $p = 1.44 \times 10^{-15}$. When restricting our nuggets to those with H I gas mass derived from clean 21 cm detections, strong upper limits, or deconfused observations, the green/red and green/blue KS tests return $p = 0.06$ and $p = 6.66 \times 10^{-16}$, respectively. Using the same gas mass restriction as above, the median log G/S for blue, green, and red nuggets are -0.01 , -1.40 , and -1.44 , respectively, and the standard deviations of log G/S are 0.63, 0.40, and 0.34. These findings are consistent with a scenario where green nuggets, like red nuggets, have reduced gas content due to halo quenching, i.e., virial shocks suppressing cold-gas refueling, explaining their suppressed sSFRs. However, green nugget G/S are not very intermediate, being roughly as low as those of red nuggets. Either their intermediate colors reflect very recent gas exhaustion, or they may have significant cold molecular

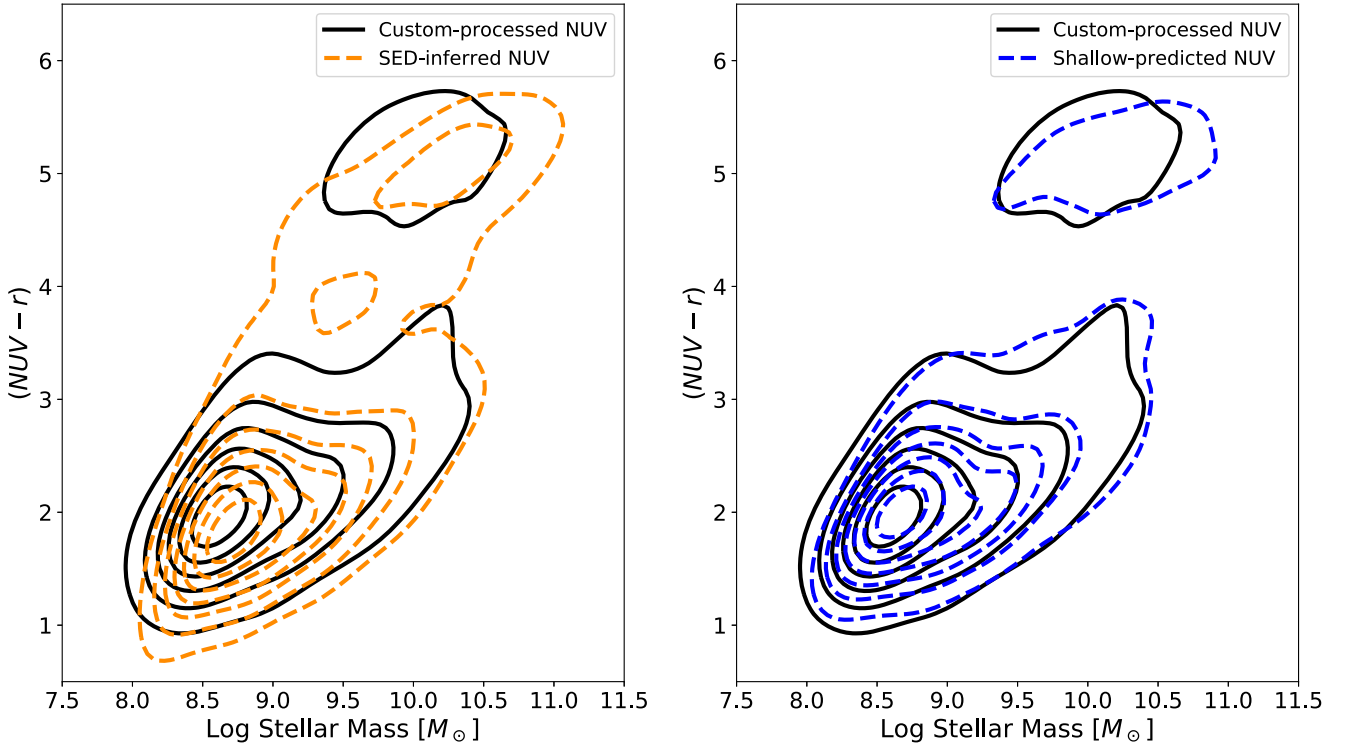


Figure 3. $\text{NUV} - r$ vs. stellar mass for ECO galaxies. In both panels, the solid black contours represent ECO galaxies with custom-processed m_{NUV} from K. D. Eckert et al. (2016) as tabulated in Z. L. Hutchens et al. (2023), while the dashed contours are galaxies that do not have custom-processed m_{NUV} but do have shallow-predicted m_{NUV} . Stellar masses and m_r are from Z. L. Hutchens et al. (2023) in both panels. Left: the dashed contours use m_{NUV} from preexisting SED fitting for ECO (K. D. Eckert et al. 2016), which did not use any NUV input data for those galaxies lacking deep GALEX data. Significant offsets between the two sets of contours can be seen. Right: the dashed contours use shallow-predicted RF m_{NUV} . The stronger agreement demonstrates that our RF models provide much more reliable NUV predictions than can be inferred from SED fits with no input NUV.

gas. Compact starbursts do tend to convert H I to H₂, but also deplete and disperse H₂ via feedback, so molecular gas data would be very interesting as a window into green nugget evolution (see, e.g., D. V. Stark et al. 2013).

4.2. Active Galactic Nuclei in Nuggets above the Threshold Scale

Above the threshold scale ($M_{\text{halo}} \geq 10^{11.4} M_{\odot}$), we find that green nuggets have a higher AGN frequency than blue or red nuggets. C24 evaluated the AGN frequency in each of the nugget categories above the threshold scale, and found AGN to be most common in blue nuggets, but the C24 green nugget sample consisted of only eight objects above $M_{\text{halo}} \geq 10^{11.4} M_{\odot}$. With the new RESOLVE+ECO nugget candidates, we performed the same analysis in the $M_{\text{halo}} \geq 10^{11.4} M_{\odot}$ mass regime, and found that the AGN frequencies (and their associated 1σ binomial uncertainties) in blue, green, and red nuggets are 117/298 (39.2% $^{+2.9\%}_{-2.8\%}$), 41/85 (48.2% $^{+5.4\%}_{-5.4\%}$), and 69/235 (29.3% $^{+3.1\%}_{-2.9\%}$), respectively. Thus, C24 missed that green nuggets have a higher overall AGN frequency than either blue or red nuggets, hinting at AGN playing a role in blue-to-red nugget quenching.

To probe the mass dependence of AGN in nuggets, we evaluated the AGN frequencies of blue, green, and red nuggets in three halo mass bins above the threshold scale: $10^{11.4} M_{\odot} < M_{\text{halo}} < 10^{12} M_{\odot}$, $10^{12} M_{\odot} < M_{\text{halo}} < 10^{12.5} M_{\odot}$, and $M_{\text{halo}} \geq 10^{12.5} M_{\odot}$. Figure 6 shows the AGN frequency of nuggets in the described mass regimes. Green nuggets have a higher AGN frequency (53.8% $^{+7.8\%}_{-8.0\%}$) than

blue/red nuggets (blue = 34.1% $^{+3.1\%}_{-2.9\%}$, red = 30.7% $^{+4.0\%}_{-3.8\%}$) in the bin between the threshold and bimodality scales. In the $10^{12} M_{\odot} < M_{\text{halo}} < 10^{12.5} M_{\odot}$ bin, blue nuggets have a higher AGN frequency (61.4% $^{+7.0\%}_{-7.5\%}$) than green/red nuggets (green = 38.2% $^{+8.6\%}_{-7.9\%}$, red = 26.8% $^{+5.1\%}_{-4.6\%}$). The highest mass bin shows similar but noisier results.

We also compare the SF activity and baryonic content of nuggets with AGN to nuggets without AGN. The bottom panels of Figure 7 show the sSFR and G/S distribution of nuggets in the $M_{\text{halo}} \geq 10^{11.4} M_{\odot}$ mass regime. The p -values from two-sample KS tests for the G/S distribution and the sSFR distribution of these two nugget subpopulations are 0.001 and 0.0009, respectively. When restricting our nuggets to those with H I gas mass derived from clean 21 cm detections, strong upper limits, or deconfused observations, the G/S p -values become 0.001 and 0.001, respectively. While nuggets without AGN above the threshold scale show bimodal distributions in both sSFR and G/S space, nuggets with AGN appear to have a somewhat higher tendency to occupy intermediate regions of sSFR and G/S, both of which are where transitional objects are likely to be found.

Overall, these results for nuggets above the threshold scale are consistent with the suggestion by J. Nogueira-Cavalcante et al. (2019) that green nuggets are experiencing fast-mode quenching, possibly due to AGN feedback, in the presence of virial shocks that prevent cold-mode refueling. The green nugget AGN frequency may be highest right above the threshold scale because AGN feedback accelerates quenching shortly after the nugget begins experiencing virial shocks. It may also be possible that optical AGN are easier to detect in

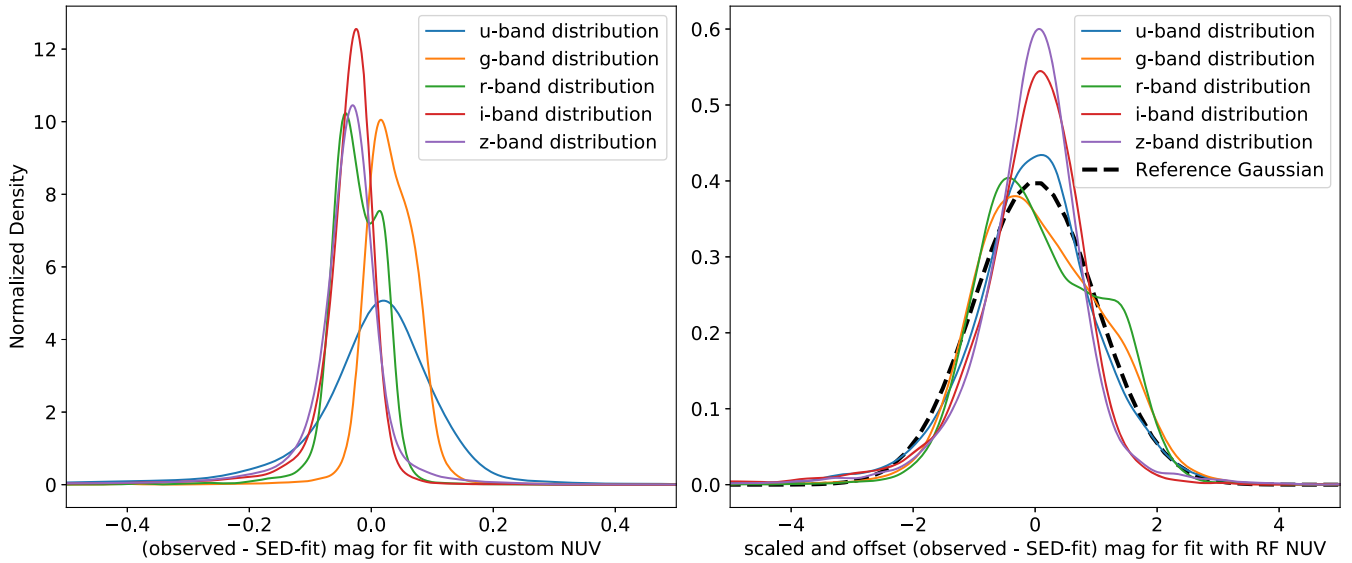


Figure 4. Comparison of *ugriz* input magnitudes and output magnitudes from SED fitting for ECO galaxies. For this comparison, k -corrections and internal extinction corrections inherent in SED-output results were removed, while foreground Milky Way extinctions were matched. Left: input *ugriz* minus SED-output *ugriz* magnitudes for ECO galaxies with custom-processed m_{NUV} . The distributions, which were created using kernel density estimation, show minor systematic <0.1 mag offsets (see Section 3.3 for discussion). Right: input *ugriz* minus SED-output *ugriz* magnitudes for ECO galaxies with RF m_{NUV} , with the systematic offsets measured in the left panel removed and the remaining residuals scaled by the inverse of the quadrature-summed systematic and observational uncertainties (see Section 3.3 for discussion). A Gaussian with a standard deviation of 1 has been included for reference. The standard deviations of the *ugriz* distributions are only marginally larger than 1.0, which suggests that the quadrature-summed observational and systematic input uncertainty is a reasonable proxy for the SED-output magnitude uncertainty.

galaxies with star formation activity below the star-forming main sequence (S. L. Ellison et al. 2016). That said, blue and green nugget AGN frequencies are consistent with each other above $M_{\text{halo}} \geq 10^{12.5} M_{\odot}$, which indicates the need for an even larger sample to improve statistics. We do find that the AGN frequency for blue nuggets is consistent with other studies of high-mass blue nuggets (G. Barro et al. 2013; D. D. Kocevski et al. 2017; E. Wang et al. 2018).

4.3. Active Galactic Nuclei in Nuggets below the Threshold Scale

Below the threshold scale, we find plausible evidence of AGN being associated with temporary quenching in some nuggets. In C24, there were only four AGN-hosting nuggets out of 54 nuggets below the threshold scale, too few to assess relative AGN frequencies for blue/green/red dwarf nuggets. In the study, we find that 44 out of 464 ($9.5^{+1.4}_{-1.3}\%$) nuggets with $M_{\text{halo}} < 10^{11.4} M_{\odot}$ host AGN (31 blue, 10 green, three red), which is roughly consistent with the reported AGN frequency in C24 within much smaller uncertainties. The blue, green, and red nugget AGN fractions below the threshold scale are 31/416 ($7.5^{+1.4}_{-1.2}\%$), 10/32 ($31.3^{+8.7}_{-7.5}\%$), and 3/16 ($18.8^{+11.5}_{-7.8}\%$), respectively. We conclude that for nuggets below the threshold scale, AGN are much more common in partially or fully quenched systems.

We also evaluated whether differences in nugget halo mass functions may impact nugget AGN fractions. In Section 4.1, it can be seen that the three nugget categories each follow distinct halo mass functions. As the blue nugget halo mass function rises toward lower mass, uncertainties in halo mass will cause more blue nuggets in low-mass halos to be scattered toward higher halo masses compared to blue nuggets in high-mass halos scattered toward lower halo masses. The scattering effect for red and green is less severe because blue nuggets have the strongest slope in their halo mass function; so, for brevity, we describe what happens to

the blue nugget population. To test the robustness of our findings to asymmetrical scatter, we added random errors to the halo masses for our final nugget sample, sampled from a Gaussian distribution with $\mu = 0$ and $\sigma = 0.3$ dex (which roughly corresponds to 1σ uncertainties per Figure 14 of Z. L. Hutchens et al. 2023). We performed this sampling 1000 times and evaluated the mean and standard deviation of the AGN fraction in each halo mass bin. After performing these Monte Carlo simulations, the mean AGN fractions of the $M_{\text{halo}} < 10^{11.4} M_{\odot}$, $10^{11.4} M_{\odot} \leq M_{\text{halo}} < 10^{12} M_{\odot}$, $10^{12} M_{\odot} \leq M_{\text{halo}} < 10^{12.5} M_{\odot}$, and $M_{\text{halo}} \geq 10^{12.5} M_{\odot}$ halo mass bins are 12.1%, 22.4%, 42.6%, and 57.1%, respectively. The standard deviations of AGN fractions for those same bins are 0.9%, 1.8%, 4.6%, and 6.7%. The trend of increasing AGN fraction with respect to halo mass seen in our results (AGN fractions of 7.5%, 34.1%, 61.3%, and 75% for the $M_{\text{halo}} < 10^{11.4} M_{\odot}$, $10^{11.4} M_{\odot} \leq M_{\text{halo}} < 10^{12} M_{\odot}$, $10^{12} M_{\odot} \leq M_{\text{halo}} < 10^{12.5} M_{\odot}$, and $M_{\text{halo}} \geq 10^{12.5} M_{\odot}$ halo mass bins, respectively) is still present, albeit weaker when adding artificial extra errors on top of the true errors to the blue nugget halo masses. Therefore, we find that the nugget halo mass function cannot produce the trend we see in nugget AGN fractions as a function of halo mass, and the underlying trend is likely stronger than we measured.

The top panels of Figure 7 show sSFR and G/S distributions for nuggets below the threshold scale ($M_{\text{halo}} < 10^{11.4} M_{\odot}$) with and without AGN. In general, dwarf nuggets with AGN have lower G/S and sSFR compared to dwarf nuggets without AGN, as seen for more massive nuggets in Section 4.2. However, AGN/non-AGN differences are more extreme at low mass: Dwarf nuggets with AGN show bimodal traits in G/S and somewhat bimodal traits in sSFR, whereas dwarf nuggets without AGN are almost entirely unquenched. Performing a two-sample KS test on the G/S distributions of nuggets with AGN and nuggets without AGN returns a p -value of 0.0005. The same test performed on the sSFR distributions of nuggets with AGN and nuggets without AGN returns a p -value of

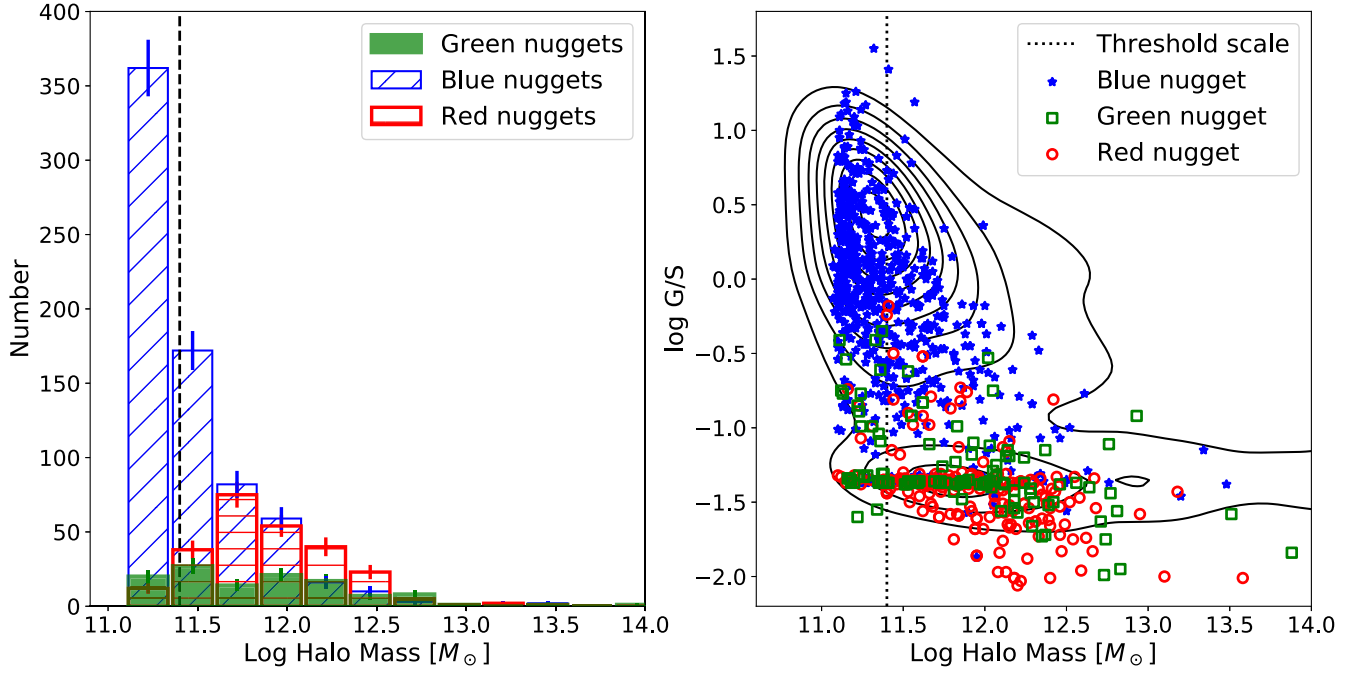


Figure 5. Halo mass distributions and G/S as a function of halo mass for ECO+RESOLVE nugget candidates. Black dashed lines in both panels represent the threshold scale. Two-sample KS tests show that green nuggets are much more similar to red nuggets than to blue nuggets in both halo mass and G/S (Section 4.1). Left: error bars indicate 1σ Poisson uncertainties. As our halo masses are derived from halo abundance matching, which assumes a monotonic relationship between group halo mass and group luminosity, the lowest-mass halos are almost exclusively single-galaxy groups, and the galaxy luminosity floor described in Section 2.1 creates an apparent halo mass floor of $M_{\text{halo}} \sim 10^{11.1} M_{\odot}$. Right: contours represent the parent ECO+RESOLVE survey. The tight locus of green and red nuggets at $\log G/S \sim 1.3$ reflects H I mass estimates based on upper limits clustered at $\sim 5\%$ of the stellar mass, as described in Z. L. Hutchens et al. (2023).

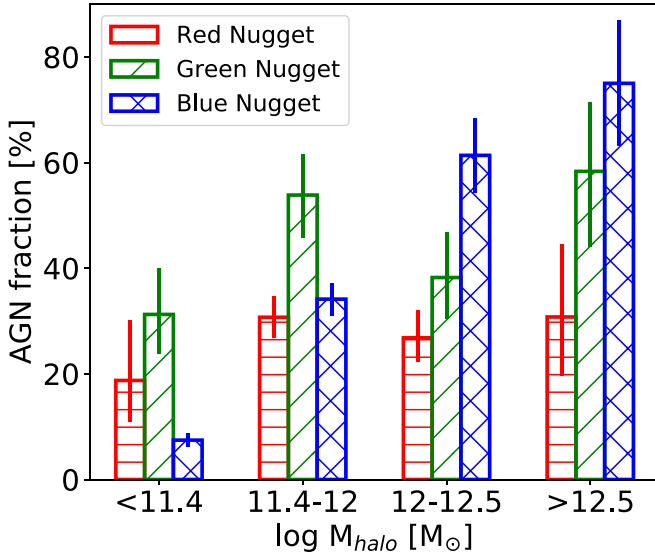


Figure 6. AGN frequencies of blue, green, and red nuggets in four halo mass bins. Binomial 1σ uncertainties are provided for each bar. In the first mass bin, blue nuggets have a notably lower AGN frequency than red or green nuggets. Green nuggets have notably higher AGN frequency than red or blue nuggets in the second mass bin, possibly due to AGN feedback accelerating quenching just above the threshold scale, or conversely quenching allowing AGN to be more active/detectable.

3.37×10^{-6} . Restricting nuggets with/without AGN below the threshold scale to those with H I gas mass derived from clean 21 cm detections, strong upper limits, or deconfused observations reduces the nuggets with AGN sample to just 10 nuggets, too few to effectively compare to nuggets without AGN.

While our results link AGN to temporary quenching in nuggets in low-mass halos ($M_{\text{halo}} < 10^{11.4} M_{\odot}$), this result has multiple interpretations. AGN feedback may cause quenching, but our result is also consistent with the notion that the abatement of star formation feedback could allow for an AGN to turn on (D. Anglés-Alcázar et al. 2017; M. Habouzit et al. 2017; J. D. Bradford et al. 2018; M. S. Polimera et al. 2022). The latter scenario is further degenerate with AGN being easier to detect in galaxies with lower star formation (M. S. Polimera et al. 2022). We note that only $\sim 25\%$ of nuggets below the threshold scale with $\log \text{sSFR} < -1 \text{ Gyr}^{-1}$ have detectable AGN, so it may be that other internal mechanisms, such as stellar feedback, are the primary drivers for temporary quenching.

4.4. Considering Cosmic Variance

We have also probed the potential impact that cosmic variance can have on our key results regarding nuggets. We used the same jackknife approach of iteratively removing subvolumes of the parent ECO+RESOLVE survey as described in Section 3.3 to recompute our statistical results.

Concerning halo mass distributions, green nuggets still differ from blue nuggets with KS test p -values less than 1×10^{-11} , while green and red nuggets differ less, with p -values of 4×10^{-4} to 6×10^{-3} . These results confirm that green nuggets have a halo mass distribution strongly distinct from blue nuggets and somewhat distinct from red nuggets, consistent with active quenching.

Pivoting to AGN fractions, we find that green nuggets still exhibit high AGN fractions between the threshold and bimodality scales. The green nugget AGN fraction between

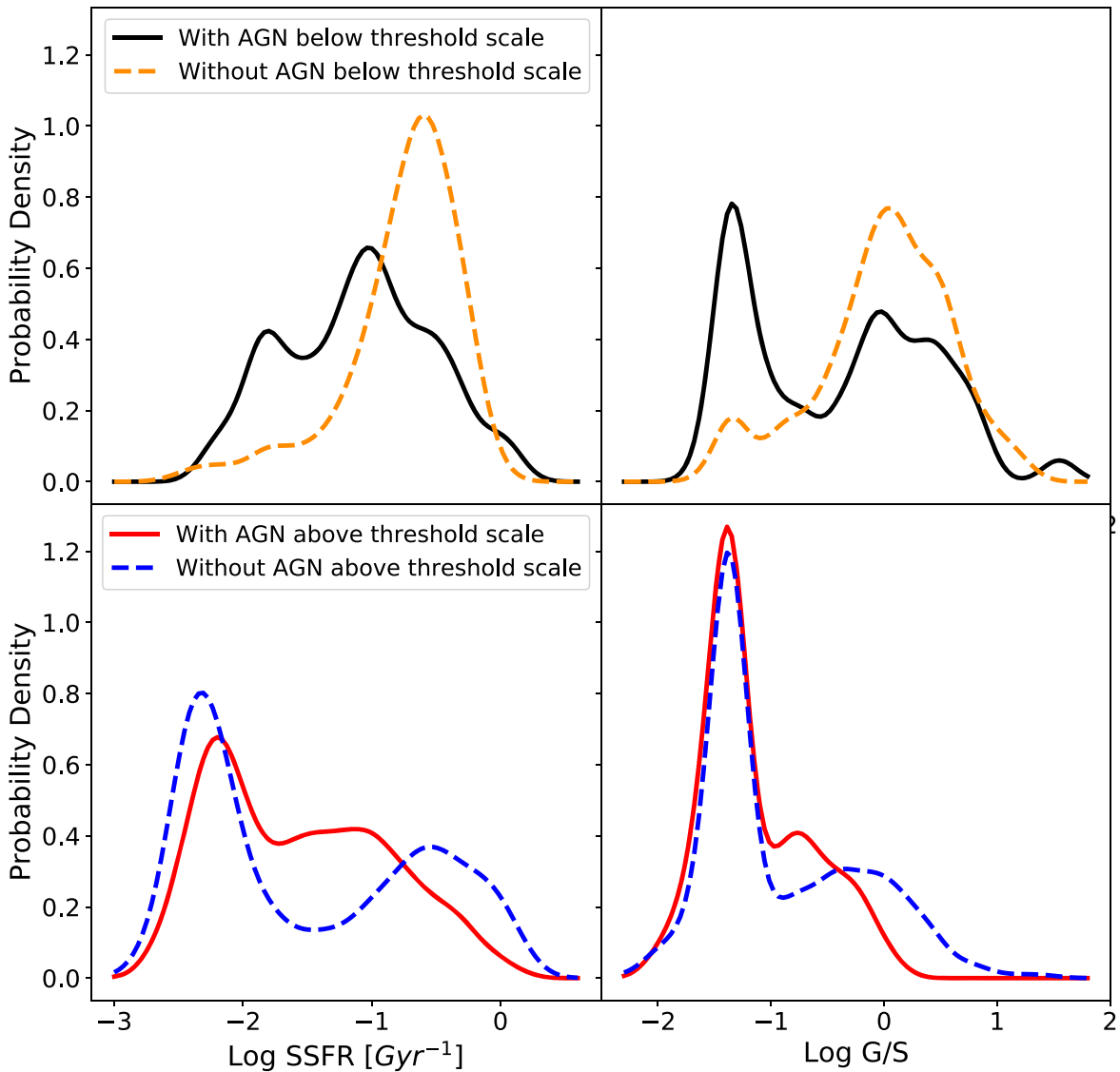


Figure 7. sSFR and G/S distributions for nuggets with AGN and nuggets without AGN in halos both above and below the threshold scale ($M_{\text{halo}} = 10^{11.4} M_{\odot}$). The bandwidths of the kernel density estimations are the typical sSFR uncertainty (0.20 dex) and G/S uncertainty (0.15 dex) for the left and right columns, respectively. At all masses, nuggets with AGN typically have lower G/S and sSFRs compared to nuggets without AGN, but the quenching is more pronounced below the threshold scale.

the threshold and bimodality scales is equal to or greater than 50% in every jackknifed volume. Additionally, the green nugget AGN fraction between the threshold and bimodality scales is at least $1.5\times$ higher than the blue or red AGN fraction between the same scales in every jackknifed volume. Blue nuggets also still have the lowest AGN fraction below the threshold scale compared to green or red nuggets in every jackknifed volume. These results agree with our results in Sections 4.2 and 4.3.

Finally, as in Sections 4.2 and 4.3, we compared the G/S and sSFR distributions of AGN-hosting nuggets and nuggets without AGN, first for just dwarfs and then for just giants, and found KS test p -values between 0.003 and 9.69×10^{-7} for every comparison of the G/S and sSFR distributions. This check confirms our results of AGN-hosting nuggets and non-AGN-hosting nuggets having distinct G/S and sSFR distributions.

In summary, given that our key results persist in jackknifed volumes of the parent RESOLVE+ECO survey, we conclude

that our results for the full nugget census are not significantly impacted by cosmic variance.

5. Conclusions

We have extended the nugget sample of C24 by using machine learning to expand the availability of high-quality NUV data for the ECO survey, thereby creating a parent RESOLVE+ECO survey $\sim 8\times$ the size of the RESOLVE survey used in C24. This larger $z=0$ sample of nuggets at all evolutionary phases has enabled us to better understand how nuggets quench. Specifically, we aimed to answer these three key questions:

1. Are the halo mass and G/S distributions of green nuggets consistent with their being nuggets caught in the process of halo quenching?
2. Above the threshold scale ($M_{\text{halo}} > 10^{11.4} M_{\odot}$), can AGN be linked to the permanent blue-to-red nugget transition?

- Below the threshold scale, can AGN be linked to temporary quenching in nuggets?

In summary:

- We used four random forest models, each calibrated on ECO galaxies with high-quality custom-processed m_{NUV} , to predict high-quality random forest m_{NUV} analogous to custom-processed m_{NUV} for the $\sim 50\%$ of ECO galaxies that lack custom-processed m_{NUV} . The predicted m_{NUV} eliminates the ~ 0.17 mag offset and reduces the ~ 0.35 mag scatter between shallow GALEX pipeline m_{NUV} and custom-processed m_{NUV} . Combining the random forest and existing m_{NUV} estimates in SED fitting, we have doubled the number of ECO galaxies with high-quality extinction-corrected SFRs and blue/green/red classifications based on sSFR versus stellar mass (Section 3 and Figure 2).
- Combining the random forest-enhanced ECO and RESOLVE surveys, we selected nuggets using the same selection criteria as C24. In short, we required that nuggets (1) be offset to sizes smaller than the red stellar mass–effective radius relation, (2) be central galaxies, and (3) not be flyby galaxies (Section 2.7).
- We find that the halo mass distribution of green nuggets is more similar to that of red nuggets, which favors halos above the threshold scale, than to that of blue nuggets, which favors low-mass halos. Green nuggets and red nuggets also have similar atomic gas content. These results suggest that green nuggets are experiencing halo quenching with suppression of cold-gas accretion, though the intermediate colors of green nuggets suggest very recent or incomplete quenching. We lack data on molecular gas, which may be important in this context (Section 4.1 and Figure 5).
- Above the threshold scale ($M_{\text{halo}} \geq 10^{11.4} M_{\odot}$), green nuggets have an overall higher AGN frequency ($48.2\%_{-5.4\%}^{+5.4\%}$) than seen for either blue ($39.2\%_{-2.8\%}^{+2.9\%}$) or red nuggets ($29.3\%_{-2.9\%}^{+3.1\%}$). In the narrow range between the threshold and bimodality scales, $10^{11.4} M_{\odot} < M_{\text{halo}} < 10^{12} M_{\odot}$, green nuggets are nearly $2\times$ more likely to host an AGN than either blue or red nuggets are, hinting at the possibility that AGN play a role in permanent quenching. While AGN-hosting nuggets typically have lower G/S and sSFR than nuggets without AGN at all halo masses, we find that intermediate mildly quenched values are more common above the threshold scale (Section 4.2, Figures 6 and 7).
- AGN are less common in nuggets below the threshold scale, with the AGN frequency for blue nuggets ($7.5\%_{-1.2\%}^{+1.4\%}$) lower than for green nuggets ($31.3\%_{-7.5\%}^{+8.7\%}$) or red nuggets ($18.8\%_{-7.8\%}^{+11.5\%}$). Below the threshold scale, AGN-hosting nuggets are associated with more extreme quenching and gas depletion than seen above the threshold scale. This result may reflect either AGN feedback or enhancement of AGN activity/detectability when star formation abates (Section 4.3, Figures 6 and 7).

The expansion of ECO using machine learning allowed us to draw stronger conclusions regarding comparatively rare galaxy populations, such as green nuggets and nuggets with AGN. Using existing data in the ECO DR3 database together with GALEX pipeline m_{NUV} values, our machine learning models have effectively unified the halves of ECO with and without

custom-processed m_{NUV} . This enhancement provides uniform and robust star formation data as well as multiband extinction corrections for future studies. Thanks to machine learning, we have constructed the first data set able to show that $z = 0$ green nuggets are primary sites of AGN activity, correlated with both temporary quenching and permanent halo quenching. Our enhanced nugget sample showed that green nuggets have the highest AGN frequency above the threshold scale of all nuggets, especially between the threshold and bimodality scales. C24 missed these results due to having too few green nuggets and AGN to robustly analyze. The green nugget–AGN link may imply that AGN play a pivotal role in accelerating the blue-to-red transition, although alternative interpretations need further investigation. With nuggets being the early progenitors of nearly all bulged and early-type galaxies today (I. G. de la Rosa et al. 2016; Y. Gao & L.-L. Fan 2020), this insight into the blue-to-red nugget transition is vital for understanding the full galaxy evolution picture.

Acknowledgments

We thank Adrienne Erickcek, Carl Rodriguez, Fabian Heitsch, and Gerald Cecil for valuable feedback regarding this project. We also thank the anonymous referee for their feedback, which has made the paper significantly more compelling. This research was supported by the National Science Foundation under award AST-2007351. This work was also made possible through additional support from a UNC-CH Graduate School Summer Research Fellowship and a North Carolina Space Grant Graduate Research Fellowship.

Software: MATPLOTLIB, a Python library for publication quality graphics (J. D. Hunter 2007); SCIPY (P. Virtanen et al. 2020); the IPYTHON package (F. Pérez & B. E. Granger 2007); ASTROPY, a community-developed core Python package for Astronomy (Astropy Collaboration et al. 2013, 2018, 2022); NUMPY (C. R. Harris et al. 2020); PANDAS (W. McKinney 2010, 2011).

ORCID iDs

Derrick S. Carr  <https://orcid.org/0009-0007-5354-3817>
 Sheila J. Kannappan  <https://orcid.org/0000-0002-3378-6551>
 Zackary L. Hutchens  <https://orcid.org/0000-0002-8574-5495>
 Mugdha S. Polimera  <https://orcid.org/0000-0001-6162-3963>
 Kathleen D. Eckert  <https://orcid.org/0000-0002-1407-4700>

References

- Anglés-Alcázar, D., Faucher-Giguère, C.-A., Quataert, E., et al. 2017, *MNRAS: Letters*, **472**, L109
- Astropy Collaboration, Price-Whelan, A. M., Lim, P. L., et al. 2022, *ApJ*, **935**, 167
- Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., et al. 2018, *AJ*, **156**, 123
- Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, *A&A*, **558**, A33
- Baldwin, J. A., Phillips, M. M., & Terlevich, R. 1981, *PASP*, **93**, 5
- Barro, G., Faber, S., Pérez-González, P. G., et al. 2013, *ApJ*, **765**, 104
- Blanton, M. R., Kazin, E., Muna, D., Weaver, B. A., & Price-Whelan, A. 2011, *AJ*, **142**, 31
- Bradford, J. D., Geha, M. C., Greene, J. E., Reines, A. E., & Dickey, C. M. 2018, *ApJ*, **861**, 50
- Bruzual, G., & Charlot, S. 2003, *MNRAS*, **344**, 1000
- Buat, V., Giovannoli, E., Heinis, S., et al. 2011, *A&A*, **533**, A93
- Carr, D. S., Kannappan, S. J., Norris, M. A., et al. 2024, *ApJ*, **968**, 18
- Chabrier, G. 2003, *PASP*, **115**, 763
- Charbonnier, A., Huertas-Company, M., Gonçalves, T. S., et al. 2017, *MNRAS*, **469**, 4523

- Cimatti, A., Daddi, E., Renzini, A., et al. 2004, *Natur*, **430**, 184
- Damjanov, I., Hwang, H. S., Geller, M. J., & Chilingarian, I. 2014, *ApJ*, **793**, 39
- Dashyan, G., Silk, J., Mamon, G. A., Dubois, Y., & Hartwig, T. 2018, *MNRAS*, **473**, 5698
- de la Rosa, I. G., La Barbera, F., Ferreras, I., et al. 2016, *MNRAS*, **457**, 1916
- Dekel, A., & Birnboim, Y. 2006, *MNRAS*, **368**, 2
- Dekel, A., & Burkert, A. 2014, *MNRAS*, **438**, 1870
- Dey, A., Schlegel, D. J., Lang, D., et al. 2019, *AJ*, **157**, 168
- Eckert, K. D., Kannappan, S. J., Stark, D. V., et al. 2016, *ApJ*, **824**, 124
- Eckert, K. D., Kannappan, S. J., Stark, D. V., et al. 2015, *ApJ*, **810**, 166
- Eckert, K. D., Kannappan, S. J., Stark, D. V., et al. 2017, *ApJ*, **847**, 2
- Ellison, S. L., Teimoorinia, H., Rosario, D. J., & Mendel, J. T. 2016, *MNRAS: Letters*, **458**, L34
- Fan, Z., & Wang, S. 2017, *Ap&SS*, **362**, 193
- Gao, Y., & Fan, L.-L. 2020, *RAA*, **20**, 106
- Habouzit, M., Volonteri, M., & Dubois, Y. 2017, *MNRAS*, **468**, 3935
- Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, *Natur*, **585**, 357
- Haynes, M. P., Giovanelli, R., Kent, B. R., et al. 2018, *ApJ*, **861**, 49
- Hunter, J. D. 2007, *CSE*, **9**, 90
- Hutchens, Z. L., Kannappan, S. J., Berlind, A. A., et al. 2023, *ApJ*, **956**, 51
- Jarrett, T., Masci, F., Tsai, C., et al. 2012, *AJ*, **145**, 6
- Kannappan, S. J., & Gawiser, E. 2007, *ApJL*, **657**, L5
- Kannappan, S. J., Guie, J. M., & Baker, A. J. 2009, *AJ*, **138**, 579
- Kannappan, S. J., Stark, D. V., Eckert, K. D., et al. 2013, *ApJ*, **777**, 42
- Kannappan, S. J., & Wei, L. H. 2008, in AIP Conf. Proc. 1035, The Evolution of Galaxies through the Neutral Hydrogen Window (Melville, NY: AIP), 163
- Kocevski, D. D., Barro, G., Faber, S., et al. 2017, *ApJ*, **846**, 112
- Krick, J. E., Fraine, J., Ingalls, J., & Deger, S. 2020, *AJ*, **160**, 99
- Lang, D., Hogg, D. W., & Schlegel, D. 2016, *AJ*, **151**, 36
- Manzano-King, C. M., Canalizo, G., & Sales, L. V. 2019, *ApJ*, **884**, 54
- McKinney, W. 2010, in Proc. 9th Python in Science Conf. 445, ed. S. van der Walt & J. Millman (Austin, TX: SciPy), 51
- McKinney, W. 2011, Python for High Performance and Scientific Computing, 14
- Mezcua, M., & Sánchez, H. D. 2024, *MNRAS*, **528**, 5252
- Moffett, A. J., Kannappan, S. J., Berlind, A. A., et al. 2015, *ApJ*, **812**, 89
- Morrissey, P., Conrow, T., Barlow, T. A., et al. 2007, *ApJS*, **173**, 682
- Mucesh, S., Hartley, W. G., Palmese, A., et al. 2021, *MNRAS*, **502**, 2770
- Munoz-Mateos, J., de Paz, A. G., Zamorano, J., et al. 2009, *ApJ*, **703**, 1569
- Nogueira-Cavalcante, J., Gonçalves, T., Menéndez-Delmeire, K., de la Rosa, I., & Charbonnier, A. 2019, *MNRAS*, **484**, 3022
- Palumbo, M. L., Kannappan, S. J., Frazer, E. M., et al. 2020, *MNRAS*, **494**, 4730
- Penny, S. J., Masters, K. L., Smethurst, R., et al. 2018, *MNRAS*, **476**, 979
- Pérez, F., & Granger, B. E. 2007, *CSE*, **9**, 21
- Polimera, M. S., Kannappan, S. J., Richardson, C. T., et al. 2022, *ApJ*, **931**, 44
- Primack, J., Dekel, A., Koo, D., et al. 2018, *ApJ*, **858**, 114
- Salim, S., Lee, J. C., Janowiecki, S., et al. 2016, *ApJS*, **227**, 2
- Saulder, C., van den Bosch, R. C., & Mieske, S. 2015, *A&A*, **578**, A134
- Stark, D. V., Kannappan, S. J., Eckert, K. D., et al. 2016, *ApJ*, **832**, 126
- Stark, D. V., Kannappan, S. J., Wei, L. H., et al. 2013, *ApJ*, **769**, 82
- Tacchella, S., Dekel, A., Carollo, C. M., et al. 2016, *MNRAS*, **457**, 2790
- Trujillo, I., Feulner, G., Goranova, Y., et al. 2006, *MNRAS: Letters*, **373**, L36
- van Dokkum, P. G., Franx, M., Kriek, M., et al. 2008, *ApJL*, **677**, L5
- Veilleux, S., & Osterbrock, D. E. 1987, *ApJS*, **63**, 295
- Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, *NatMe*, **17**, 261
- Wang, E., Kong, X., & Pan, Z. 2018, *ApJ*, **865**, 49
- Wilkins, S. M., Gonzalez-Perez, V., Lacey, C. G., & Baugh, C. M. 2012, *MNRAS*, **427**, 1490
- Zolotov, A., Dekel, A., Mandelker, N., et al. 2015, *MNRAS*, **450**, 2327