



**HAL**  
open science

## Multivariate Predictors of Lyman Continuum Escape. II. Predicting Lyman Continuum Escape Fractions for High-redshift Galaxies

Anne E. Jaskot, Anneliese C. Silveyra, Anna Plantinga, Sophia R. Flury,  
Matthew Hayes, John Chisholm, Timothy Heckman, Laura Pentericci, Daniel  
Schaerer, Maxime Trebitsch, et al.

► **To cite this version:**

Anne E. Jaskot, Anneliese C. Silveyra, Anna Plantinga, Sophia R. Flury, Matthew Hayes, et al.. Multivariate Predictors of Lyman Continuum Escape. II. Predicting Lyman Continuum Escape Fractions for High-redshift Galaxies. *The Astrophysical Journal*, 2024, 973, 10.3847/1538-4357/ad5557. insu-04838845

**HAL Id: insu-04838845**

**<https://insu.hal.science/insu-04838845v1>**

Submitted on 15 Dec 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



# Multivariate Predictors of Lyman Continuum Escape. II. Predicting Lyman Continuum Escape Fractions for High-redshift Galaxies\*

Anne E. Jaskot<sup>1</sup>, Anneliese C. Silveyra<sup>1,2</sup>, Anna Plantinga<sup>3</sup>, Sophia R. Flury<sup>4</sup>, Matthew Hayes<sup>5</sup>, John Chisholm<sup>6</sup>, Timothy Heckman<sup>7</sup>, Laura Pentericci<sup>8</sup>, Daniel Schaerer<sup>9</sup>, Maxime Trebitsch<sup>10</sup>, Anne Verhamme<sup>9,11</sup>, Cody Carr<sup>12,13</sup>, Henry C. Ferguson<sup>14</sup>, Zhiyuan Ji<sup>15</sup>, Mauro Giavalisco<sup>4</sup>, Alaina Henry<sup>14</sup>, Rui Marques-Chaves<sup>9</sup>, Göran Östlin<sup>5</sup>, Alberto Saldana-Lopez<sup>5</sup>, Claudia Scarlata<sup>16</sup>, Gábor Worsceck<sup>17</sup>, and Xinfeng Xu<sup>18</sup>

<sup>1</sup>Department of Astronomy, Williams College, Williamstown, MA 01267, USA

<sup>2</sup>Department of Physics, University of Nevada, Reno, NV 89557, USA

<sup>3</sup>Department of Mathematics & Statistics, Williams College, Williamstown, MA 01267, USA

<sup>4</sup>Department of Astronomy, University of Massachusetts Amherst, Amherst, MA 01002, USA

<sup>5</sup>Department of Astronomy, Oskar Klein Centre, Stockholm University, AlbaNova, SE-10691 Stockholm, Sweden

<sup>6</sup>Department of Astronomy, University of Texas at Austin, Austin, TX 78712, USA

<sup>7</sup>Department of Physics and Astronomy, Johns Hopkins University, Baltimore, MD 21218, USA

<sup>8</sup>INAF—Osservatorio Astronomico di Roma, via Frascati 33, 00078, Monteporzio Catone, Italy

<sup>9</sup>Observatoire de Genève, Université de Genève, Chemin Pegasi 51, 1290 Versoix, Switzerland

<sup>10</sup>Astronomy, Kapteyn Astronomical Institute, Landleven 12, 9747 AD Groningen, The Netherlands

<sup>11</sup>Univ. Lyon, Univ. Lyon 1, ENS de Lyon, CNRS, Centre de Recherche Astrophysique de Lyon UMR5574, 69230 Saint-Genis-Laval, France

<sup>12</sup>Center for Cosmology and Computational Astrophysics, Institute for Advanced Study in Physics, Zhejiang University, Hangzhou 310058, People's Republic of China

<sup>13</sup>Institute of Astronomy, School of Physics, Zhejiang University, Hangzhou 310058, People's Republic of China

<sup>14</sup>Space Telescope Science Institute, 3700 San Martin Dr., Baltimore, MD 21218, USA

<sup>15</sup>Steward Observatory, University of Arizona, Tucson, AZ 85721, USA

<sup>16</sup>Minnesota Institute for Astrophysics, School of Physics and Astronomy, University of Minnesota, 316 Church St. SE, Minneapolis, MN 55455, USA

<sup>17</sup>VDI/VDE Innovation+Technik, Berlin, Germany

<sup>18</sup>Center for Interdisciplinary Exploration and Research in Astrophysics, Northwestern University, Evanston, IL 60201, USA

Received 2024 January 12; revised 2024 May 30; accepted 2024 June 1; published 2024 September 24

## Abstract

JWST is uncovering the properties of ever-increasing numbers of galaxies at  $z > 6$ , during the epoch of reionization. Connecting these observed populations to the process of reionization requires understanding how efficiently they produce Lyman continuum (LyC) photons and what fraction ( $f_{\text{esc}}$ ) of these photons escape into the intergalactic medium. By applying the Cox proportional hazards model, a survival analysis technique, to the Low-redshift Lyman Continuum Survey (LzLCS), we develop new, empirical, multivariate predictions for  $f_{\text{esc}}$ . The models developed from the LzLCS reproduce the observed  $f_{\text{esc}}$  for  $z \sim 3$  samples, which suggests that LyC emitters may share similar properties at low and high redshift. Our best-performing models for the  $z \sim 3$  galaxies include information about dust attenuation, ionization, and/or morphology. We then apply these models to  $z \gtrsim 6$  galaxies. For large photometric samples, we find a median predicted  $f_{\text{esc}} = 0.047\text{--}0.14$ . For smaller spectroscopic samples, which may include stronger emission-line galaxies, we find that  $\geq 33\%$  of the galaxies have  $f_{\text{esc}} > 0.2$ , and we identify several candidate extreme leakers with  $f_{\text{esc}} \geq 0.5$ . The current samples show no strong trend between predicted  $f_{\text{esc}}$  and UV magnitude, but limited spectroscopic information makes this result uncertain. Multivariate predictions can give significantly different results from single-variable predictions, and the predicted  $f_{\text{esc}}$  for high-redshift galaxies can differ significantly depending on whether star formation rate surface density or radius is used as a measure of galaxy morphology. We provide all parameters necessary to predict  $f_{\text{esc}}$  for additional samples of high-redshift galaxies using these models.

*Unified Astronomy Thesaurus concepts:* [Astrostatistics \(1882\)](#); [Reionization \(1383\)](#); [High-redshift galaxies \(734\)](#); [Starburst galaxies \(1570\)](#); [Interstellar medium \(847\)](#); [Ultraviolet astronomy \(1736\)](#); [Radiative transfer \(1335\)](#)

*Materials only available in the [online version of record](#): machine-readable tables*

## 1. Introduction

Reionization represents a fundamental transformation of the Universe's hydrogen gas and dramatically illustrates the effect galaxies can have on their surroundings. The presence of Ly $\alpha$  absorption in quasar spectra suggests that reionization ended sometime near  $z \sim 6$  (e.g., Fan et al. 2006; McGreer et al. 2015; Robertson 2022). The electron scattering optical depth of the cosmic microwave background is consistent with this general picture, constraining the midpoint of reionization to  $z \sim 7.8$  (Planck Collaboration et al. 2020). However, the exact timeline of reionization remains uncertain, in part because of the

\* Based on observations made with the NASA/ESA Hubble Space Telescope, obtained at the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS 5-26555. These observations are associated with programs GO-15626, GO-13744, GO-14635, GO-15341, and GO-15639.

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

inhomogeneous nature of the reionization process (e.g., Eilers et al. 2018; Jung et al. 2020; Becker et al. 2021).

The recently launched James Webb Space Telescope (JWST) is rapidly expanding our knowledge of galaxy properties during the epoch of reionization. Surveys are detecting more bright galaxies than anticipated, challenging our understanding of early galaxy evolution (e.g., Castellano et al. 2022; Naidu et al. 2022; Donnan et al. 2023; Finkelstein et al. 2023; Harikane et al. 2023; McLeod et al. 2024). JWST images suggest that galaxies in the reionization era are morphologically compact (e.g., Morishita et al. 2024; Robertson et al. 2023; Ormerod et al. 2024), and spectroscopic observations are tracing the evolution of nebular ionization and metallicity to redshifts well above  $z = 7$  (e.g., Schaerer et al. 2022; Backhaus et al. 2024; Curti et al. 2023; Fujimoto et al. 2023; Sanders et al. 2023; Tang et al. 2023). Measurements of nebular emission lines in early galaxies are also constraining their ionizing, Lyman continuum (LyC) photon production rate,  $\xi_{\text{ion}}$ , a key quantity needed to understand reionization. Such observations suggest that faint galaxies with bursty star formation histories may have elevated  $\xi_{\text{ion}}$  (e.g., Atek et al. 2024; Simmonds et al. 2024; Saxena et al. 2024).

Several complementary approaches can give insights as to which galaxy populations dominated the reionization process. A galaxy’s contribution to reionization depends on its  $\xi_{\text{ion}}$  and the fraction  $f_{\text{esc}}$  of these LyC photons that escape into the intergalactic medium (IGM). Using constraints on galaxy number densities at  $z > 6$ , some studies explore different distributions of  $\xi_{\text{ion}}$  and  $f_{\text{esc}}$  among the galaxy population and seek to match the observed constraints on the reionization timeline. Depending on the model assumptions, either moderately bright or very faint galaxies may provide the majority of the ionizing photons (e.g., Finkelstein et al. 2019; Naidu et al. 2020).

Cosmological hydrodynamical simulations can also model the progress of reionization over time and can explore how and why LyC photons escape simulated galaxies. Simulations consistently find that feedback plays an essential role in clearing out obscuring material (e.g., Wise & Cen 2009; Cen & Kimm 2015; Paardekooper et al. 2015; Trebitsch et al. 2017). LyC input may peak shortly following a burst of star formation after supernovae explode (e.g., Ma et al. 2015; Trebitsch et al. 2017), although other factors such as binary star evolution or an extended star formation history may affect this timeline (e.g., Ma et al. 2015; Barrow et al. 2020; Katz et al. 2023). The geometry and mechanism of LyC escape may also vary across the galaxy population (e.g., Bremer & Dayal 2023; Katz et al. 2023). Despite general agreement on the importance of feedback, however, simulation predictions have yet to be confirmed observationally, and the galaxies that are the primary drivers of reionization have not yet been conclusively identified.

Directly detecting LyC and constraining  $f_{\text{esc}}$  observationally becomes difficult above  $z > 4$  (e.g., Inoue et al. 2014) due to high IGM attenuation. Fortunately, LyC observations at lower redshifts can help test simulation predictions regarding which galaxy properties regulate LyC escape. Over the past decade, the number of LyC detections has grown rapidly, with dozens of LyC emitters (LCEs) now known at  $z \sim 2\text{--}3$  (e.g., Mostardi et al. 2015; Shapley et al. 2016; Vanzella et al. 2016, 2018; Bassett et al. 2019; Fletcher et al. 2019; Rivera-Thorsen et al. 2019; Ji et al. 2020; Saxena et al. 2022), multiple detections of

LyC emission in stacked samples at  $z \sim 2\text{--}4$  (e.g., Marchi et al. 2018; Steidel et al. 2018; Bian & Fan 2020; Nakajima et al. 2020), and more than 50 known LCEs at  $z < 0.5$  (e.g., Leitert et al. 2011, 2013; Borthakur et al. 2014; Izotov et al. 2016b; Leitherer et al. 2016; Izotov et al. 2018b; Wang et al. 2019; Izotov et al. 2021; Flury et al. 2022a). At both low and intermediate redshift, LCEs appear deficient in absorbing gas and dust (e.g., Chisholm et al. 2018; Gazagnes et al. 2018; Steidel et al. 2018; Ji et al. 2020; Saldana-Lopez et al. 2022), morphologically compact (e.g., Borthakur et al. 2014; Vanzella et al. 2016; Izotov et al. 2018b; Marchi et al. 2018; Rivera-Thorsen et al. 2019; Flury et al. 2022b), and bright in higher-ionization emission lines such as [O III]  $\lambda 5007$  (e.g., Vanzella et al. 2016; Izotov et al. 2018b; Fletcher et al. 2019; Nakajima et al. 2020; Flury et al. 2022b).

Building on these observational and theoretical efforts, several studies have proposed diagnostics to predict  $f_{\text{esc}}$  at  $z > 6$  based on observable properties. Using simulated galaxies from the SPHINX cosmological radiation hydrodynamics simulation, Choustikov et al. (2024) develop a method to predict  $f_{\text{esc}}$  from a linear combination of observables, including the UV slope,  $E(B - V)$ ,  $H\beta$  luminosity, UV magnitude, and nebular line ratios. Most other studies have taken an empirical approach, constructing  $f_{\text{esc}}$  diagnostics based on the large observational sample of LCEs at  $z \sim 0.3$  (e.g., Verhamme et al. 2017; Wang et al. 2021; Flury et al. 2022b; Xu et al. 2023). The largest of these  $z \sim 0.3$  samples is the Low-redshift Lyman Continuum Survey (LzLCS; Flury et al. 2022a), a set of 66 galaxies with LyC measurements from the Hubble Space Telescope (HST) and ancillary ultraviolet and optical data from HST and the Sloan Digital Sky Survey (SDSS; Blanton et al. 2017). Combined with archival data sets (Izotov et al. 2016a, 2016b, 2018a, 2018b; Wang et al. 2019; Izotov et al. 2021), this  $z \sim 0.3$  sample, hereafter the LzLCS+, consists of 89 galaxies with measured LyC or stringent upper limits. Based on an analysis of the LzLCS+, Chisholm et al. (2022) find that the UV slope  $\beta_{1550}$  shows one of the strongest correlations with  $f_{\text{esc}}$  and propose that this single variable can serve as a predictor of  $f_{\text{esc}}$  at high redshift. Other recent studies of the LzLCS+ consider  $f_{\text{esc}}$  diagnostics that incorporate information from multiple variables. Saldana-Lopez et al. (2022) generate an equation to predict  $f_{\text{esc}}$  from  $E(B - V)$  and low-ionization UV absorption lines. By tracing the gas and dust that destroy LyC photons, these parameters closely track  $f_{\text{esc}}$ . Employing this method at high redshift may be a challenge, however, as measuring weak absorption lines requires high signal-to-noise observations of galaxy continua (but see Saldana-Lopez et al. 2023). Emission lines and photometry offer a simpler, if less direct, means of predicting  $f_{\text{esc}}$ . Mascia et al. (2023) propose a new  $f_{\text{esc}}$  diagnostic using  $\beta_{1550}$ ,  $[\text{O III}] \lambda 5007 / [\text{O II}] \lambda 3727 = \text{O}32$ , and half-light radius, three of the variables that correlate strongly with  $f_{\text{esc}}$  in the LzLCS+. By applying this diagnostic at high redshift using JWST observations from the GLASS and CEERS surveys, they find predicted  $f_{\text{esc}}$  of  $\sim 0.1$  for galaxies at  $z > 6$  (Mascia et al. 2023, 2024). With a similar combination of variables ( $\beta_{1550}$ , O32, and UV magnitude), Lin et al. (2024) develop a regression model for the probability of LyC escape. Their model suggests that  $f_{\text{esc}}$  may be high ( $\sim 0.2$ ) in brighter galaxies in the epoch of reionization (Lin et al. 2024). These studies all agree that plausible LCE candidates exist at high redshift, although they differ in their  $f_{\text{esc}}$  prediction methods.

In Jaskot et al. (2024), hereafter Paper I, we use the LzLCS+ to develop new empirical multivariate models for predicting  $f_{\text{esc}}$ . Because the LzLCS+ contains both LyC detections and upper limits, we adopt the statistical techniques of survival analysis, which are suitable for such censored data. Specifically, we employ the semiparametric Cox proportional hazards model to generate  $f_{\text{esc}}$  predictions based on a desired set of input observables. We show that a model limited to observables accessible at  $z > 6$  can reproduce the observed  $f_{\text{esc}}$  in the LzLCS+ with an rms scatter of 0.46 dex. Of these observables, three variables (O32,  $\beta_{1550}$ , and the star formation rate, hereafter SFR, surface density) are statistically significant in the fit, and a model limited to these three input observables predicts  $f_{\text{esc}}$  as well as the full model. The Cox model technique can be customized to include any combination of variables that are available for most of the LzLCS+ galaxies and hence offers a flexible tool for predicting  $f_{\text{esc}}$  at high redshift.

In this paper, we apply Cox models for  $f_{\text{esc}}$  to samples of high-redshift galaxies. Following the techniques in Paper I, which we summarize in Section 2, we generate new models for the sets of variables in published high-redshift samples. In Section 3, we first test the models’ performance at  $z \sim 3$  using samples with published LyC measurements. With published samples at  $z \gtrsim 6$ , we then generate  $f_{\text{esc}}$  predictions for galaxies in the epoch of reionization in Section 4. In Section 5, we compare our models with alternative proposed  $f_{\text{esc}}$  prediction methods from the literature, and we discuss the implications of our results for studies of reionization in Section 6. We summarize our conclusions in Section 7. In the Appendix, we provide parameters for all models as well as examples on how to apply these models to future samples. We adopt a cosmology of  $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$ ,  $\Omega_m = 0.3$ , and  $\Omega_\Lambda = 0.7$ .

## 2. Methods

### 2.1. Sample: The Low-redshift Lyman Continuum Survey

We derive our empirical predictions of  $f_{\text{esc}}$  from the LzLCS+, a combined, homogeneously processed data set consisting of LzLCS (Flury et al. 2022a) and archival samples with HST Cosmic Origins Spectrograph (COS) LyC observations (Izotov et al. 2016a, 2016b, 2018a, 2018b; Wang et al. 2019; Izotov et al. 2021). A full description of the data processing and UV and optical measurements appears in Flury et al. (2022a), Saldana-Lopez et al. (2022), and Paper I, but we summarize key points here.

The LzLCS+ contains 89 galaxies at  $z \sim 0.3$ , a distance where the LyC redshifts into a sensitive wavelength range for the COS detector. The LzLCS targets were selected on properties proposed to indicate LyC escape: high O32 ratios ( $\text{O32} \geq 3$ ), high SFR surface densities ( $\Sigma_{\text{SFR}} > 0.1 M_\odot \text{ yr}^{-1} \text{ kpc}^{-2}$ ), and/or blue UV slopes ( $\beta < -2$ ). The LzLCS+ covers a wide range of parameter space, spanning  $\sim 2$  dex in O32,  $\sim 2$  dex in  $\Sigma_{\text{SFR}}$ , and an observed (not dust-corrected) UV absolute magnitude range of  $M_{1500} = -18.3$  to  $-21.5$ . In Paper I and this study, we exclude one galaxy from the LzLCS+ sample: J1333+6246 (Izotov et al. 2016b). This galaxy has visibly truncated emission lines in its SDSS spectrum and unphysical Balmer line ratios, which together suggest that its nebular line flux measurements may be inaccurate. For this work, our total sample therefore includes 88  $z \sim 0.3$  galaxies, 49 of which have detected LyC.

To measure the LyC, we use COS G140L observations, processed using the `calcos` pipeline (v3.3.9) and the FAINT-COS software routines (Worseck et al. 2016; Makan et al. 2021). We measure the LyC in a 20 Å wide wavelength bin near rest-frame 900 Å, while excluding any wavelengths above 1180 Å in the observed frame because of telluric emission. We follow the definitions of LyC detections and upper limits from Flury et al. (2022a), where detections are observations with a probability  $< 0.02275$  of originating from background counts, and the upper limit for nondetections represents the the 84th percentile of the background count distribution. To correct for Milky Way attenuation, we adopt the Green et al. (2018) dust maps and Fitzpatrick (1999) attenuation law.

Flury et al. (2022a) investigate several different measures of  $f_{\text{esc}}$ . Here, as in Paper I, we adopt the absolute  $f_{\text{esc}}$ , with STARBURST99 (Leitherer et al. 2011, 2014) spectral energy distribution (SED) fits to the UV continuum providing the estimate of the initial intrinsic LyC (Chisholm et al. 2019; Saldana-Lopez et al. 2022). Alternative estimates of the intrinsic LyC from H $\beta$  neglect LyC photons absorbed by dust, assume an isotropic geometry, and require an assumed simplistic star formation history (e.g., Flury et al. 2022a). Another alternative measure of LyC escape is the  $F_{\text{ALYC}}/F_{\lambda 1100}$  flux ratio. As a ratio of two observed fluxes, this method requires no model assumptions. However, we found in Paper I that this quantity was more difficult to predict from easily accessible observables with the survival analysis models. Hence, for the  $f_{\text{esc}}$  models in this paper, we proceed with the absolute  $f_{\text{esc}}$  estimates from the UV SED fits for the LzLCS+.

Optical and UV observations supply information about numerous other properties of the LzLCS+ galaxies (see Flury et al. 2022a; and Saldana-Lopez et al. 2022 for details). We obtain stellar mass ( $M_*$ ) estimates from PROSPECTOR (Leja et al. 2017; Johnson et al. 2019) fits to the SDSS and Galaxy Evolution Explorer (Martin et al. 2003) photometry (Flury et al. 2022a). With multi-Gaussian fits to nebular lines, we measure nebular line fluxes and equivalent widths (EWs), and we estimate the nebular dust attenuation,  $E(B - V)_{\text{neb}}$ , from Balmer line ratios. We calculate the oxygen abundance using the direct method and the `pyneb` package (Luridiana et al. 2015), adopting  $n_e = 100 \text{ cm}^{-3}$  and the estimated [O III]  $\lambda 4363$  flux from the Pilyugin et al. (2006) “ff-relation” in cases where the [S II]  $\lambda\lambda 6716, 6731$  doublet or [O III]  $\lambda 4363$  auroral line are undetected. We estimate SFRs from the Kennicutt & Evans (2012) SFR calibration using the dust-corrected H $\beta$  luminosities and Case B H $\alpha$ /H $\beta$  ratio (Storey & Hummer 1995). The COS near-UV (NUV) acquisition images allow us to measure the UV half-light radius  $r_{50, \text{NUV}}$ , which we then use to calculate  $\Sigma_{\text{SFR}}$  as

$$\Sigma_{\text{SFR}} = \frac{\text{SFR}}{2\pi r_{50, \text{NUV}}^2}. \quad (1)$$

In addition to  $f_{\text{esc}}$ , we derive several other parameters from the HST-COS UV spectra. We obtain estimates of the dust attenuation  $E(B - V)$  from the UV spectrum STARBURST99 SED fits; we label this parameter as  $E(B - V)_{\text{UV}}$  to distinguish it from the nebular dust attenuation  $E(B - V)_{\text{neb}}$  derived from the Balmer lines. Although the G140L spectra do not extend to rest-frame 1500 Å, we can estimate the “observed” (nonextinction corrected) absolute magnitude at 1500 Å ( $M_{1500}$ ) and the power-law index slope at 1550 Å ( $\beta_{1550}$ ) by extrapolating the

SED fits to longer wavelengths (Saldana-Lopez et al. 2022). The inferred  $\beta_{1550}$  values do match the observed values for the few galaxies with existing longer wavelength UV spectra (Chisholm et al. 2022). We also measure Ly $\alpha$  properties from the COS spectra. We derive Ly $\alpha$  fluxes and EWs by linearly fitting the continuum within 100 Å of Ly $\alpha$ , excluding regions affected by nebular or stellar features, and integrating all emission above the continuum. We then derive the Ly $\alpha$  escape fraction ( $f_{\text{esc,Ly}\alpha}$ ) using the dust-corrected H $\beta$  flux and intrinsic Case B Ly $\alpha$ /H $\beta$  ratio (Storey & Hummer 1995) appropriate for the galaxies’ measured electron temperatures and densities. We note that the Ly $\alpha$  measurements represent the net sum of both underlying absorption along the line of sight and scattered Ly $\alpha$  emission within the aperture. Nine galaxies have detectable Ly $\alpha$  absorption troughs that overlap with the Si III  $\lambda$ 1206 absorption feature. We increase our uncertainties to account for the change in flux from omitting wavelengths within 500 km s $^{-1}$  of the SiIII line. The inclusion or exclusion of this region has only minor effects, changing the Ly $\alpha$  EW by  $<3$  Å and with typical changes to  $f_{\text{esc,Ly}\alpha}$  of only 0.001.

## 2.2. Multivariate Survival Analysis: The Cox Proportional Hazards Model

In order to generate multivariate diagnostics for  $f_{\text{esc}}$ , we need to incorporate information from both the  $f_{\text{esc}}$  detections and the upper limits. Within the field of statistics, survival analysis techniques are appropriate for censored data sets that contain limits. One such survival analysis method is the Cox proportional hazards model (Cox 1972; see Clark et al. 2003; and Bradburn et al. 2003; for reviews and Feigelson & Nelson 1985; and Isobe et al. 1986 for examples in astronomy). Here, we describe the basic form of the Cox model and its assumptions. We refer the reader to Paper I for a more thorough discussion of this model and its application to the LzLCS+. To implement the Cox model, we use the CoxPHFitter routine in the lifelines Python package (Davidson-Pilon 2019).

The Cox proportional hazards model predicts the probability of a particular  $f_{\text{esc}}$  value given a set of input variables. Like many other survival analysis techniques, implementations of the Cox model typically assume the data set contains measurements and lower limits (so-called “right-censored” data). In contrast, our dependent variable data consist of  $f_{\text{esc}}$  measurements and associated upper limits. Consequently, we transform our  $f_{\text{esc}}$  values to the absorbed fraction of LyC,  $f_{\text{abs}} = 1 - f_{\text{esc}}$ , for use in the Cox model. However, for ease of interpretation, we put the  $f_{\text{abs}}$  results back into the form of  $f_{\text{esc}}$  in all figures and in the tabulated results in the Appendix.

As applied to the LzLCS+, the Cox proportional hazards regression model fits for the probability of an LyC detection in an infinitesimally small increment of  $f_{\text{abs}}$ , for a given set of independent variables and assuming no detection at a lower value of  $f_{\text{abs}}$  (higher value of  $f_{\text{esc}}$ ). The Cox model assumes a particular functional form for this  $f_{\text{esc}}$  probability, which is known as the “hazard function”:

$$h(f_{\text{abs}}|x) = h_0(f_{\text{abs}}) \exp \left[ \sum_{i=1}^n b_i (x_i - \bar{x}_i) \right]. \quad (2)$$

In this equation,  $b_i$  are the best-fit coefficients for each input variable  $x_i$ , and  $\bar{x}_i$  is the mean value of each input variable in the reference LzLCS+ data set. The term  $h_0(f_{\text{abs}})$  is the baseline hazard function, the probability of having  $f_{\text{abs}}$  in the case where all input variables match their average values within the LzLCS+. The Cox model is semiparametric, in that the dependence on the input variables has a fixed exponential functional form, but the baseline hazard function,  $h_0$ , is estimated nonparametrically. Despite its fixed functional form for the input variable dependence, the Cox model allows the input variables to take any form. For instance, we could define an input variable  $x_i$  as a measured value  $j$ , the logarithm of that measurement  $\log_{10}(j)$ , the square of that measurement  $j^2$ , or any other functional form of our choosing. In our case, we opt to use logarithmic forms of our input variables where possible (e.g.,  $\log_{10}(\text{O}32)$ ,  $M_{1500}$ ,  $\log_{10}(M_*)$ ). In this logarithmic form, most variables have a similar order of magnitude and scale in the same manner, and an order of magnitude increase in a particular input variable simply translates into increasing the probability  $h(f_{\text{abs}}|x)$  by a factor of  $e^{b_i}$ . In this paper, the only variables that we do not use in a logarithmic fashion are EW(Ly $\alpha$ ) and  $f_{\text{esc,Ly}\alpha}$ , because they range from negative (Ly $\alpha$  absorption) to positive (Ly $\alpha$  emission) within the LzLCS+.

Although Equation (2) gives the probability of  $f_{\text{abs}}$ , we would like to predict the expected values of  $f_{\text{abs}}$  and  $f_{\text{esc}}$  and their associated uncertainty. We adopt the median of the probability distribution as this expected value (e.g., Bradburn et al. 2003; Davidson-Pilon 2019); the model predicts that  $f_{\text{esc}}$  will be above this value 50% of the time and below it 50% of the time. To determine this median  $f_{\text{abs}}$  mathematically, we first calculate the survival function,  $S(f_{\text{abs}})$ , the probability that we do not detect LyC at  $f_{\text{abs,detect}} < f_{\text{abs}}$  (i.e., at  $f_{\text{esc,detect}} > f_{\text{esc}}$ ). In the Cox model, the survival function is

$$S(f_{\text{abs}}) = \exp[-\text{HF}_0(f_{\text{abs}}) \cdot \text{ph}(x)], \quad (3)$$

where  $\text{HF}_0$  is the baseline cumulative hazard function

$$\text{HF}_0(f_{\text{abs}}) = \int_0^{f_{\text{abs}}} h_0(f) df, \quad (4)$$

and  $\text{ph}(x)$  is the partial hazards function, which describes how the probability scales with the set of input variables  $x$ :

$$\text{ph}(x) = \exp \left[ \sum_{i=1}^n b_i (x_i - \bar{x}_i) \right] \quad (5)$$

(e.g., Cox 1972; Bradburn et al. 2003; Davidson-Pilon 2019; McLernon et al. 2023). The median  $f_{\text{abs}}$ , which corresponds to our predicted  $f_{\text{esc}}$  value, is the  $f_{\text{abs}}$  value where  $S(f_{\text{abs}}) = 0.5$ . In some cases,  $S(f_{\text{abs}}) > 0.5$ , even for our largest tabulated value of  $f_{\text{abs}}$ , which implies that  $f_{\text{abs}}$  is closer to 1 than we can determine, and  $f_{\text{esc}} \sim 0$ . In this circumstance, we infer that  $f_{\text{esc}}$  is arbitrarily small and report a predicted  $f_{\text{esc}} = 0$ .

The lifelines CoxPhFitter returns the best-fit coefficients  $b_i$  and the cumulative baseline hazard  $\text{HF}_0(f_{\text{abs}})$  for each value of  $f_{\text{abs}}$  corresponding to an LyC detection in the LzLCS+ (see Paper I for more details about the lifelines methodology). In the Appendix, we provide these best-fit parameters for the models in Paper I and the models in this paper, and we give examples of how to use these models to

calculate the predicted  $f_{\text{esc}}$  for a set of observed input variables. The coefficients and hazards can be used to predict the expected  $f_{\text{esc}}$  for any galaxy, as long as it has estimated values for each of the independent variables used in the model.

### 2.2.1. Uncertainty in the Predicted $f_{\text{esc}}$

Because the survival function represents a probability distribution, we can also use it to calculate the expected uncertainty in our predicted  $f_{\text{esc}}$  estimate. The  $f_{\text{abs}}$  values where  $S(f_{\text{abs}}) = 0.159$  and  $S(f_{\text{abs}}) = 0.841$  represent the range in the predicted  $f_{\text{esc}}$  corresponding to the normal theory  $1\sigma$  uncertainty. This uncertainty range reflects the inherent scatter of the relationships between  $f_{\text{esc}}$  and the input variables, where this scatter can come both from measurement uncertainties and from genuine variation among the galaxy population. In Paper I, we tested the effect of measurement uncertainty by performing a Monte Carlo (MC) resampling of each independent and dependent variable according to its observational uncertainty. We found that the distribution of predicted  $f_{\text{esc}}$  from the resampled inputs was nearly always smaller than the  $1\sigma$  bounds inferred from the survival function. In other words, the inherent scatter in the correlations, not the measurement error, is the dominant source of uncertainty in the predicted  $f_{\text{esc}}$ , and the survival function bounds serve as a reasonable estimate of the uncertainty in the  $f_{\text{esc}}$  predictions.

### 2.2.2. Goodness-of-fit Metrics

As in Paper I, we evaluate the Cox models' performance using several complementary metrics. The concordance index,  $C$ , is particularly useful, as it is appropriate for censored data. The concordance index assesses whether the model  $f_{\text{esc}}$  predictions correctly sort the data set in the order of its observed  $f_{\text{esc}}$ . To evaluate this sorting, the concordance index calculation compares each possible pair of data points. Concordant pairs are those where the galaxy with higher observed  $f_{\text{esc}}$  also has higher predicted  $f_{\text{esc}}$ , discordant pairs are the opposite, and tied pairs have an identical predicted  $f_{\text{esc}}$ . Some pairs with upper limits in  $f_{\text{esc}}$  lead to ambiguous rankings and do not appear in the concordance index calculation. With all pairs evaluated, the concordance index is calculated as

$$C = \frac{n_c + 0.5n_t}{n_c + n_d + n_t}, \quad (6)$$

where  $n_c$ ,  $n_d$ , and  $n_t$  are the number of concordant, discordant, and tied pairs, respectively. A value of  $C = 1.0$  indicates a perfect rank ordering, 0.5 is perfectly random ordering, and 0 is perfect disagreement. The major advantage of  $C$  is that it includes both LyC detections and nondetections. However, it only assesses the relative order of the predicted  $f_{\text{esc}}$  values, not their accuracy. To measure the latter, we turn to alternative quantities. One such metric is the  $R^2$  statistic

$$R^2 = 1 - \frac{\sum_i (y_i - f_i)^2}{\sum_i (y_i - \bar{y})^2}, \quad (7)$$

where  $y_i$  are the observed values of  $\log_{10}(f_{\text{esc}})$ ,  $\bar{y}$  is their mean value, and  $f_i$  are the predicted  $f_{\text{esc}}$  values from the Cox model. As in Paper I and Maji et al. (2022), we also calculate a variant of  $R^2$ , the adjusted  $R^2$ , which accounts for the number of free

parameters  $p$  in the model and number of data points  $n$ :

$$R_{\text{adj}}^2 = 1 - (1 - R^2) \frac{n - 1}{n - p - 1}. \quad (8)$$

When increasing the number of variables,  $R_{\text{adj}}^2$  increases only if a variable improves the predictions more than expected by chance. Finally, we report the rms dispersion

$$\text{RMS} = \sqrt{\frac{\sum_i (y_i - f_i)^2}{n}}. \quad (9)$$

As explained in Paper I, we evaluate these three quantities ( $R^2$ ,  $R_{\text{adj}}^2$ , and rms) using  $\log_{10}(f_{\text{esc}})$ . The scatter in the  $\log_{10}(f_{\text{esc}})$  predictions is relatively consistent across the full range of observed  $f_{\text{esc}}$ , whereas the scatter in the linear  $f_{\text{esc}}$  changes systematically across this range. We can only calculate the  $R^2$ ,  $R_{\text{adj}}^2$ , and rms metrics for the LzLCS+ galaxies with both detected LyC and nonzero predicted  $f_{\text{esc}}$ , since these metrics require observed values of  $\log_{10}(f_{\text{esc}})$  and finite predicted  $\log_{10}(f_{\text{esc}})$ . Hence, these metrics only indicate the accuracy of the model predictions for LCEs. However, the most successful models according to the  $R^2$ ,  $R_{\text{adj}}^2$ , and/or rms metrics also tend to have high  $C$  and vice versa (see Paper I).

### 2.2.3. Choice of Input Variables

In generating Cox models from the LzLCS+, we consider a limited set of input variables. Although the dependent variable in the Cox model can contain measurement limits, the independent variables cannot. We therefore only use independent variables that have measurements for nearly all the LzLCS+ galaxies, which requires us to exclude fainter emission-line measurements (e.g., [O I]  $\lambda 6300$  and [S II]  $\lambda \lambda 6716, 6731$ ) or measurements not widely available for the full sample (e.g., Ly $\alpha$  velocity peak separation). In this paper, all 88 galaxies in the LzLCS+ have measurements for our chosen input variables with one exception. The nonleaker J1046+5827 does not have a reported  $\Sigma_{\text{SFR}}$  or  $r_{50,\text{NUV}}$ , and we do not include it in deriving Cox models that use these variables. The Cox model may also fail to converge if we include highly collinear variables, which trace nearly identical properties. Given this limitation, we choose to use only one measure of UV dust attenuation ( $E(B - V)_{\text{UV}}$  or  $\beta_{1550}$ ) per model, and we do not include both  $\Sigma_{\text{SFR}}$  and  $r_{50,\text{NUV}}$  in a single model.

We explored a variety of variable combinations in Paper I, and our best model attained  $R^2 = 0.69$ ,  $R_{\text{adj}}^2 = 0.60$ , rms = 0.31 dex, and  $C = 0.91$ . We provide the parameters for this model in the Appendix. However, this model included the average EW of Lyman-series absorption lines (EW(H I, abs)) and  $f_{\text{esc,Ly}\alpha}$ , both of which will be affected by the IGM at  $z > 6$ . Models in Paper I without UV absorption lines or Ly $\alpha$  showed higher scatter but an overall ability to identify LCEs, with  $R^2 = 0.29\text{--}0.40$ ,  $R_{\text{adj}}^2 = 0.14\text{--}0.35$ , rms = 0.44–0.47 dex, and  $C = 0.83$  (Paper I; see the Appendix for these model parameters). In this paper, we are specifically concerned with variables that are easily observable at  $z > 6$  or measured in large samples at  $z \sim 3$ . Hence, we likewise omit UV absorption line measurements, and we avoid Ly $\alpha$  measurements for most of the models tailored to  $z > 6$  galaxies.

**Table 1**  
Cox Models for High-redshift Predictions

Model	Variables				
	Dust	Ly $\alpha$	Nebular	Luminosity	Morphology
TopThree	$E(B - V)_{\text{UV}}$	...	$\log_{10}(\text{O32})$	...	$\log_{10}(\Sigma_{\text{SFR}})$
LAE	$E(B - V)_{\text{UV}}$	EW(Ly $\alpha$ )	...	$M_{1500}$	...
LAE-O32	$E(B - V)_{\text{UV}}$	EW(Ly $\alpha$ )	$\log_{10}(\text{O32})$	$M_{1500}, \log_{10}(M_*)$	...
LAE-O32-nodust	...	EW(Ly $\alpha$ )	$\log_{10}(\text{O32})$	$M_{1500}$	...
ELG-EW	$E(B - V)_{\text{UV}}$	...	$\log_{10}(\text{EW}([\text{O III}] + \text{H}\beta))$	$M_{1500}, \log_{10}(M_*)$	...
ELG-O32	$E(B - V)_{\text{UV}}$	...	$\log_{10}(\text{O32})$	$M_{1500}, \log_{10}(M_*)$	...
ELG-O32- $\beta$	$\beta_{1550}$	...	$\log_{10}(\text{O32})$	$M_{1500}$	...
ELG-O32- $\beta$ -Ly $\alpha$	$\beta_{1550}$	$f_{\text{esc,Ly}\alpha}$	$\log_{10}(\text{O32})$	$M_{1500}, \log_{10}(M_*)$	...
R50- $\beta$	$\beta_{1550}$	...	...	$M_{1500}, \log_{10}(M_*)$	$\log_{10}(r_{50,\text{NUV}})$
$\beta$ -Metals	$\beta_{1550}$	...	$12 + \log_{10}(\text{O/H})$	$M_{1500}, \log_{10}(M_*)$	...

### 3. Testing the Cox Models at $z \sim 3$

We apply the Cox models developed on the LzLCS+ sample to galaxies at  $z \sim 3$  and  $z \gtrsim 6$ . We use the  $z \sim 3$  galaxies to test whether the models based on the low-redshift LzLCS+ galaxies can successfully predict  $f_{\text{esc}}$  for galaxies at high redshift. We then apply the Cox models to  $z \gtrsim 6$  samples to predict  $f_{\text{esc}}$  for galaxies in the epoch of reionization.

#### 3.1. High-redshift Data Sets and Models

First, we compile samples of  $z \sim 3$  galaxies with reported global absolute LyC  $f_{\text{esc}}$  and at least three of the input variables from Paper I, which include stellar mass  $M_*$ ,  $M_{1500}$ , nebular EWs, metallicity, optical nebular line ratios,  $\Sigma_{\text{SFR}}$ ,  $r_{50,\text{NUV}}$ ,  $E(B - V)_{\text{UV}}$ ,  $\beta_{1550}$ , and Ly $\alpha$  measurements. The  $z \sim 3$  LyC measurements include individual detections, reported LyC upper limits, and stacked samples that average over variations in IGM attenuation (Hainline et al. 2009; de Barros et al. 2016; Vanzella et al. 2016; Vasei et al. 2016; James et al. 2018; Steidel et al. 2018; Bassett et al. 2019; Fletcher et al. 2019; Bian & Fan 2020; Nakajima et al. 2020; Marques-Chaves et al. 2021; Pahl et al. 2021; Marques-Chaves et al. 2022; Kerutt et al. 2024; Liu et al. 2023). We exclude one active galactic nucleus (AGN) from the Fletcher et al. (2019) sample and one shock-dominated galaxy from the Bassett et al. (2019) sample. Some well-known  $z \sim 3$  LCEs do not appear in our sample, because they lack published global absolute  $f_{\text{esc}}$  measurements (e.g., Ion3 and the Sunburst Arc). We also note that some of the high-redshift samples fall outside of the parameter space probed by the LzLCS+, such that our model predictions will extrapolate for these galaxies. The  $z \sim 3$  data have limitations as well; the compiled  $z \sim 3$  samples differ in the methods used for  $f_{\text{esc}}$  and input variable measurements, and individual  $z \sim 3$   $f_{\text{esc}}$  measurements have high uncertainty due to unknown variations in IGM attenuation. We discuss the limitations of the  $z \sim 3$  comparison further in the following sections.

In Paper I, we introduced a JWST Cox model, which could be applied to an ideal  $z > 6$  sample, with eight relevant variables from the LzLCS+ included. We also found that a ‘‘limited JWST Model’’ fit using only the three top-ranked

variables ( $\beta_{1550}$ ,  $\log_{10}(\text{O32})$ , and  $\log_{10}(\Sigma_{\text{SFR}})$ ) performed equally well for predicting  $f_{\text{esc}}$  in the LzLCS+ sample. Unfortunately, many of the required variables for both the full and limited JWST models have not been measured for large samples of  $z \sim 3$  LCEs or  $z > 6$  galaxies. Consequently, we run new models limited to the variables available for the  $z \sim 3$  and for  $z \gtrsim 6$  samples. In Table 1 and Table 2, we list these models, the variables they include, and the samples to which they apply.

For the  $z \gtrsim 6$  samples, we prioritize models that can apply to large samples (e.g., Endsley et al. 2021, 2023; Morishita et al. 2024), models that can apply to faint galaxies (Atek et al. 2024), and models that have at least two of the most statistically significant variables in Paper I (a measure of dust attenuation plus a measurement of ionization or morphology). The ‘‘TopThree’’ model includes only the three top-ranked variables from Paper I: dust, O32, and  $\Sigma_{\text{SFR}}$ . In the TopThree model, we use  $E(B - V)_{\text{UV}}$  rather than  $\beta_{1550}$ , since it enables us to compare with a larger  $z \sim 3$  sample. As noted in Chisholm et al. (2022), the  $\beta_{1550}$  and  $E(B - V)_{\text{UV}}$  measurements for the LzLCS+ track each other almost perfectly and provide equivalent information. Most models in Table 1 use  $E(B - V)_{\text{UV}}$ ; models with ‘‘ $\beta$ ’’ in their name use  $\beta_{1550}$  instead, and one model, LAE-O32-nodust, has no dust attenuation measurement. All models except the TopThree model include  $M_{1500}$ , and most models include  $M_*$  as well. The only models without  $M_*$  are the LAE, LAE-O32-nodust, and ELG-O32- $\beta$  models. Models that include EW(Ly $\alpha$ ) have names that begin with ‘‘LAE’’ for Ly $\alpha$  emitter. Models that lack EW(Ly $\alpha$ ) but include rest-frame optical emission line variables have names that begin with ‘‘ELG’’ for emission-line galaxy.

We provide the best-fit coefficients and cumulative baseline hazards for each model in the Appendix, which can be used to derive the median predicted  $f_{\text{esc}}$  for a given galaxy and the uncertainty from the 16th to 84th percentiles of the  $f_{\text{esc}}$  probability distribution. As with the LzLCS+ galaxies, we use an MC method to sample the variable uncertainties for the  $z \sim 3$  and  $z \gtrsim 6$  observations and regenerate the predicted  $f_{\text{esc}}$ . We again find that the uncertainty estimated from the Cox model survival function dominates over the uncertainty from sampling the input variables in nearly all cases.

**Table 2**  
Cox Model High-redshift Samples

Model	$z \sim 3$ Samples		$z \sim 6$ Samples
TopThree	1 Detection 1 Nondetection 1 with Input Limits	Ion2 <sup>a</sup> Cosmic Horseshoe <sup>b</sup> J1316+2614 <sup>c</sup>	
LAE	28 Detections+13 Stacks  6 Nondetections+5 Stacks 3 with Input Limits	Ion2 <sup>a</sup> , Bassett et al. (2019) Sample, Fletcher et al. (2019) Sample <sup>d</sup> , Liu et al. (2023) Sample <sup>e</sup> , Kerutt et al. (2024) Sample, J0121+0025 <sup>f</sup> , J1316+2614 <sup>c</sup> , Steidel et al. (2018) Stacks <sup>g</sup> Steidel et al. (2018) Stacks <sup>g</sup> , Bassett et al. (2019) Sample, Bian & Fan (2020) Stack Fletcher et al. (2019) Sample <sup>d</sup>	JADES-GS-z7-LA
LAE-O32	4 Detections  5 Nondetections 5 with Input Limits	Ion2 <sup>a</sup> , Bassett et al. (2019) Sample, Fletcher et al. (2019) Sample <sup>d</sup> Bassett et al. (2019) Sample Bassett et al. (2019) Sample, Fletcher et al. (2019) Sample <sup>d</sup> , J1316+2614 <sup>c</sup>	JADES-GS-z7-LA
LAE-O32-nodust	5 Detections+1 Stack  14 Nondetections+2 Stacks 19 with Input Limits	Ion2 <sup>a</sup> , Bassett et al. (2019) Sample, Nakajima et al. (2020) Sample and Stacks, J1316+2614 <sup>c</sup> Bassett et al. (2019) Sample, Nakajima et al. (2020) Sample and Stacks Bassett et al. (2019) Sample, Nakajima et al. (2020) Sample	JADES-GS-z7-LA
ELG-EW	4 Detections 4 with Input Limits	Ion2 <sup>a</sup> , Fletcher et al. (2019) Sample <sup>d</sup> Fletcher et al. (2019) Sample <sup>d</sup> , J1316+2614 <sup>c</sup>	Endsley et al. (2021, 2023) Samples, Bouwens et al. (2023) Sample, Tang et al. (2023) Sample, <sup>h</sup> JADES-GS-z7-LA, Fujimoto et al. (2023) Sample
ELG-O32	4 Detections  5 Nondetections 5 with Input Limits	Ion2 <sup>a</sup> , Bassett et al. (2019) Sample, Fletcher et al. (2019) Sample <sup>d</sup> Bassett et al. (2019) Sample Bassett et al. (2019) Sample, Fletcher et al. (2019) Sample <sup>d</sup> , J1316+2614 <sup>c</sup>	Tang et al. (2023) Sample <sup>h</sup> , 11027 JADES-GS-z7-LA, Fujimoto et al. (2023) Sample
ELG-O32- $\beta$	2 Detections	Ion2 <sup>a</sup> , J1316+2614 <sup>c</sup>	Schaerer et al. (2022) Sample <sup>i</sup> , 11027, Mascia et al. (2023) Sample, Saxena et al. (2024) Sample, JADES-GS-z7-LA Mascia et al. (2023), Saxena et al. (2024, 2024)
ELG-O32- $\beta$ -Ly $\alpha$	...	...	JADES-GS-z7-LA
R50- $\beta$	1 Detection	Ion2 <sup>a</sup>	Mascia et al. (2023) Sample, Morishita et al. (2024) Sample <sup>j</sup>
$\beta$ -Metals	1 Detection 1 with Input Limits	Ion2 <sup>a</sup> J1316+2614 <sup>c</sup>	Schaerer et al. (2022) Sample <sup>i</sup> , Atek et al. (2024) Sample <sup>k</sup> , 11027

**Notes.**

<sup>a</sup> From de Barros et al. (2016), Vanzella et al. (2016, 2020).

<sup>b</sup> From Hainline et al. (2009), Vasei et al. (2016), James et al. (2018). We adopt the average dust-corrected O32 ratio of the two similar-flux regions in Hainline et al. (2009). The  $E(B - V)_{UV}$  and  $\Sigma_{SFR}$  values are flux-weighted averages of the regions in James et al. (2018).

<sup>c</sup> From Marques-Chaves et al. (2022).

<sup>d</sup> From Fletcher et al. (2019), Nakajima et al. (2020).

<sup>e</sup> We adopt the  $f_{esc}$  determined from the SED-fitting method in Liu et al. (2023).

<sup>f</sup> From Marques-Chaves et al. (2021).

<sup>g</sup> From Steidel et al. (2018), Pahl et al. (2021).

<sup>h</sup> We adopt  $M_*$  calculated using the bursty nonparametric star formation history, but include the alternative values in the uncertainties. We use the direct method metallicities for the two galaxies with secure [O III]  $\lambda 4363$  detections and the modeled metallicities for the others.

<sup>i</sup> We correct  $M_*$  for magnification, using the values in Schaerer et al. (2022). For the source without a direct method metallicity, we adopt the strong-line method  $12 + \log_{10}(O/H)$ .

<sup>j</sup> We exclude sources flagged as having significant residuals in the morphological fits.

<sup>k</sup> Metallicities calculated using strong-line methods.



**Table 3**  
Metrics for High-redshift Cox Models

LzLCS+ Sample Model	$N_{\text{Gal}}$	$N_{\text{Detect}}$	$R^2$	$R_{\text{adj}}^2$	rms	$C$
TopThree	87	45	0.38	0.34	0.44	0.82
LAE	88	46	0.21	0.15	0.50	0.81
LAE-O32	88	44	0.37	0.29	0.45	0.82
LAE-O32-nodust	88	43	0.02	-0.06	0.59	0.77
ELG-EW	88	43	0.14	0.05	0.53	0.79
ELG-O32	88	42	0.42	0.36	0.44	0.79
ELG-O32- $\beta$	88	42	0.30	0.25	0.48	0.79
ELG-O32- $\beta$ -Ly $\alpha$	88	43	0.40	0.32	0.45	0.81
R50- $\beta$	87	45	0.36	0.30	0.44	0.84
$\beta$ -Metals	88	44	0.11	0.02	0.55	0.77
$z \sim 3$ Sample Model	$N_{\text{Gal}}$	$N_{\text{Detect}}$	$R^2$	$R_{\text{adj}}^2$	rms	$C$
TopThree	2	1	...	...	0.15	1.00
LAE	52	35	-4.18	...	0.92	0.49
LAE-O32	9	4	-0.35	...	0.32	0.71
LAE-O32-nodust	22	6	-0.12	...	0.31	0.80
ELG-EW	4	4	-4.81	...	0.65	0.50
ELG-O32	9	4	0.09	...	0.27	0.71
ELG-O32- $\beta$	2	2	-0.91	...	0.04	1.00
ELG-O32- $\beta$ -Ly $\alpha$	...	...	...	...	...	...
R50- $\beta$	1	1	...	...	0.09	...
$\beta$ -Metals	1	1	...	...	0.43	...
Combined Sample Model	$N_{\text{Gal}}$	$N_{\text{Detect}}$	$R^2$	$R_{\text{adj}}^2$	rms	$C$
TopThree	89	46	0.43	0.39	0.44	0.83
LAE	140	81	-0.47	-0.53	0.71	0.70
LAE-O32	97	48	0.42	0.35	0.44	0.82
LAE-O32-nodust	110	49	0.19	0.13	0.56	0.78
ELG-EW	92	47	0.16	0.08	0.54	0.79
ELG-O32	97	46	0.48	0.42	0.43	0.80
ELG-O32- $\beta$	90	44	0.40	0.36	0.47	0.80
ELG-O32- $\beta$ -Ly $\alpha$	...	...	...	...	...	...
R50- $\beta$	88	46	0.41	0.35	0.43	0.84
$\beta$ -Metals	89	45	0.17	0.08	0.55	0.77

**Note.**  $N_{\text{Gal}}$  is the number of galaxies assessed in the fit and does not include any galaxies with limits for input variables.  $N_{\text{Detect}}$  is the number of galaxies with LyC detections, finite  $f_{\text{esc}}$  predictions, and no upper or lower limits for input variables; the  $R^2$  and rms metrics use only these galaxies. The  $R^2$  statistic measures how well the predictive model explains the observed variance in the  $\log_{10}(f_{\text{esc}})$  data, with higher  $R^2$  values corresponding to more accurate models. The  $R_{\text{adj}}^2$  metric accounts for the number of parameters used in the model and increases only if a variable improves the fit more than expected by chance. rms is the rms dispersion of the predicted versus observed  $\log_{10}(f_{\text{esc}})$  for the LyC detections. Higher  $C$  values indicate that the model more accurately sorts the observations in the correct order of increasing  $f_{\text{esc}}$ .  $C$  includes both detections and galaxies with  $f_{\text{esc}}$  upper limits. See Section 2.2.2 for a full description of these statistics.

We first test the Cox model predictions on  $z \sim 3$  samples to see whether the high-redshift LCEs behave similarly to the LzLCS+ sample. We derive Cox models for the variable sets in Table 1 using only the LzLCS+ data; the high-redshift samples are not included in the fitting process. We then use the models to predict  $f_{\text{esc}}$  for each set of  $z \sim 3$  measurements, and we calculate goodness-of-fit metrics for the LzLCS+ sample alone, the high-redshift sample alone, and the combined sample of low- and high-redshift galaxies. We list these goodness-of-fit metrics in Table 3.

### 3.2. Model Performance

Figures 1 and 2 show the predicted versus observed  $f_{\text{esc}}$  for the LzLCS+ and  $z \sim 3$  samples. As seen from the plots and the goodness-of-fit statistics in Table 3, these models do not perform as well as the fiducial model from Paper I, but several (TopThree, LAE-O32, ELG-O32, ELG-O32- $\beta$ , ELG-O32- $\beta$ -Ly $\alpha$ , R50- $\beta$ ) are comparable to or better than the Paper I JWST model (see Tables 11–13 for the list of input variables used in the Paper I models). The metrics for the LzLCS+ sample are  $R^2 = 0.02$ –0.42 (versus 0.59 for the fiducial model and 0.29 for JWST),  $R_{\text{adj}}^2 = -0.06$  to 0.36 (versus 0.48 for the fiducial model and 0.14 for JWST), rms = 0.44–0.59 (versus 0.37 for the fiducial model and 0.47 for JWST), and  $C = 0.77$  to 0.84 (versus 0.88 for the fiducial model and 0.83 for JWST).

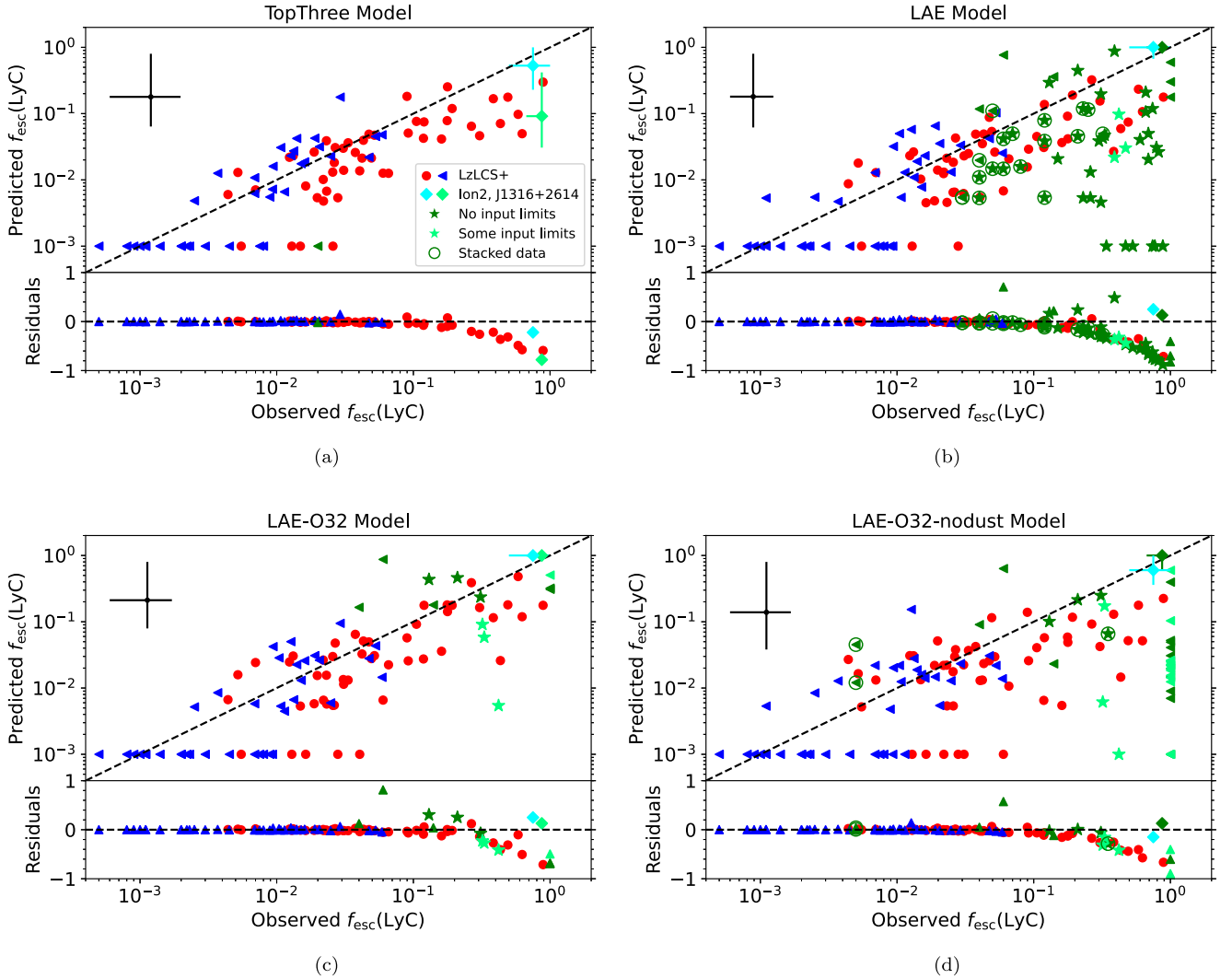
The models developed on the LzLCS+ sample generally reproduce the observed  $f_{\text{esc}}$  values of the  $z \sim 3$  LCEs. In fact, the high-redshift samples often have a lower rms scatter than the predictions for the LzLCS+, and in several models, the  $R^2$  and  $C$  values for the combined high- and low-redshift sample are comparable to or higher than the  $R^2$  and  $C$  values for the LzLCS+ sample alone. Notably, predictions for the two strongest high-redshift LCEs, Ion2 and J1316+2614, match the observations nearly perfectly in most cases (Figures 1 and 2), demonstrating the success of the model in identifying extreme LCEs.

The goodness-of-fit metrics give clues as to which variables are most important in predicting  $f_{\text{esc}}$ . As seen in Table 3, the models with the highest  $R^2$  and lowest rms for the LzLCS+ sample are the TopThree, LAE-O32, ELG-O32 and variants, and the R50- $\beta$  models. These same models are the only ones that include both a measure of UV dust attenuation and either O32 or radius as variables. Turning to the  $C$  metric, which includes nondetections, the best-performing models for the LzLCS+ sample are the R50- $\beta$ , TopThree, LAE-O32, ELG-O32- $\beta$ -Ly $\alpha$ , and LAE models. To properly estimate  $f_{\text{esc}}$  for weak and non-LCEs, UV dust attenuation again appears important, but each of the best-performing models also includes either morphological information or Ly $\alpha$  measurements. These additional variables may help distinguish non-LCEs from LCEs.

By all metrics, the LAE-O32-nodust and  $\beta$ -Metals models perform the worst for the LzLCS+ galaxies. The LAE-O32-nodust model is the only one that lacks  $E(B - V)_{\text{UV}}$  or  $\beta_{1550}$ , which highlights the crucial role of dust extinction in LyC escape. Conversely, the  $\beta$ -Metals model relies almost exclusively on dust attenuation to infer  $f_{\text{esc}}$  and lacks information on O32, Ly $\alpha$ , or morphology, which may better constrain the LyC absorption due to HI.

#### 3.2.1. Model Performance for $z \sim 3$ Samples

Comparing the models' performance for the high-redshift and combined low- and high-redshift samples is difficult, because each high-redshift model applies to a different set of galaxies. For example, the ELG-O32- $\beta$  model has the lowest rms for the high-redshift galaxies, but this rms is based on a single galaxy: Ion2. Ion2 is the only galaxy that is included in most of the models (all but ELG-O32- $\beta$ -Ly $\alpha$ , where it has an input limit). For Ion2, we adopt an observed absolute  $f_{\text{esc}} = 0.75$ , in the middle of the 0.5–1 bounds reported by Vanzella et al. (2016). Using this  $f_{\text{esc}}$ , Ion2's LyC escape is predicted most accurately by the ELG-O32- $\beta$  (rms = 0.02 dex) and ELG-



**Figure 1.** The  $f_{\text{esc}}$  predictions from the TopThree (a), LAE (b), LAE-O32 (c), and LAE-O32-nodust (d) models. See Table 1 for model descriptions. Red circles represent LzLCS+ LyC detections, and blue triangles represent upper limits. We plot  $z \sim 3$  galaxies in green, with stars denoting LyC detections and triangles denoting upper limits. We identify the two strongest high-redshift LCEs, Ion2 and J1316+2614, by a teal and a green diamond, respectively. Light green symbols indicate high-redshift galaxies that have a limit for one or more input variables. We draw circles around data points representing high-redshift galaxy stacks, which are less subject to uncertainty in IGM attenuation. The error bar in the upper left corner indicates the median size of the uncertainties in the observed and predicted  $f_{\text{esc}}$  for the combined sample of low- and high-redshift galaxies. The dashed line shows a one-to-one correspondence. Several of the Cox models predict  $f_{\text{esc}}$  for both low- and high-redshift galaxies with comparable accuracy.

O32 (rms = 0.04 dex) models. For Ion2, dust attenuation, luminosity, and ionization appear to be key factors in predicting its  $f_{\text{esc}}$ , although most other models are still consistent with its reported range of  $f_{\text{esc}}$  (rms = 0.09–0.15 dex for the next six models). The  $\beta$ -Metals model, which lacks information about O32 and morphology, is the only inaccurate model (rms = 0.43 dex). More surprisingly, however, the TopThree model gives the lowest predicted  $f_{\text{esc}} = 0.53$  (rms = 0.15 dex), which highlights the role of UV magnitude in the  $f_{\text{esc}}$  predictions, a point we discuss further in Section 4.

Nine galaxies, including four LCEs, appear in each of the LAE, LAE-O32, LAE-O32-nodust, and ELG-O32 model test samples. Although the LAE-O32-nodust model performed worst for the LzLCS+, for this common subsample of nine high-redshift galaxies, the LAE-O32-nodust model has the lowest rms (0.09 dex) and highest  $C$  (0.76), compared to rms values of 0.27–0.32 dex and  $C$  values of 0.65–0.71 for these galaxies in the other three models. Given the generally successful performance of all four of these models, models

with O32 and/or EW(Ly $\alpha$ ) appear able to predict  $f_{\text{esc}}$  for high-redshift galaxies equally well. Hence, while models with  $E(B - V)_{\text{UV}}$  and O32 or morphology work best for the LzLCS+ sample, we cannot rule out the possibility that EW(Ly $\alpha$ ) may also be important in predicting  $f_{\text{esc}}$  at  $z \sim 3$ .

Nevertheless, using EW(Ly $\alpha$ ) as the sole nebular measurement is not sufficient to accurately predict  $f_{\text{esc}}$ . The LAE model (Figure 1b), which applies to the largest set of  $z \sim 3$  LyC measurements, is one of the least successful models at predicting  $f_{\text{esc}}$  in the high-redshift galaxies. As seen in Table 3, this model has the largest rms scatter for the  $z \sim 3$  sample (0.9 dex), lowest  $C$  (0.49), and the second worst  $R^2$  (−4.2). The other model with comparably poor  $R^2$  and  $C$  for  $z \sim 3$  galaxies is the ELG-EW model, which is the only other model that exclusively relies on an emission-line EW for its input nebular information. Both models also have some of the lowest  $R^2$  values for the LzLCS+ sample, which reflects the fact that these models substantially underpredict the  $f_{\text{esc}}$  of the strongest LCEs in the LzLCS+. The LAE model shows this

same tendency toward underprediction for the  $z \sim 3$  LCEs as well; 13  $z \sim 3$  galaxies with observed  $f_{\text{esc}} \geq 0.2$  are predicted to have  $f_{\text{esc}}$  of only  $\leq 0.03$  by the model.

The poor performance of the LAE model may result from limitations in both the model itself and in the  $z \sim 3$  data. The fact that both the LAE and ELG-EW models fare poorly for the LzLCS+ and for the high-redshift galaxies, especially for strong LCEs, suggests that nebular EW is a flawed tracer of  $f_{\text{esc}}$ . Indeed, [O III] EW and Ly $\alpha$  EW should be low both for the weakest and for the strongest LCEs. In nonleakers, a high optical depth and corresponding lack of Ly $\alpha$  escape should lead to a low Ly $\alpha$  EW. Low [O III] and Ly $\alpha$  EWs could also indicate a weak current burst of star formation, without significant feedback or LyC production, which might result in undetected LyC. At lower optical depths and moderate  $f_{\text{esc}}$ , Ly $\alpha$  EW may increase due to enhanced escape, and [O III] EWs may likewise increase if these lower optical depths are preferentially associated with stronger starbursts. However, the [O III] and Ly $\alpha$  EWs will decrease again in the strongest LCEs, as a general lack of nebular gas results in the limited production of nebular emission lines (e.g., Zackrisson et al. 2013; Nakajima & Ouchi 2014). Thus, EW may be an ambiguous indicator of LyC escape.

The  $z \sim 3$  data also have limitations. The strong LCEs that are most severely underpredicted by the LAE model all come from the Liu et al. (2023) and Kerutt et al. (2024) samples of individual LAEs with LyC detections. Both papers note that their LyC detections represent the extreme of the population and may not be representative of the average galaxy population with these parameters, whose  $f_{\text{esc}}$  may be much lower. In Paper I, we found a similar result for the strongest LCEs in the LzLCS+, whose only distinguishing feature compared to more moderate LCEs is their low line-of-sight HI content, suggestive of a favorable orientation. Another limitation of the high-redshift data is the uncertainty in the IGM attenuation. At  $z \sim 3$ , this attenuation is significant and varies along different lines of sight (e.g., Rudie et al. 2013; Inoue et al. 2014; Vanzella et al. 2016; Steidel et al. 2018). This unknown sight line dependence leads to additional uncertainty in the reported  $f_{\text{esc}}$  for individual galaxies and could cause some genuine LCEs to appear as nondetections or to have overestimated  $f_{\text{esc}}$ .

The stacked samples (Steidel et al. 2018; Bian & Fan 2020; Nakajima et al. 2020), shown circled in Figure 1, average over these variations and are less subject to this uncertainty. In the LAE model in Figure 1(b), we see that the stacked samples do indeed show less scatter than the individual high-redshift detections. However, the model does not perfectly predict  $f_{\text{esc}}$  for the stacked samples, and their scatter is still comparable to the scatter in the individual LzLCS+ galaxies. In addition to averaging over IGM attenuation variations, the stacks also average over any galaxy-to-galaxy variation in physical properties. Hence, the input variables may not represent the true set of properties of an individual system, leading to some uncertainty in the predicted  $f_{\text{esc}}$ .

Unfortunately, the small sample sizes in most models and the lack of a common high-redshift sample across the models make it difficult to discern which model parameters are most important for predicting  $f_{\text{esc}}$  at high redshift. Our results suggest that if high-redshift galaxies behave like their lower-redshift counterparts in the LzLCS+ sample,  $E(B - V)_{\text{UV}}$  or  $\beta_{1550}$ , O32, and  $\Sigma_{\text{SFR}}$  or half-light radius are essential variables to include. For the existing  $z \sim 3$  samples, models with at least

two of these parameters do successfully predict the  $f_{\text{esc}}$  of the available high-redshift galaxies to within  $\sim 0.3$  dex.

This performance also shows that the LzLCS+ galaxies may indeed be reasonable analogs for high-redshift galaxies and that the same observable and physical properties may correlate with LyC escape at both low and high redshift (e.g., Schaerer et al. 2022; Mascia et al. 2023; Saldana-Lopez et al. 2023). This agreement is not trivial, as, in principle, the  $z \sim 0.3$  and  $z \sim 3$  galaxies could have different star formation histories, dust properties, morphologies, or other properties, any of which could affect  $f_{\text{esc}}$ . The LzLCS+ parameter space does cover the properties of the  $z \sim 3$  samples, with some exceptions. 17  $z \sim 3$  targets (eight LCEs and nine non-LCEs) are brighter than the LzLCS+ UV magnitude range by 0.1–3.18 mag, including the strong LCEs Ion2 and J1316+2614. An additional two LCEs in the LAE model are fainter than the LzLCS+ by 0.5–0.7 mag. The predictions for most models, all those with  $M_{1500}$  as a variable, therefore extrapolate to an unobserved part of parameter space, yet still perform well.

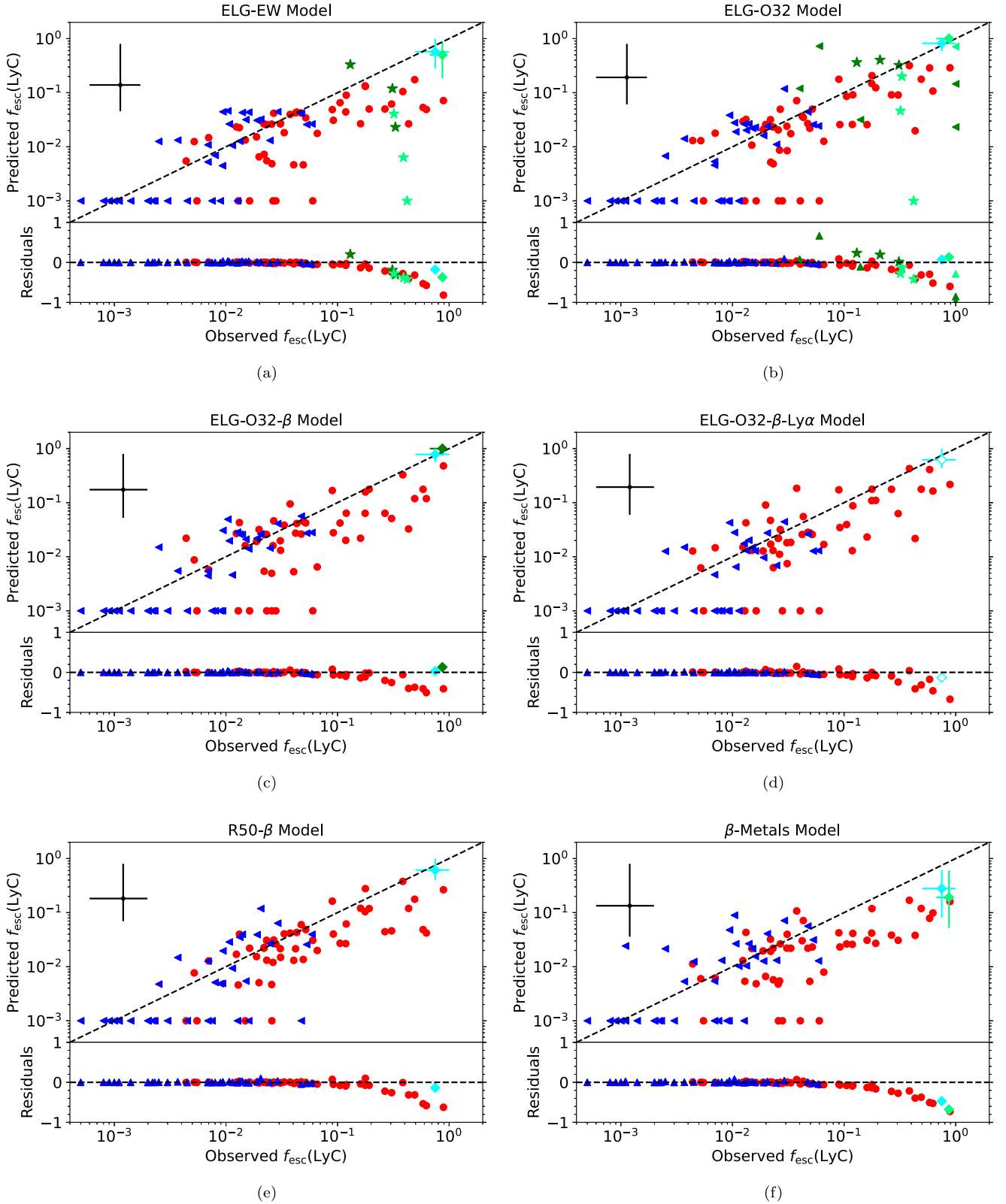
The data in Figures 1 and 2 and metrics in Table 3 include  $z \sim 3$  galaxies that have measured values for all of the input independent variables in the models.<sup>19</sup> However, additional galaxies in these samples have upper or lower limits for some independent variables, which can still provide potentially useful constraints on  $f_{\text{esc}}$ . For these galaxies, we adopt the limit as the value of the independent variable and predict their  $f_{\text{esc}}$  using the LzLCS+ models. We plot these approximate estimates in Figures 1 and 2 as light green symbols and an open symbol for Ion2, but we do not include these galaxies in the goodness-of-fit metrics.

The galaxies with limits have one or more of the following: upper limits in  $M_*$ , lower limits in  $\Sigma_{\text{SFR}}$ , lower limits in EW([O III]+H $\beta$ ), lower limits in O32, lower limits in EW(Ly $\alpha$ ), lower limits in  $M_{1500}$ , and lower limits in  $f_{\text{esc,Ly}\alpha}$ . One galaxy has an upper limit in EW([O III]+H $\beta$ ) rather than a lower limit. Given the coefficients for these variables in the models, the mass and magnitude limits would cause the model to overestimate  $f_{\text{esc}}$ , while the lower limits in the nebular emission lines should generally cause the model to underestimate  $f_{\text{esc}}$ . For example, a galaxy with a lower limit of EW(Ly $\alpha$ ) > 100Å is an even stronger Ly $\alpha$  emitter than we assume, and its predicted  $f_{\text{esc}}$  will be an underestimate.

Most of the  $z \sim 3$  LCEs with input limits in the LAE-O32 and LAE-O32-nodust models only have lower limits in O32, such that their predicted  $f_{\text{esc}}$  should be an underestimate. For galaxies with lower limits in both a nebular property and  $M_{1500}$ , we cannot easily interpret the predicted  $f_{\text{esc}}$  as a lower or upper limit. However, the best-fit coefficients in the Appendix show that the models have a steeper dependence on the nebular lines than they do on  $M_{1500}$ . Figures 1 and 2 show that, for most of the galaxies with input limits, the measured limits still constrain  $f_{\text{esc}}$  as accurately as for the rest of the sample. However, a few galaxies (90675, 101846, 105937 from Fletcher et al. 2019) have strongly underpredicted  $f_{\text{esc}}$ , which suggests that their nebular lines may be much stronger than the reported limit.

Finally, we note that our model predictions do not account for systematic uncertainties, including differences in methodology. We adopt the published values for all data. However, each

<sup>19</sup> The reported  $E(B - V)$  for Ion2 is a stringent upper limit of  $E(B - V) < 0.04$ . For consistency with Ion2's other reported measurements, we follow Vanzella et al. (2016) in adopting  $E(B - V) = 0.0$ , and we choose to include Ion2 in our goodness-of-fit calculations.



**Figure 2.** The  $f_{\text{esc}}$  predictions from the ELG-EW (a), ELG-O32 (b), ELG-O32- $\beta$  (c), ELG-O32- $\beta$ -Ly $\alpha$  (d), R50- $\beta$  (e), and  $\beta$ -Metals (f) models. See Table 1 for model descriptions. Ion2 is shown by an open symbol in panel (d), because its prediction is based on a limit in  $f_{\text{esc,Ly}\alpha}$ . Other symbols are the same as in Figure 1.

paper makes different assumptions, which could affect the tabulated values. For instance, papers differ in their adopted IGM transmission models, the models used in SED fits, and adopted dust attenuation laws. These assumptions could lead to systematic differences in properties like  $f_{\text{esc}}$ ,  $M_*$ , and

$E(B - V)_{\text{UV}}$  among the different samples. In addition, LyC measurements may be photometric or spectroscopic and may cover different wavelength ranges. Despite these systematics, the models work well in predicting  $f_{\text{esc}}$ . Adopting consistent methodologies could potentially result in better predictions;

**Table 4**  
Predictions for  $z \gtrsim 6$  Galaxies from the ELG-EW Model

Source ID	$z_{\text{phot}}$	$z_{\text{spec}}$	$f_{\text{esc}}$	$f_{\text{esc,min}}$	$f_{\text{esc,max}}$	References
COS-83688	6.70	...	0.190	0.056	0.602	Endsley et al. (2021)
COS-87259	6.66	...	0.024	0.005	0.119	Endsley et al. (2021)
COS-237729	6.87	...	0.119	0.033	0.488	Endsley et al. (2021)
COS-312533	6.85	...	0.043	0.016	0.265	Endsley et al. (2021)
COS-400019	6.88	...	0.109	0.031	0.463	Endsley et al. (2021)

**Note.** Predicted  $f_{\text{esc}}$  values from the ELG-EW model for the  $z \gtrsim 6$  galaxies from Endsley et al. (2021, 2023), Bouwens et al. (2023), Fujimoto et al. (2023), Tang et al. (2023), and Saxena et al. (2024).  $z_{\text{phot}}$  and  $z_{\text{spec}}$  are the photometric and spectroscopic redshifts.  $f_{\text{esc,min}}$  and  $f_{\text{esc,max}}$  represent the 15.9 and 84.1 percentiles of the model  $f_{\text{esc}}$  predictions. For galaxies with upper limits on EW([O III]+H $\beta$ ), the  $f_{\text{esc}}$  predictions are also upper limits and are marked accordingly. The reference column lists the publication used for the model input variables.

(This table is available in its entirety in machine-readable form in the [online article](#).)

**Table 5**  
Predictions for  $z \gtrsim 6$  Galaxies from the R50- $\beta$  Model

Source ID	$z_{\text{phot}}$	$z_{\text{spec}}$	$f_{\text{esc}}$	$f_{\text{esc,min}}$	$f_{\text{esc,max}}$	References
J1235-15534	5.9	...	0.09	0.03	0.41	Morishita et al. (2024)
JADESGDS-6734	...	5.92	0.11	0.03	0.44	Morishita et al. (2024)
JADESGDS-11449	...	5.94	0.51	0.20	0.88	Morishita et al. (2024)
JADESGDS-18169	...	5.94	1.00	0.61	1.00	Morishita et al. (2024)
JADESGDS-33803	...	5.97	0.18	0.05	0.56	Morishita et al. (2024)

**Note.** Predicted  $f_{\text{esc}}$  values from the R50- $\beta$  model.  $z_{\text{phot}}$  and  $z_{\text{spec}}$  are the photometric and spectroscopic redshifts.  $f_{\text{esc,min}}$  and  $f_{\text{esc,max}}$  represent the 15.9 and 84.1 percentiles of the model  $f_{\text{esc}}$  predictions. The reference column lists the publication used for the model input variables.

(This table is available in its entirety in machine-readable form in the [online article](#).)

however, as shown by the scatter in the LzLCS+ sample, which has a consistent methodology, the inherent uncertainty in the model itself also limits the possible accuracy of predictions.

### 3.2.2. Summary of $z \sim 3$ Results

In conclusion, Cox models derived using the  $z \sim 0.3$  LzLCS+ sample can successfully predict  $f_{\text{esc}}$  in  $z \sim 3$  LCEs. This agreement suggests that LCEs may have similar physical properties at both low and high redshift (e.g., Saxena et al. 2023). We find that the most accurate models include  $E(B - V)_{\text{UV}}$  and O32 or morphology measurements as variables. However, larger high-redshift samples with a full suite of measurements are required to test this result. Future observations of  $z \sim 3$  LCEs with JWST will further clarify whether the relationship between  $f_{\text{esc}}$  and physical properties evolves with redshift or remains constant. Preliminary JWST observations suggest that  $z \gtrsim 6$  galaxies may indeed share numerous physical properties with low-redshift analog samples like the LzLCS (e.g., Schaerer et al. 2022; Mascia et al. 2023; Endsley et al. 2023; Lin et al. 2024). The models developed on the LzLCS+ sample and tested on  $z \sim 3$  LCEs may therefore apply equally well at the reionization epoch.

## 4. $f_{\text{esc}}$ Predictions at $z \gtrsim 6$

### 4.1. Model Predictions for $z \gtrsim 6$ Samples

Given the success of our models in predicting  $f_{\text{esc}}$  at  $z \sim 3$ , we now use the Cox models to estimate  $f_{\text{esc}}$  for galaxies at  $z \gtrsim 6$ , where LyC is not detectable because of the IGM opacity. We consider several models: the ‘‘ELG’’ models, R50- $\beta$ , and  $\beta$ -Metals. The  $\beta$ -Metals model is the most limited, as it only includes measurements of luminosity, mass, UV slope, and metallicity and has no information on nebular line strength or

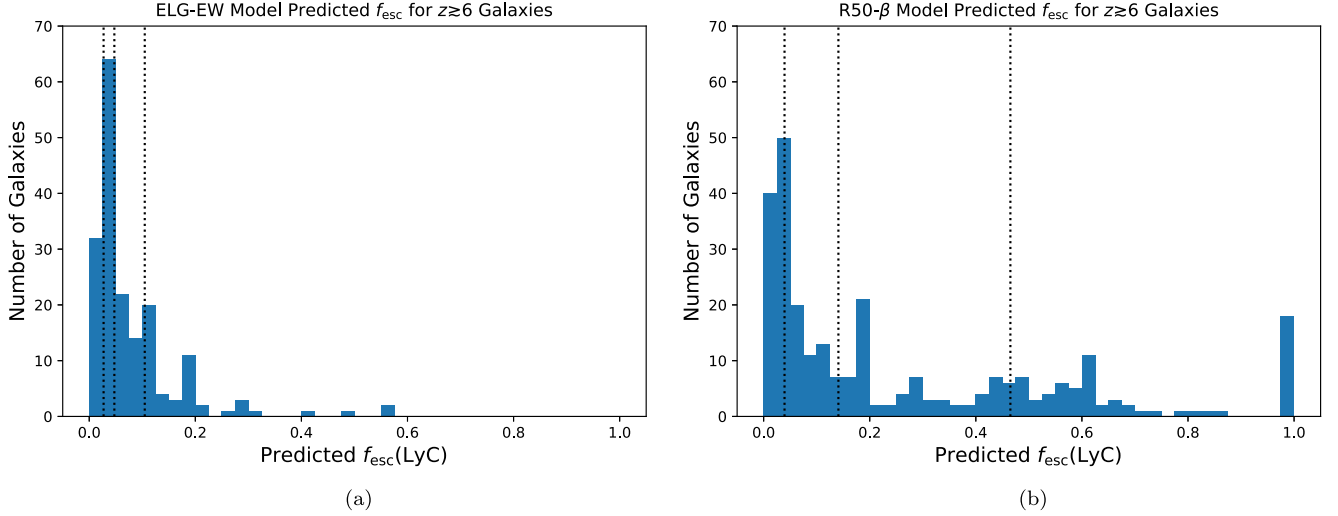
morphology. Based on the LzLCS+ galaxies, these models have an rms scatter of 0.44–0.55 dex in  $f_{\text{esc}}$  (see Section 3 and Table 3), with the ELG-O32 or R50- $\beta$  models giving the most accurate predictions, depending on the metric considered.

We can apply these models to several samples of reionization-era galaxies, which span  $z \sim 6$ –14: photometric samples from Endsley et al. (2021), Bouwens et al. (2023), Endsley et al. (2023), and Morishita et al. (2024) and spectroscopic samples from Schaerer et al. (2022), Mascia et al. (2023), Saxena et al. (2024, 2024), Fujimoto et al. (2023), Tang et al. (2023), Williams et al. (2023), and Atek et al. (2024). The Endsley et al. (2021) sample is based on Spitzer, HST, and ground-based observations; all other samples incorporate JWST NIRCcam photometry, and several (Schaerer et al. 2022; Mascia et al. 2023; Saxena et al. 2023, 2024; Atek et al. 2024; Fujimoto et al. 2023; Tang et al. 2023; Williams et al. 2023) use JWST NIRSpec spectroscopy as well. We select all galaxies at  $z > 5.9$  from these samples to investigate  $f_{\text{esc}}$  in the epoch of reionization.

#### 4.1.1. Photometric Samples

We list the predicted  $f_{\text{esc}}$  values and their associated uncertainties for the  $z \gtrsim 6$  galaxies in Tables 4–8. The ELG-EW and R50- $\beta$  models (Tables 4 and 5) apply to the largest compilations of  $z \gtrsim 6$  galaxies, with 183 measurements corresponding to 180 unique galaxies for the ELG-EW samples and 278 measurements at  $z > 5.9$  for the R50- $\beta$  samples, although most of the galaxies for both models only have photometric redshifts. Most of the galaxies in the ELG-EW samples are from Endsley et al. (2021, 2023), and most of the R50- $\beta$  sample galaxies come from Morishita et al. (2024).

We show histograms of the ELG-EW and R50- $\beta$  model  $f_{\text{esc}}$  predictions in Figure 3. The median of the  $f_{\text{esc}}$  predictions for



**Figure 3.** The distribution of predicted  $f_{\text{esc}}$  from the ELG-EW model (a) for  $z \gtrsim 6$  galaxies from Bouwens et al. (2023), Endsley et al. (2021, 2023), Fujimoto et al. (2023), Tang et al. (2023), and Saxena et al. (2023) and from the R50- $\beta$  model (b) for  $z \gtrsim 6$  galaxies from Morishita et al. (2024) and Mascia et al. (2023). The vertical dotted lines show the median and quartiles of the distributions ( $f_{\text{esc}} = 0.027, 0.047, 0.105$  for ELG-EW and  $f_{\text{esc}} = 0.039, 0.141, 0.467$  for R50- $\beta$ ). Both models have some limitations but suggest that many, but not all, high-redshift galaxies may have  $f_{\text{esc}} > 0.1$ .

**Table 6**  
Predictions for  $z \gtrsim 6$  Galaxies from the ELG-O32 Model

Source ID	Alternate Name	$z_{\text{spec}}$	$f_{\text{esc}}$	$f_{\text{esc,min}}$	$f_{\text{esc,max}}$	References
CEERS-2	CEERS1-3858	8.807	>0.178	>0.053	>0.567	Fujimoto et al. (2023)
CEERS-3	CEERS1-3908	8.005	>0.049	>0.023	>0.267	Fujimoto et al. (2023)
CEERS-3	CEERS1-3908	8.00	>0.006	>0	>0.032	Tang et al. (2023)
CEERS-7	CEERS1-6059	7.993	>0.119	>0.038	>0.453	Fujimoto et al. (2023)
CEERS-20	CEERS3-1748	7.769	>0	>0	>0	Fujimoto et al. (2023)
CEERS-23	CEERS6-7603	8.881	>0.008	>0	>0.041	Fujimoto et al. (2023)
CEERS-23	CEERS6-7603	8.881	>0.007	>0	>0.038	Tang et al. (2023)
CEERS-24	CEERS6-7641	8.998	>0.012	>0	>0.043	Fujimoto et al. (2023)
CEERS-24	CEERS6-7641	8.999	>0.013	>0	>0.049	Tang et al. (2023)
CEERS-44	...	7.10	>0.402	>0.159	>0.731	Tang et al. (2023)
CEERS-407	...	7.028	0.023	0.005	0.093	Tang et al. (2023)
CEERS-498	...	7.18	>0.499	>0.201	>0.869	Tang et al. (2023)
CEERS-499	...	7.168	>0	>0	>0	Tang et al. (2023)
CEERS-698	...	7.47	0.722	0.498	1	Tang et al. (2023)
CEERS-1019	...	8.678	1	0.644	1	Tang et al. (2023)
CEERS-1025	...	8.715	0.301	0.110	0.622	Tang et al. (2023)
CEERS-1027	...	7.819	>0.650	>0.453	1	Tang et al. (2023)
CEERS-1038	...	7.194	>0.007	>0	>0.038	Tang et al. (2023)
JADES-GS-z7-LA	...	7.278	0	0	0.012	Saxena et al. (2024)
11027	...	9.51	0.021	0.005	0.071	Williams et al. (2023)

**Note.** Predicted  $f_{\text{esc}}$  values from the ELG-O32 model.  $z_{\text{spec}}$  is the spectroscopic redshift.  $f_{\text{esc,min}}$  and  $f_{\text{esc,max}}$  represent the 15.9 and 84.1 percentiles of the model  $f_{\text{esc}}$  predictions. For galaxies with lower limits on O32, the  $f_{\text{esc}}$  predictions are also lower limits and are marked accordingly. The reference column lists the publication used for the model input variables.

the ELG-EW model samples is fairly low,  $f_{\text{esc}} = 0.047$ , which is lower than the average  $f_{\text{esc}}$  value of 0.1–0.2 required to reionize the Universe assuming a canonical ionizing photon production efficiency  $\log_{10}(\xi_{\text{ion}}) \sim 25.3$  (e.g., Finkelstein et al. 2015; Robertson et al. 2015; Finkelstein et al. 2019; Naidu et al. 2020). Of these  $z \gtrsim 6$  galaxies, 27% have  $f_{\text{esc}} \geq 0.1$ , and only 6% have  $f_{\text{esc}} \geq 0.2$ . However, as shown in Figure 2a, the ELG-EW model is one of the less accurate models, with a greater tendency to underpredict the true  $f_{\text{esc}}$  of the LzLCS+ sample. This underprediction implies that the ELG-EW variable set, which includes EW([O III]+H $\beta$ ) but not O32, does not distinguish strong LCEs from weaker LCEs. The more

accurate R50- $\beta$  model (see Table 3 and Figure 2) implies a higher fraction of LCEs, with a median value of  $f_{\text{esc}} = 0.14$ , and 56% and 39% of the galaxies with  $f_{\text{esc}} > 0.1$  and  $f_{\text{esc}} > 0.2$ , respectively. Both models find a substantial fraction of weak or nonleakers, and 25% of the galaxies in both models have  $f_{\text{esc}} < 0.03$ –0.04. Taken at face value, our preliminary results would suggest that  $f_{\text{esc}}$  values  $> 0.1$  are common but not ubiquitous among moderate to bright galaxies ( $M_{1500} < -18$ ) in the epoch of reionization (see also Mascia et al. 2023, 2024).

We caution that some of the  $z \gtrsim 6$  galaxies in the ELG-EW sample and many in the R50- $\beta$  sample fall outside of the parameter space probed by the LzLCS+, although a similar

**Table 7**  
Predictions for  $z \gtrsim 6$  Galaxies from the ELG-O32- $\beta$  Model

Source ID	$z_{\text{spec}}$	$f_{\text{esc}}$	$f_{\text{esc,min}}$	$f_{\text{esc,max}}$	References
21842	7.982	0.09	0.03	0.41	Saxena et al. (2024)
10013682	7.276	0.009	0.00	0.04	Saxena et al. (2024)
16625	6.631	0.37	0.12	0.68	Saxena et al. (2024)
18846	6.336	0.60	0.37	1.00	Saxena et al. (2024)
19342	5.974	0.47	0.18	0.82	Saxena et al. (2024)
9422	5.937	0.85	0.59	1.00	Saxena et al. (2024)
6002	5.937	0.11	0.03	0.44	Saxena et al. (2024)
12637	7.66	0.21	0.07	0.60	Saxena et al. (2024)
15362	6.794	0.00	0.00	0.03	Saxena et al. (2024)
13607	6.622	0.00	0.00	0.00	Saxena et al. (2024)
14123	6.327	0.07	0.03	0.34	Saxena et al. (2024)
58850	6.263	0.35	0.12	0.66	Saxena et al. (2024)
17138	6.204	0.00	0.00	0.02	Saxena et al. (2024)
9365	5.917	0.28	0.09	0.61	Saxena et al. (2024)
11027	9.51	0.01	0.00	0.05	Williams et al. (2023)
4590	8.495	0.01	0.00	0.05	Schaerer et al. (2022)
6355	7.664	0.06	0.03	0.35	Schaerer et al. (2022)
10612	7.66	0.17	0.05	0.53	Schaerer et al. (2022)
JADES-GS-z7-LA	7.278	0.00	0.00	0.02	Saxena et al. (2023)
10025	7.875	0.01	0.00	0.06	Mascia et al. (2023)
100004	7.884	>0.01	>0.00	>0.05	Mascia et al. (2023)
10000	7.884	0.05	0.02	0.26	Mascia et al. (2023)
10021	7.288	0.42	0.17	0.76	Mascia et al. (2023)
100001	7.875	0.006	0.00	0.04	Mascia et al. (2023)
100003	7.88	0.47	0.18	0.82	Mascia et al. (2023)
100005	7.883	0.02	0.005	0.09	Mascia et al. (2023)
150008	6.23	>0.12	>0.04	>0.47	Mascia et al. (2023)

**Note.** Predicted  $f_{\text{esc}}$  values from the ELG-O32- $\beta$  model.  $z_{\text{spec}}$  is the spectroscopic redshift.  $f_{\text{esc,min}}$  and  $f_{\text{esc,max}}$  represent the 15.9 and 84.1 percentiles of the model  $f_{\text{esc}}$  predictions. For galaxies with lower limits on O32, the  $f_{\text{esc}}$  predictions are also lower limits and are marked accordingly. The reference column lists the publication used for the model input variables.

**Table 8**  
Predictions for  $z \gtrsim 6$  Galaxies from the  $\beta$ -Metals Model

Source ID	$z_{\text{spec}}$	$f_{\text{esc}}$	$f_{\text{esc,min}}$	$f_{\text{esc,max}}$	References
18924	7.7	0.05	0.02	0.30	Atek et al. (2024)
16155	6.87	0.03	0.006	0.18	Atek et al. (2024)
23920	6.00	0.10	0.03	0.46	Atek et al. (2024)
12899	6.88	0.17	0.04	0.55	Atek et al. (2024)
8613	6.38	0.18	0.04	0.57	Atek et al. (2024)
23619	6.72	0.08	0.03	0.42	Atek et al. (2024)
38335	6.23	0.02	0.00	0.11	Atek et al. (2024)
27335	6.76	0.08	0.03	0.42	Atek et al. (2024)
11027	9.51	0.03	0.005	0.15	Williams et al. (2023)
4590	8.495	0.05	0.02	0.35	Schaerer et al. (2022)
6355	7.664	0.03	0.006	0.18	Schaerer et al. (2022)
10612	7.66	0.05	0.02	0.36	Schaerer et al. (2022)

**Note.** Predicted  $f_{\text{esc}}$  values from the  $\beta$ -Metals model.  $z_{\text{spec}}$  is the spectroscopic redshift.  $f_{\text{esc,min}}$  and  $f_{\text{esc,max}}$  represent the 15.9 and 84.1 percentiles of the model  $f_{\text{esc}}$  predictions. The reference column lists the publication used for the model input variables.

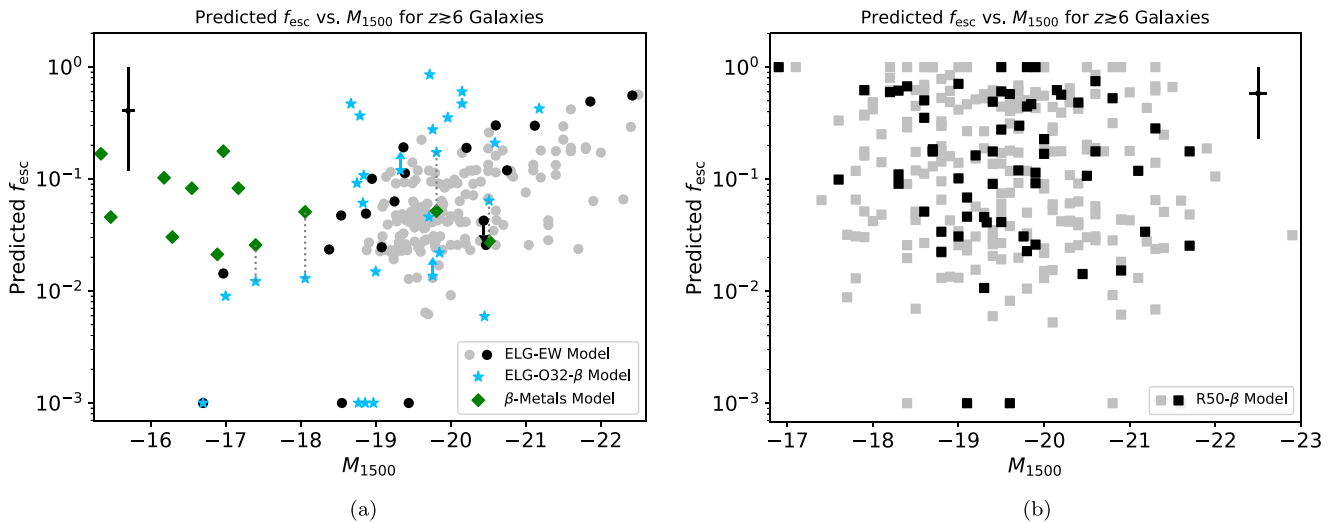
extrapolation did not seem to adversely affect the  $f_{\text{esc}}$  predictions for the  $z \sim 3$  galaxies (Section 3). Of the 183 measurements in the ELG-EW model sample, 14 galaxies have brighter UV magnitudes than the LzLCS+ galaxies, one is fainter, and 17 have higher EW([O III]+H $\beta$ ). Likewise, the R50- $\beta$  sample contains eight galaxies brighter than and 30 galaxies fainter than the LzLCS+ sample. More concerningly, 114 of the 278 galaxies in the R50- $\beta$  compilation are more

compact than the LzLCS+ galaxies, which could mean that their high inferred  $f_{\text{esc}}$  values result from an incorrect extrapolation into this compact regime.

#### 4.1.2. Spectroscopic Samples

The ELG-O32 and ELG-O32- $\beta$  model predictions (Tables 6 and 7) apply to fewer  $z \gtrsim 6$  galaxies, only 17 and 27 galaxies, respectively, but all of these galaxies have spectroscopic redshifts. Half the galaxies in the ELG-O32 model have predicted  $f_{\text{esc}} < 0.05$ , although most of these are only lower limits in  $f_{\text{esc}}$ , and 25% of the ELG-O32- $\beta$  sample galaxies have  $f_{\text{esc}} \leq 0.01$ . Like the R50- $\beta$  model predictions, low  $f_{\text{esc}}$  is common, but the overall distribution also extends to very high  $f_{\text{esc}}$ . Both models include some galaxies with lower limits in O32, whose predicted  $f_{\text{esc}}$  values therefore also correspond to lower limits. Depending on how high the true  $f_{\text{esc}}$  for these galaxies are, 35%–82% of the ELG-O32 model galaxies and 33%–41% of the ELG-O32- $\beta$  galaxies have  $f_{\text{esc}} \geq 0.2$ .

Because these galaxies are all spectroscopically confirmed, these samples could be biased toward galaxies with stronger emission lines and hence higher  $f_{\text{esc}}$ ; the median EW([O III]+H $\beta$ ) of the spectroscopic ELG-O32 sample is 1790 Å (Saxena et al. 2023; Fujimoto et al. 2023; Tang et al. 2023), compared with 690 Å for the photometric samples in the ELG-EW model, for example (Endsley et al. 2021; Bouwens et al. 2023; Endsley et al. 2023). In addition, these models may not underpredict  $f_{\text{esc}}$  for strong LCEs to the same extent as the ELG-EW model (see Figure 2 and Table 3). For the six galaxies with predicted  $f_{\text{esc}}$



**Figure 4.** Predicted  $f_{\text{esc}}$  as a function of  $M_{1500}$  for  $z \gtrsim 6$  galaxies. Panel (a) shows the predicted  $f_{\text{esc}}$  from the ELG-EW model (circles) for galaxies with photometric redshifts (gray) and spectroscopic redshifts (black), from the ELG-O32- $\beta$  model (blue stars), and from the  $\beta$ -Metals model (green diamonds). Panel (b) shows the predicted  $f_{\text{esc}}$  from the R50- $\beta$  model (squares) with photometric (gray) and spectroscopic (black) redshifts. The crosses in the upper corners show the median uncertainties for each panel. Dark gray dotted lines connect the predictions for the same galaxy in different models. We see no strong trend between  $f_{\text{esc}}$  and  $M_{1500}$ , but many of the predictions suffer from a limited set of variables (ELG-EW and  $\beta$ -Metals models) or require extrapolation outside the LzLCS+ parameter space (R50- $\beta$  model).

$> 0.2$  in the ELG-O32 model, the ELG-EW predictions are indeed lower by  $\Delta f_{\text{esc}} = 0.26$  on average, which suggests that model differences also account for some of the higher  $f_{\text{esc}}$  values compared to the ELG-EW distribution.

While we have generated initial predictions for  $z \gtrsim 6$  galaxies using our models, predictions for the largest samples are limited by a lack of spectroscopic information. We do find evidence of high  $f_{\text{esc}} > 0.2$  among smaller spectroscopic samples at  $z \gtrsim 6$ , but these samples may be biased toward stronger emission-line galaxies. To obtain more accurate predictions, high-redshift studies should prioritize observations of nebular line ratios and galaxy sizes for larger samples. As JWST surveys continue, larger, more representative spectroscopic samples will improve estimates of the average  $f_{\text{esc}}$  at  $z > 6$ .

#### 4.2. Trends with UV Magnitude

We examine the predicted  $f_{\text{esc}}$  values as a function of magnitude in Figure 4. We plot predictions from the three models that apply to the largest sample sizes (ELG-EW, ELG-O32- $\beta$ , and R50- $\beta$ ) and from the model that applies to the faintest galaxies ( $\beta$ -Metals; Table 8). We see no strong trend with magnitude, although the few brightest sources with  $M_{1500} \lesssim -21.5$  do appear to have higher  $f_{\text{esc}}$  on average, with a median  $f_{\text{esc}} = 0.18$  versus 0.07 for fainter sources. Aside from these rare sources, the ELG-EW, ELG-O32- $\beta$ , and  $\beta$ -Metals samples tend to show  $f_{\text{esc}}$  within  $\sim 0.02$ – $0.3$  across a wide range of magnitudes, from  $-16$  to  $-22$ . The R50- $\beta$  model (Figure 4(b)) also shows a flat distribution of  $f_{\text{esc}}$  with magnitude, but it predicts many more galaxies at higher  $f_{\text{esc}}$ . However, as previously noted, 41% of the galaxies in this model require extrapolating predictions to galaxies with extremely compact morphologies, which makes these predictions uncertain.

Any trends with magnitude or lack thereof in the other models should also be taken with caution. All the models have a tendency to underpredict  $f_{\text{esc}}$ , but this tendency is most pronounced in the ELG-EW and  $\beta$ -Metals models (Figure 2),

which trace the extremes of the magnitude range in Figure 4. The more reliable ELG-O32- $\beta$  model, which better reproduces the LzLCS+  $f_{\text{esc}}$  values, suggests that some of the galaxies in the  $M_{1500} = -19$  to  $-21$  mag range may have significantly high  $f_{\text{esc}}$ . The ELG-O32- $\beta$  sample does not include enough faint galaxies to reveal whether or not a trend between  $f_{\text{esc}}$  and magnitude exists, however.

At the faintest magnitudes, the Atek et al. (2024) sample, included in the  $\beta$ -Metals model, includes several extremely low-luminosity lensed galaxies observed as part of the JWST UNCOVER survey (Bezanson et al. 2022; Weaver et al. 2024). For these galaxies, the  $\beta$ -Metals model therefore extrapolates the LzLCS+ sample to an unobserved parameter space. The only statistically significant predictor variable in the  $\beta$ -Metals model is  $\beta_{1550}$ , and Figure 4 illustrates that these galaxies' blue slopes suggest at least moderate  $f_{\text{esc}}$  (0.02–0.18; Table 8). Crucially, even these moderate  $f_{\text{esc}}$  values, which may be underpredicted, are still greater than what is required for such faint galaxies to dominate the reionization given their high LyC photon production rates (Atek et al. 2024).

While moderate or high  $f_{\text{esc}}$  for faint, blue galaxies matches past predictions (e.g., Chisholm et al. 2022), the high  $f_{\text{esc}}$  in the brighter sources in Figure 4 may seem surprising at first glance. In all four models, the fitted coefficients (see the Appendix) show that a brighter  $M_{1500}$  results in a higher  $f_{\text{esc}}$  at fixed values of the other inputs (fixed  $E(B - V)$ , fixed O32, etc.). This effect does not come from the association of higher observed UV luminosity with lower dust content. The data do not show such a correlation, and, because the models already include dust measurements, the model implies that brighter UV luminosities increase  $f_{\text{esc}}$  at fixed dust attenuation. In addition, we see a similar relationship with  $M_*$ , where higher  $M_*$  is also associated with higher  $f_{\text{esc}}$  at fixed values of the other parameters. Lin et al. (2024) find a similar result using a different technique. Using the LzLCS+ data, they fit a logistic regression model for the probability of having LyC escape as a function of  $M_{1500}$ ,  $\beta_{1550}$ , and O32 and likewise find a brighter  $M_{1500}$  increases the probability of LyC escape for fixed values



of the other variables. This scaling does not imply that higher UV luminosities correlate with  $f_{\text{esc}}$  in the LzLCS+ sample; if anything, the LzLCS+ indicates a slight trend in the opposite direction (Flury et al. 2022b). Instead, this scaling implies that a bright galaxy with LCE-like properties, such as high O32 and blue UV slope, is a more extreme object than a faint galaxy with identical properties. For instance, at fixed O32, the LzLCS+ galaxies with higher luminosities tend to have higher  $f_{\text{esc}}$ , but high O32 values are also uncommon among the bright galaxies as a whole, leading to an overall trend of decreasing  $f_{\text{esc}}$  with luminosity (e.g., Flury et al. 2022b).

The enhanced  $f_{\text{esc}}$  for bright LCE candidates affects the predictions in Figure 4 by boosting the  $f_{\text{esc}}$  of a bright, dust-poor emission-line galaxy or compact galaxy relative to an otherwise similar faint galaxy. For instance, for the ELG-EW model, the brightest galaxies should have higher predicted  $f_{\text{esc}}$  only if they have equally low  $E(B - V)$  and equally high EW as the fainter galaxies. Indeed, the ELG-EW samples, mostly galaxies from Endsley et al. (2021, 2023), show no trend between  $E(B - V)_{\text{UV}}$  and  $M_{1500}$  or between  $\text{EW}([\text{O III}] + \text{H}\beta)$  and  $M_{1500}$ ; with comparable dust and emission-line strengths, the brighter galaxies therefore end up with higher predicted  $f_{\text{esc}}$ . Endsley et al. (2023) note that lower metallicities or higher  $f_{\text{esc}}$  among the fainter galaxies could suppress their EWs and account for the lack of an observed trend with  $M_{1500}$ . Hence, other properties, such as O32, might be necessary to distinguish faint galaxies with high  $f_{\text{esc}}$ . The EWs in the Endsley et al. (2021, 2023) samples also come from photometry, which may be more uncertain than spectroscopic measurements (e.g., Duan et al. 2024).

The ELG-O32- $\beta$  samples have O32 measurements, but they similarly show no strong trend between  $\beta_{1550}$  and  $M_{1500}$  or between O32 and  $M_{1500}$ . Likewise, the Morishita et al. (2024) sample, which constitutes the bulk of the R50- $\beta$  model sample, exhibits a relatively flat trend between radius and  $M_{1500}$ . Only the  $\beta$ -Metals samples, mostly from Atek et al. (2024), show any trend between physical properties and  $M_{1500}$ , with the lowest metallicities and bluest slopes appearing among the faintest galaxies.

The limitations of the models and the available  $z \gtrsim 6$  data sets make it difficult to discern any trends between  $M_{1500}$  and  $f_{\text{esc}}$ . Limited measurements for galaxies at the brightest and faintest magnitudes restrict us to some of the less accurate Cox models. The models also require extrapolating to make predictions for galaxies outside the magnitude or radius range of the LzLCS+ sample. Better constraints on  $f_{\text{esc}}$  within the  $z > 6$  population will require larger spectroscopic samples across a wide range of galaxy magnitudes as well as better observational constraints on  $f_{\text{esc}}$  in a more diverse set of galaxies at lower redshifts.

### 4.3. Notes on Individual Sources

#### 4.3.1. Strong LCEs

Several galaxies in the ELG-O32 and ELG-O32- $\beta$  models appear to be extremely strong LCEs (Tables 6 and 7). One unusual galaxy, CEERS-1019, has a predicted  $f_{\text{esc}} = 1$ . Such a high  $f_{\text{esc}}$  would seemingly conflict with the presence of strong emission lines, since it should correspond to a complete absence of absorbing gas. However, given the uncertainty in the model predictions, the  $f_{\text{esc}}$  of CEERS-1019 could be as low as 0.6. Two other galaxies in the ELG-O32 model sample,

CEERS-698 and CEERS-1027, have  $f_{\text{esc}} \geq 0.5$ , and  $f_{\text{esc}}$  as high as 1 is consistent with their uncertainties. Interestingly, all three of these strong LCE candidates are strong Ly $\alpha$  emitters, which suggests that they reside within ionized bubbles (Tang et al. 2023). Similarly, two galaxies in the ELG-O32- $\beta$  samples have predicted  $f_{\text{esc}} > 0.5$ : 18846 and 9422 from JADES (Saxena et al. 2024). With O32 = 70.6, the ELG-O32- $\beta$  model infers  $f_{\text{esc}} = 0.85$  for 9422, much higher than the  $f_{\text{esc}} = 0.01$  value derived by Saxena et al. (2024) using the multivariate prediction method of Choustikov et al. (2024); we discuss the differences between our predictions and those of other models in Section 5.

The true  $f_{\text{esc}}$  of some of these extreme galaxies may not be quite as high as the models predict, however. Because two of these strong LCEs, CEERS-1019 ( $f_{\text{esc}} = 1$ ) and CEERS-698 ( $f_{\text{esc}} \sim 0.7$ ), have brighter UV magnitudes than the LzLCS+ sample by 0.4–0.9, their high predicted  $f_{\text{esc}}$  could be unreliable. However, these same models successfully predicted the high  $f_{\text{esc}}$  of the  $z \sim 3$  galaxies Ion2 and J1316+2614 (Figure 2), which are as bright as these  $z > 6$  galaxies. The LzLCS+ sample itself also contains only three LCEs with  $f_{\text{esc}} > 0.5$ , which limits the models’ ability to accurately determine predictive relationships in the high- $f_{\text{esc}}$  regime (Mascia et al. 2023). CEERS-1019 may differ from the LzLCS+ galaxies for additional reasons; it has unusually strong nitrogen emission and is a possible candidate for supermassive star formation (Marques-Chaves et al. 2024) or an AGN (Larson et al. 2023). The physical conditions occurring in extreme galaxies like CEERS-1019 likely are not present within the LzLCS+ sample. Empirical predictions based on  $z = 0.3$  galaxies would not be suitable for such objects.

The predictions for the other strong LCE candidates are likely more reliable, provided their star formation and nebular conditions are not significantly different from the LCEs at  $z \sim 3$ . Galaxy 9422 has a higher O32 ratio than the LzLCS+, but the other strong leaker candidates fall within the LzLCS+ parameter space, so that the model is not extrapolating to an unobserved regime. Hence, our models suggest that several strong emission-line galaxies identified in the epoch of reionization may be strong LCEs with  $f_{\text{esc}}$  well above 0.2.

#### 4.3.2. JADES-GS-z7-LA

Saxena et al. (2024) report the detection of a  $z = 7.3$  galaxy with strong Ly $\alpha$  emission, JADES-GS-z7-LA, which is presumably located within an ionized bubble. With  $f_{\text{esc,Ly}\alpha} = 0.96$ , IGM absorption seems to have had little effect on the galaxy’s Ly $\alpha$  emission. We therefore choose to incorporate the measured  $\text{EW}(\text{Ly}\alpha)$  and  $f_{\text{esc,Ly}\alpha}$  in our model predictions. We can apply most of the models in Table 1 to JADES-GS-z7-LA, omitting only the TopThree, R50- $\beta$ , and  $\beta$ -Metals models due to a lack of reported  $\Sigma_{\text{SFR}}$ , radius, and  $12 + \log_{10}(\text{O}/\text{H})$ . We compare the different  $f_{\text{esc}}$  predictions in Table 9. JADES-GS-z7-LA does fall outside of the LzLCS+ parameter space, which may limit the accuracy of the  $f_{\text{esc}}$  predictions; it is fainter than the LzLCS+ galaxies by 1.6 mag, and it has a higher  $\text{EW}(\text{Ly}\alpha)$  by 144.1 Å.

The different Cox models disagree regarding the  $f_{\text{esc}}$  of JADES-GS-z7-LA, with predicted  $f_{\text{esc}}$  ranging from 0 to 0.40. The only nonzero predictions come from models that have information about UV dust attenuation and Ly $\alpha$ , but even the extreme Ly $\alpha$  properties of JADES-GS-z7-LA do not guarantee  $f_{\text{esc}} > 0.1$ . The LAE-O32 and ELG-O32- $\beta$ -Ly $\alpha$  models include

**Table 9**  
 $f_{\text{esc}}$  Predictions for JADES-GS-z7-LA

Model	$f_{\text{esc}}$	$f_{\text{esc,min}}$	$f_{\text{esc,max}}$
LAE	0.400	0.142	0.736
LAE-O32	0.085	0.028	0.325
LAE-O32-nodust	0	0	0.028
ELG-EW	0	0	0.026
ELG-O32	0	0	0.012
ELG-O32- $\beta$	0	0	0.022
ELG-O32- $\beta$ -Ly $\alpha$	0.017	0	0.059

**Note.** Predicted  $f_{\text{esc}}$  values for JADES-GS-z7-LA from different models.  $f_{\text{esc,min}}$  and  $f_{\text{esc,max}}$  represent the 15.9 and 84.1 percentiles of the model  $f_{\text{esc}}$  predictions.

O32 and stellar mass as well as Ly $\alpha$  and find  $f_{\text{esc}} = 0.017\text{--}0.085$ . These two models are also among the best-performing models for the LzLCS+ sample (Table 3). JADES-GS-z7-LA is not devoid of dust, with  $E(B - V)_{\text{UV}} = 0.10$ , and its UV slope  $\beta = -2.1$  is not extreme. These properties are consistent with the properties of  $z \sim 0.3$  galaxies with moderate  $f_{\text{esc}}$ ; all the LzLCS+ galaxies with higher  $f_{\text{esc}} > 0.1$  have lower  $E(B - V)_{\text{UV}} < 0.1$ , and a slope of  $\beta_{1550} = -2.1$  matches the median value for the moderate LCEs with  $f_{\text{esc}} = 0.05 - 0.1$ .

Furthermore, as discussed above, given JADES-GS-z7-LA's low mass ( $\log_{10}(M_*/M_{\odot}) = 7.15$ ), its O32 value of 8.8 may not be extreme. Even its strong Ly $\alpha$  emission does not necessarily imply a low optical depth along the line of sight, since Ly $\alpha$  photons can scatter. For instance, the nonleaking  $z \sim 0.3$  galaxy J1248+4259 (Izotov et al. 2018b) has  $f_{\text{esc}} \leq 0.013$  and  $\text{EW}(\text{Ly}\alpha) = 256 \text{ \AA}$ . One piece of evidence that favors a high  $f_{\text{esc}}$  for JADES-GS-z7-LA is the small offset between the Ly $\alpha$  velocity and the systemic redshift (Saxena et al. 2023). With the low resolution of most of the LzLCS+ UV spectra, we cannot currently include this parameter in our models, and it may boost the predicted  $f_{\text{esc}}$  for JADES-GS-z7-LA. However, this low velocity offset could still be consistent with little to no LyC escape; its offset of  $120 \pm 80 \text{ km s}^{-1}$  also resembles the Ly $\alpha$  profile of the nonleaker J1248+4259, whose red peak is offset from the systemic velocity by  $< 200 \text{ km s}^{-1}$ .

In agreement with the assessment of Saxena et al. (2024), we thus find that the evidence for LyC escape in JADES-GS-z7-LA is ambiguous. The multivariate models that incorporate the most information suggest that it likely has a moderate  $f_{\text{esc}} \sim 0.017\text{--}0.085$  and may therefore not be solely responsible for producing its ionized bubble in the IGM (e.g., Witstok et al. 2024).

#### 4.3.3. Galaxies with High O32

The case of JADES-GS-z7-LA highlights the fact that the Cox multivariate predictions can differ from single-variable estimates (e.g., based on Ly $\alpha$  alone). Galaxies with high O32 provide another example of this result. For instance, Williams et al. (2023) present JWST observations of a lensed low-mass ( $\log_{10}(M_*/M_{\odot}) = 7.7$ ) galaxy, 11027, at  $z = 9.51$  and hypothesize that the galaxy has a high  $f_{\text{esc}} > 0.1$  based on its high O32 ratio of 12 ( $[\text{O III}]\lambda\lambda 5007+4959/[\text{O II}] = 16$ ). However, Lin et al. (2024) suggest that 11027 is not likely to be a strong LCE considering the combination of O32, UV magnitude, and UV slope. Our ELG-O32 and ELG-O32- $\beta$  Cox models (Tables 6–7) likewise predict  $f_{\text{esc}}$  of only 0.012–0.021. O32

alone is not sufficient to constrain  $f_{\text{esc}}$ , and high O32 values are common among the lowest-mass galaxies in the LzLCS+ sample (Flury et al. 2022b), including weak LCEs. In contrast, the ELG-O32- $\beta$  model predicts that several galaxies with lower O32 than galaxy 11027 (e.g., 6355, 10000, 10612, 12637 with  $\text{O32} = 6.3\text{--}10.6$ , Schaerer et al. 2022; Mascia et al. 2023; Saxena et al. 2024) are actually more likely to be LCEs because of their higher luminosities and/or bluer UV slopes (Table 7). The Cox predictions emphasize the fact that not all galaxies with high O32 have high  $f_{\text{esc}}$  (e.g., Izotov et al. 2018b; Flury et al. 2022b), and additional properties such as mass, UV luminosity, and dust extinction are important to consider.

Multivariate models are an important tool for predicting  $f_{\text{esc}}$  in the epoch of reionization. Our results highlight the fact that multivariate models can give substantially different predictions from single-variable estimates, and  $f_{\text{esc}}$  predictions should therefore incorporate as much information as possible.

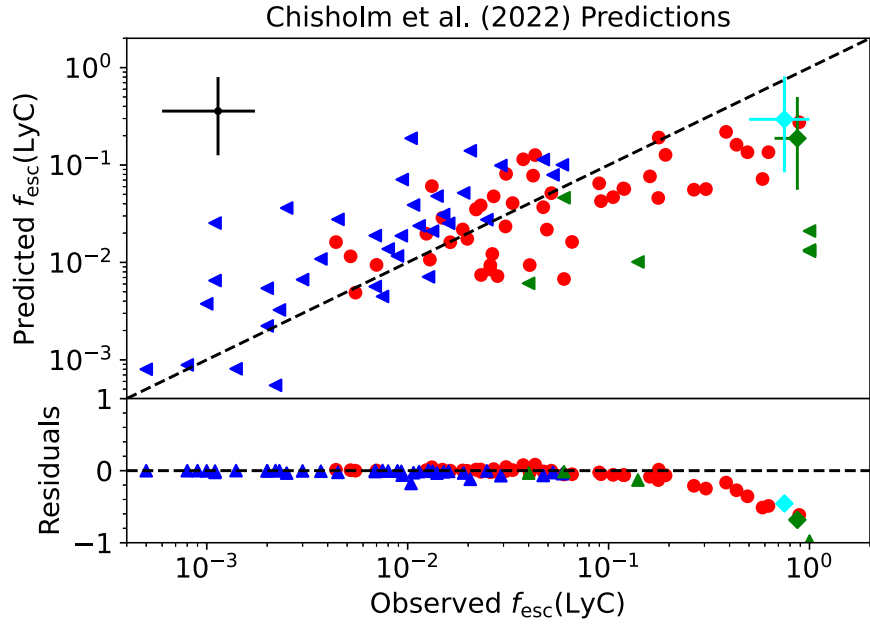
## 5. Comparison with Other Studies

### 5.1. Single-variable Predictions

Among the observable properties measured for the LzLCS+ sample, the UV slope  $\beta_{1550}$  shows one of the strongest correlations with  $f_{\text{esc}}$ . Consequently, Chisholm et al. (2022) propose that  $\beta_{1550}$  could predict  $f_{\text{esc}}$  for high-redshift galaxies and derive a relationship between  $\beta_{1550}$  and  $f_{\text{esc}}$  based on the LzLCS+ data set. In Figure 5, we compare the predicted  $f_{\text{esc}}$  from this method with the observed  $f_{\text{esc}}$  for the LzLCS+ and for  $z \sim 3$  galaxies (Vanzella et al. 2016; Bassett et al. 2019; Marques-Chaves et al. 2022). We list the goodness-of-fit statistics for the LzLCS+ with the Chisholm et al. (2022) model in Table 10. The Chisholm et al. (2022) model's high  $R^2 = 0.45$  and low rms = 0.43 are comparable to the best-performing Cox models in Tables 3 and 10, which demonstrates that  $\beta_{1550}$  alone can indeed predict  $f_{\text{esc}}$  for LCEs reasonably well. Consistent with this result, Paper I finds that, of the variables accessible at high redshift,  $\beta_{1550}$  is the most important variable to include in the Cox models.

Despite its success for LCEs, however, the Chisholm et al. (2022) model has more difficulty in constraining  $f_{\text{esc}}$  in the LzLCS+ nondetections, as indicated by its lower concordance  $C = 0.76$ . This concordance is lower than all the Cox models in Tables 3 and 10, which have  $C = 0.77 - 0.83$ . The difference between the predictions for detections and nondetections is apparent in Figure 5. For LzLCS+ galaxies with  $f_{\text{esc}}$  detections or upper limits between 0.01 and 0.1, the one-to-one relation between the Chisholm et al. (2022) model predictions and the observations runs right between the LzLCS+ detections. However, nearly all the nondetections are above this line, indicating that the model is systematically overpredicting their  $f_{\text{esc}}$ . In contrast, for one of the better performing Cox models, such as the ELG-O32 model in Figure 2(b), both detections and nondetections in this same  $f_{\text{esc}}$  range fall on either side of the one-to-one relation, indicating that predictions for detections and nondetections are comparable.

Figure 5 also applies the Chisholm et al. (2022) model to two  $z \sim 3$  LCEs: Ion2 and J1316+2614. The model correctly identifies both galaxies as LCEs but underpredicts their  $f_{\text{esc}}$ , although Ion2's observed  $f_{\text{esc}}$  is consistent with the stated uncertainties in the Chisholm et al. (2022) predictions. Comparing the Ion2 predictions with those from the Cox models discussed in Section 3, we find that the Chisholm et al. (2022) single-variable



**Figure 5.** The  $f_{\text{esc}}$  predictions for the LzLCS+ and  $z \sim 3$  galaxies from the Chisholm et al. (2022) model, which uses a single predictor variable:  $\beta_{1550}$ . Symbols are the same as in Figure 1. The model reproduces the  $f_{\text{esc}}$  of LCEs but overpredicts the  $f_{\text{esc}}$  for the LzLCS+ nondetections and underpredicts  $f_{\text{esc}}$  for the strongest LCEs.

**Table 10**  
Goodness-of-fit Statistics for Literature versus Cox Models

Model	$R^2$	$R_{\text{adj}}^2$	rms	$C$
Chisholm et al. (2022) Model	0.45	0.43	0.43	0.76
Choustikov et al. (2024) Model	-1.10	-1.40	0.83	0.57
Mascia et al. (2023) Model	0.57	0.54	0.38	0.81
JWST Cox Model (Paper I)	0.29	0.14	0.47	0.83
Cox Model with $\beta_{1550}$ , $\Sigma_{\text{SFR}}$ , O32	0.34	0.29	0.46	0.83
Cox Model with Choustikov et al. (2024) Variables	0.28	0.16	0.48	0.82
Cox Model with Mascia et al. (2023) Variables	0.40	0.35	0.44	0.82

**Note.** A comparison of the goodness-of-fit statistics for  $f_{\text{esc}}$  predictions for the LzLCS+ sample from the Chisholm et al. (2022), Choustikov et al. (2024), and Mascia et al. (2023) literature models and various Cox models. We list statistics for the JWST model from Paper I and a model limited to its three top-ranked variables. We also list statistics for Cox models run using the same sets of variables as in the Choustikov et al. (2024) and Mascia et al. (2023) models.

$\beta_{1550}$  model is one of the least accurate at predicting Ion2’s  $f_{\text{esc}}$ , with a predicted  $f_{\text{esc}} = 0.29$ , corresponding to rms = 0.40 dex. As discussed in Chisholm et al. (2022),  $\beta_{1550}$  only tracks the loss of LyC photons due to dust. The scatter below and above the Chisholm et al. (2022)  $\beta_{1550}$ - $f_{\text{esc}}$  relation results from variations in the absorbing HI column. Other properties, such as O32,  $\Sigma_{\text{SFR}}$ , and UV magnitude, may better track the HI component of LyC absorption for galaxies like Ion2.

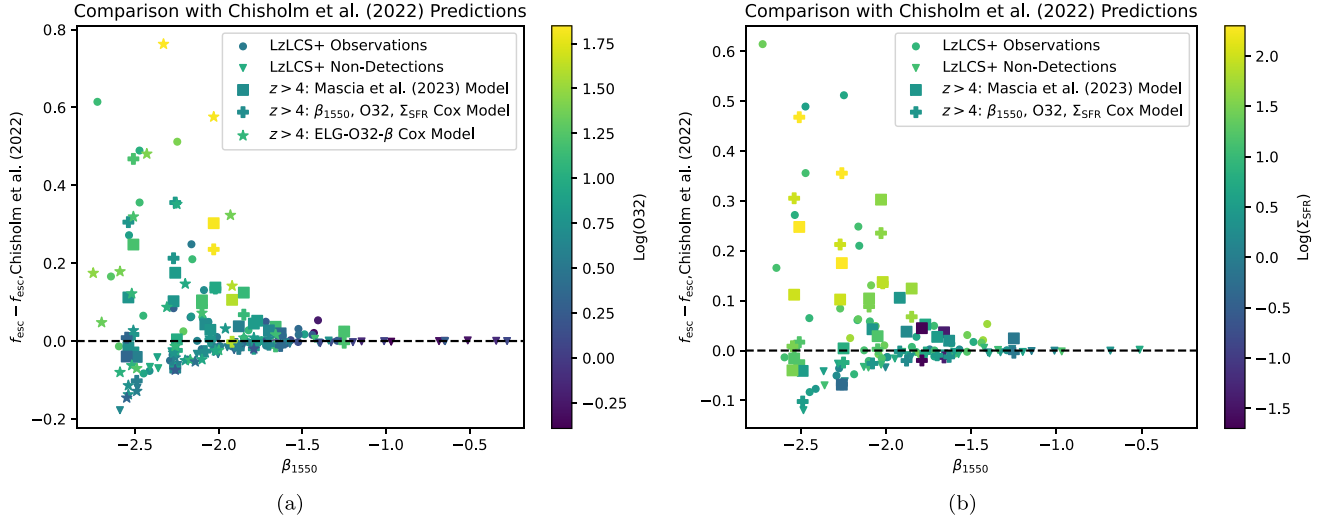
Figure 6 illustrates how including these other variables affects predictions at  $z > 4$ . We plot the difference between the observed LzLCS+  $f_{\text{esc}}$  and the Chisholm et al. (2022) predictions as a function of  $\beta_{1550}$ . Galaxies are color-coded by their O32 ratio (Figure 6(a)) or  $\Sigma_{\text{SFR}}$  (Figure 6(b)). For the reddest  $\beta_{1550}$ , all the model predictions and observations agree that galaxies have little to no LyC escape. However, at blue UV slopes, the disagreement increases. The Chisholm et al. (2022) model assigns all galaxies a single value of  $f_{\text{esc}}$ , whereas the LzLCS+ observations show a range of  $f_{\text{esc}}$  and deviate from the predictions by as much as

$\Delta f_{\text{esc}} = 0.2$  lower or  $\Delta f_{\text{esc}} = 0.6$  higher at the bluest slopes. For the LzLCS+ galaxies, this difference is associated with other differences in the galaxies’ properties; galaxies with higher O32 or higher  $\Sigma_{\text{SFR}}$  tend to be offset to higher  $f_{\text{esc}}$ , and galaxies with lower  $f_{\text{esc}}$  than the Chisholm et al. (2022) predictions tend to have lower O32 or lower  $\Sigma_{\text{SFR}}$ .

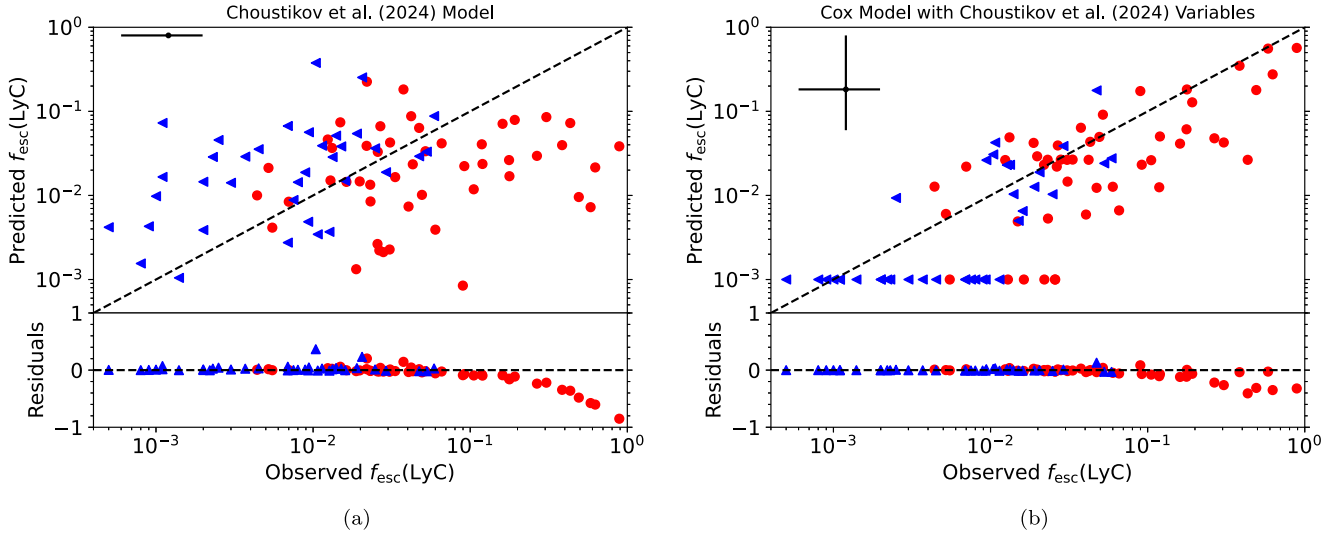
The multivariate model predictions reflect these trends with other properties and assign galaxies a higher or lower  $f_{\text{esc}}$  at fixed  $\beta_{1550}$  accordingly. In Figure 6, we also plot the difference between the predicted  $f_{\text{esc}}$  for  $z > 4$  galaxies from multivariate models versus the single-variable Chisholm et al. (2022) model. All these multivariate models use  $\beta_{1550}$  in addition to two or more other variables. The Mascia et al. (2023) model (squares) predicts  $f_{\text{esc}}$  using a linear fit to  $\beta_{1550}$ ,  $\log_{10}(\text{O32})$ , and  $\log_{10}(r_{50,\text{NUV}})$  for their sample of  $z > 4$  JWST-GLASS galaxies. We show predictions from two Cox models in crosses and stars: a Cox model using the three top-ranked, high-redshift accessible variables from Paper I ( $\beta_{1550}$ , O32, and  $\Sigma_{\text{SFR}}$ ) for the Mascia et al. (2023) sample and the ELG-O32- $\beta$  model (Sections 3 and 4), which uses  $\beta_{1550}$ , O32, and  $M_{1500}$  as inputs to predict  $f_{\text{esc}}$  for  $z > 4$  galaxies from Schaerer et al. (2022), Mascia et al. (2023), Williams et al. (2023), and Saxena et al. (2024). Like the LzLCS+ observations, the multivariate  $z > 4$  predictions can significantly deviate from the Chisholm et al. (2022) predictions, especially at blue UV slopes. High-redshift surveys relying only on  $\beta_{1550}$  could therefore miss some of the physical differences among blue galaxies that might indicate variations in  $f_{\text{esc}}$ . Properties such as O32 and  $\Sigma_{\text{SFR}}$  are sensitive to feedback in galaxies and may complement  $\beta_{1550}$  measurements by tracing a galaxy’s ability to carve low column density HI channels.

## 5.2. Multivariate Predictions

Recently, Choustikov et al. (2024) and Mascia et al. (2023) have developed multivariate linear regression models to predict  $f_{\text{esc}}$  from a set of observable variables. Choustikov et al. (2024) derive their model for  $f_{\text{esc}}$  from synthetic spectra of  $z > 4$



**Figure 6.** Circles (detections) and triangles (upper limits) show the difference between LzLCS+ observations and Chisholm et al. (2022) predicted  $f_{\text{esc}}$  as a function of observed  $\beta_{1550}$ . Other symbols show the difference between multivariate model predictions for  $z > 4$  samples and the Chisholm et al. (2022) single-variable predictions as a function of  $\beta_{1550}$ . Squares compare the  $f_{\text{esc}}$  predictions from Mascia et al. (2023), using  $\beta_{1550}$ , O32, and  $f_{50,\text{NUV}}$ . Crosses show  $f_{\text{esc}}$  predictions for the same  $z > 4$  Mascia et al. (2023) sample for a Cox model using  $\beta_{1550}$ , O32, and  $\Sigma_{\text{SFR}}$ . Stars show the ELG-O32- $\beta$  Cox model predictions for  $z > 4$  galaxies from Schaerer et al. (2022), Mascia et al. (2023), Williams et al. (2023), and Saxena et al. (2024), where the input variables are  $\beta_{1550}$ , O32, and  $M_{1500}$ . We color-code galaxies by their observed O32 ratio in panel (a) and by  $\Sigma_{\text{SFR}}$  in panel (b). At blue UV slopes, the LzLCS+ observations and high-redshift model predictions can differ significantly from the Chisholm et al. (2022) model predictions, depending on the galaxies’ other properties, such as O32 and  $\Sigma_{\text{SFR}}$ .



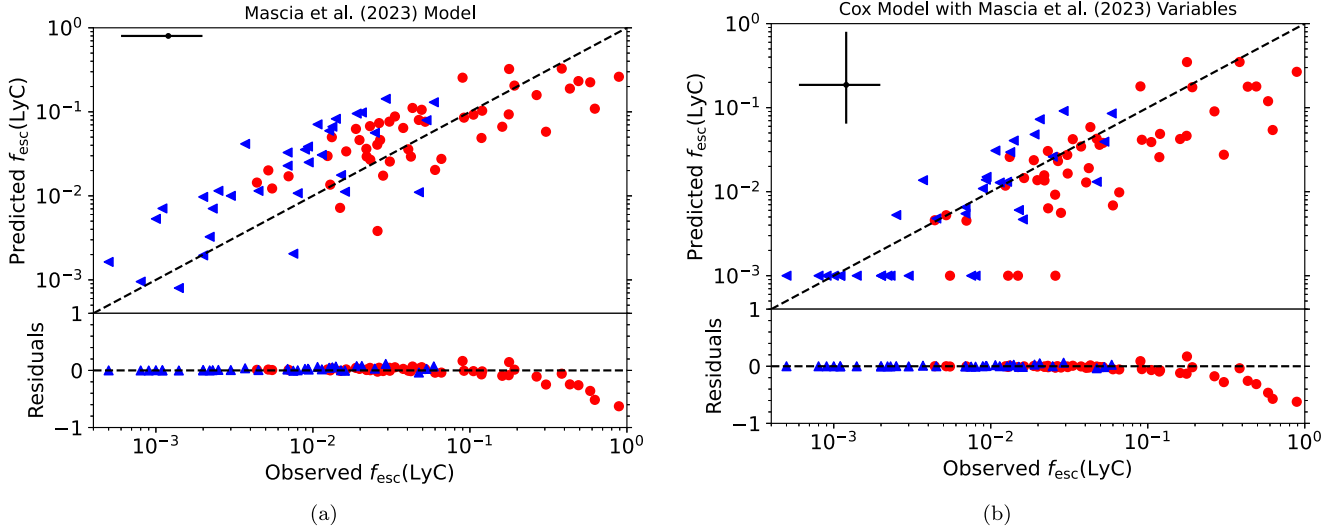
**Figure 7.** (a) The  $f_{\text{esc}}$  predictions from the Choustikov et al. (2024) literature model compared with the LzLCS+ observations. (b) The  $f_{\text{esc}}$  predictions from a Cox model run using the same variables as the Choustikov et al. (2024) model:  $\beta_{1550}$ ,  $E(B - V)_{\text{neb}}$ ,  $L(H\beta)$ ,  $M_{1500}$ , R23, and O32. Symbols are the same as in Figure 1. The Choustikov et al. (2024) model does not reproduce the  $f_{\text{esc}}$  observations from the LzLCS+, but a Cox model using the same input variables does recover the observed  $f_{\text{esc}}$ .

galaxies from the SPHINX cosmological radiation hydrodynamics simulation. Mascia et al. (2023) base their predictions on the LzLCS+ sample but adopt the upper limit in  $f_{\text{esc}}$  as the observed value for nondetections. Here, we use galaxies at low and high redshift to compare the predictions of the Cox models with the predictions from these literature models.

In Figure 7(a), we show the results of applying the Choustikov et al. (2024) model to the LzLCS+ sample, and we list the corresponding goodness-of-fit statistics in Table 10. The Choustikov et al. (2024) model fails to predict the observed  $f_{\text{esc}}$  in the LzLCS+ sample. This result does not depend on our method of calculating  $f_{\text{esc}}$ ; we find a similarly poor fit if we substitute the  $f_{\text{esc}}$  derived from  $H\beta$ , as used in the

Choustikov et al. (2024) simulations. In Figure 7(b), we show  $f_{\text{esc}}$  predictions from a Cox model run using the same set of variables ( $\beta_{1550}$ ,  $E(B - V)_{\text{neb}}$ ,  $L(H\beta)$ ,  $M_{1500}$ , R23 =  $([O\text{ III}]\lambda\lambda 5007, 4959 + [O\text{ II}]\lambda 3727)/H\beta$ , and O32) as in the Choustikov et al. (2024) linear regression mode, and we list the goodness-of-fit metrics in Table 10.

This model performs comparably to the other Cox models in Table 3, which shows that the reason for the Choustikov et al. (2024) model’s difficulty is not its set of variables. Rather, in the Choustikov et al. (2024) model,  $f_{\text{esc}}$  anticorrelates with O32, whereas the LzLCS+ data and the Cox model using the Choustikov et al. (2024) variable set indicate a correlation between  $f_{\text{esc}}$  and O32. The compact, lower-mass LCEs in the



**Figure 8.** (a) The  $f_{\text{esc}}$  predictions from the Mascia et al. (2023) literature model compared with the LzLCS+ observations. (b) The  $f_{\text{esc}}$  predictions from a Cox model run using the same variables as the Mascia et al. (2023) model:  $\beta_{1550}$ ,  $r_{50,\text{NUV}}$ , and O32. Symbols are the same as in Figure 1. Both models perform similarly, but the Cox model does slightly better at predicting  $f_{\text{esc}}$  in weak LCEs and nondetections.

LzLCS do not have counterparts in the SPHINX simulations; these compact LCEs often have high O32 ratios  $\gtrsim 10$  (Flury et al. 2022b), whereas the SPHINX galaxies with O32  $> 10$  typically have lower values of  $f_{\text{esc}}$ . Hence, while we find that high O32 increases  $f_{\text{esc}}$  in our multivariate predictions, Choustikov et al. (2024) find the opposite, that, at fixed  $\beta$  and  $E(B - V)$ , galaxies with older ages and lower O32 have higher  $f_{\text{esc}}$ . Radiative feedback with a turbulent gas structure may allow  $f_{\text{esc}}$  from high O32 galaxies at earlier ages than predicted by cosmological simulations (e.g., Kimm et al. 2019; Kakiichi & Gronke 2021; Choustikov et al. 2024). However, further studies of  $z \sim 3$  galaxies are also necessary to test whether the LzLCS population and the LzLCS-derived Cox models correctly describe the high-redshift population.

Like our Cox model predictions in Section 4, Choustikov et al. (2024) predict a low  $f_{\text{esc}} = 0.03$  for JADES-GS-z7-LA, due to its moderate UV slope and dust content. However, because of the different dependence of  $f_{\text{esc}}$  on O32, the Choustikov et al. (2024) model predicts a dramatically lower  $f_{\text{esc}}$  for CEERS-44, CEERS-698, CEERS-1019, and CEERS-1027 ( $f_{\text{esc}} = 0.006\text{--}0.1$ ) and for JADES 9422 ( $f_{\text{esc}} = 0.01$ ; Saxena et al. 2024) than our model predictions ( $f_{\text{esc}} > 0.4$  to 1). Choustikov et al. (2024) point out that their predicted  $f_{\text{esc}}$  values are consistent with  $f_{\text{esc,Ly}\alpha} > f_{\text{esc}}$  as observed locally (e.g., Flury et al. 2022b), but IGM effects may complicate any comparison with  $f_{\text{esc,Ly}\alpha}$ . As noted above (Section 4), the Cox model predictions for some of these galaxies require extrapolating outside the LzLCS+ parameter space, and their  $f_{\text{esc}}$  may not truly be as extreme as this model predicts. Still, these galaxies remain plausible candidates for high  $f_{\text{esc}}$  given the observed trends seen at  $z \sim 0.3$  and  $z \sim 3$  (Section 3). LyC observations of galaxies with similar properties are necessary to better constrain their  $f_{\text{esc}}$ .

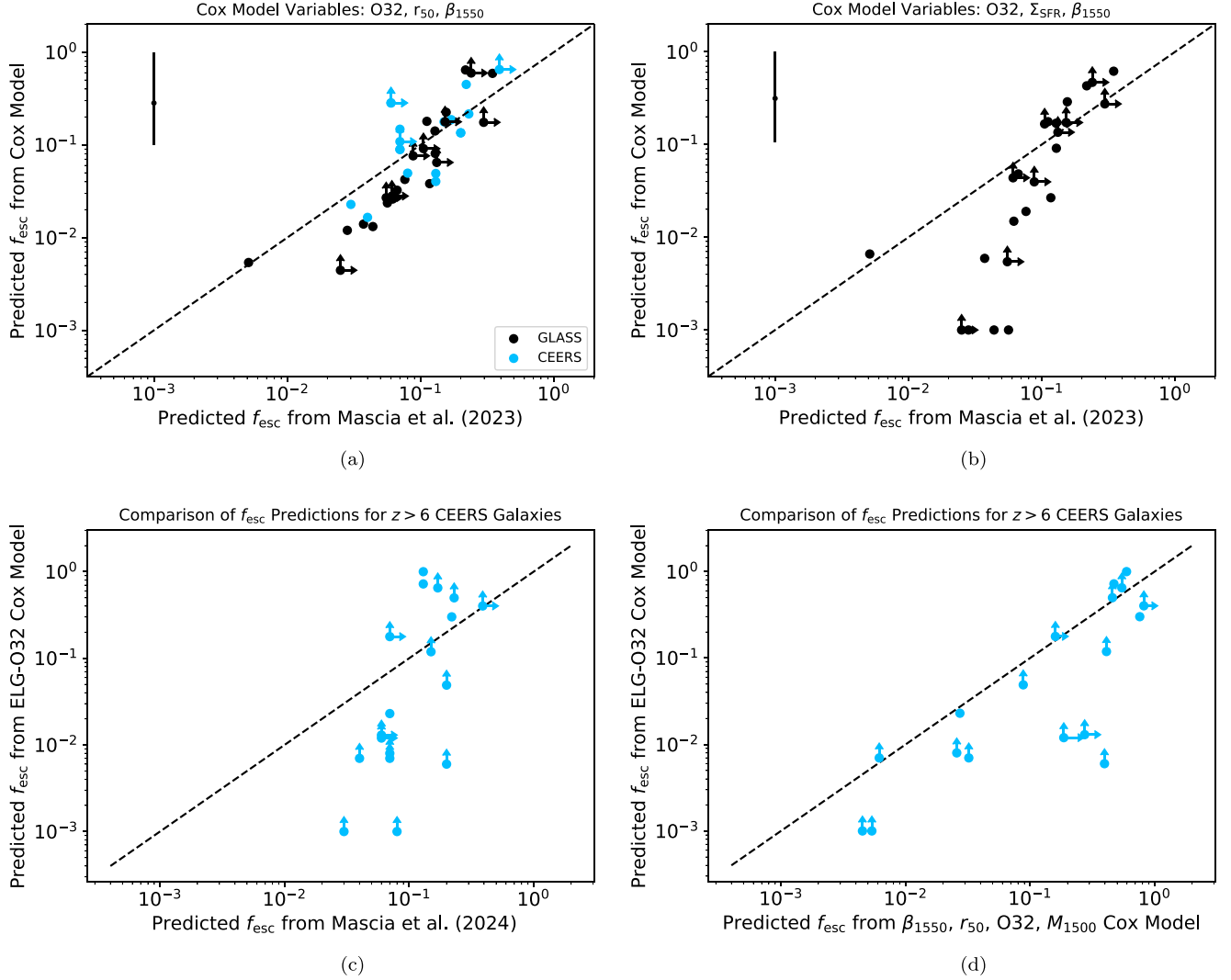
In contrast to the results for the Choustikov et al. (2024) model, the Mascia et al. (2023) model (Figure 8(a) and Table 10) reproduces  $f_{\text{esc}}$  for the LzLCS+ galaxies reasonably well. This finding is unsurprising, since the Mascia et al. (2023) model is in fact derived from the LzLCS+ data set. A Cox model run using the same variables ( $\beta_{1550}$ ,  $r_{50,\text{NUV}}$ , and O32) performs comparably to the Mascia et al. (2023) model

(Figure 8(b) and Table 10). The largest difference occurs for weak LCEs ( $f_{\text{esc}} < 0.05$ ) and nondetections. In these cases, the Mascia et al. (2023) model tends to systematically overpredict  $f_{\text{esc}}$ , whereas by incorporating information from nondetections, Cox model predictions are more evenly distributed above and below the observed values. However, we note that the overprediction of  $f_{\text{esc}}$  in the Mascia et al. (2023) model is small,  $\Delta f_{\text{esc}}$  of only a few percent.

Despite its derivation from the same data set, the Cox and Mascia et al. (2023) models give different predictions at high redshift. The Mascia et al. (2023) model predicts  $f_{\text{esc}} = 0.19$  for Ion2, the only  $z \sim 3$  LCE with the required measurements, whereas the Cox models more accurately match its observed  $f_{\text{esc}} > 0.5$  (Section 3). This difference suggests that galaxy luminosity, incorporated in the Cox models in either  $\Sigma_{\text{SFR}}$  or in  $M_{1500}$ , may be important in reproducing  $f_{\text{esc}}$  for the strongest LCEs.

In Figure 9, we compare the model predictions for GLASS-JWST galaxies (Treu et al. 2022) at  $z = 4\text{--}8$  from Mascia et al. (2023) with predictions from Cox models using similar combinations of variables. For a Cox model using the same variables as in Mascia et al. (2023), the predictions from Mascia et al. (2023) and the corresponding Cox model track each other (Figure 9(a)), aside from the aforementioned tendency of the Mascia et al. (2023) model to give higher predictions at low  $f_{\text{esc}}$ . However, the Mascia et al. (2023) and Cox models disagree more strongly when  $\Sigma_{\text{SFR}}$  is used as a variable in the Cox model instead of  $r_{50,\text{NUV}}$  (Figure 9(b)). The inclusion of  $\Sigma_{\text{SFR}}$  shifts some galaxies with compact sizes but weak star formation to much lower  $f_{\text{esc}}$ . We conclude that including some measure of luminosity or SFR may be important to accurately identify nonleakers and extreme LCEs like Ion2.

Mascia et al. (2024) also apply their multivariate model to galaxies from CEERS and find significantly different results than the predictions from our ELG-O32 Cox model for these galaxies (Section 4, Table 6). We compare these predictions in Figure 9(c). Again, the disagreement primarily arises from the inclusion of luminosity in the ELG-O32 Cox model but not the Mascia et al. (2023) model. As with the GLASS sample, when



**Figure 9.** A comparison of  $f_{\text{esc}}$  predictions for high-redshift galaxies from Cox models and the Mascia et al. (2023) model. (a) A Cox model using  $\beta_{1550}$ ,  $r_{50, \text{NUV}}$ , and O32 gives similar predicted  $f_{\text{esc}}$  as the Mascia et al. (2023) model for  $z > 4$  galaxies from the GLASS-JWST (black) and CEERS (blue) surveys (Mascia et al. 2023, 2024). (b) If  $\Sigma_{\text{SFR}}$  is substituted for  $r_{50, \text{NUV}}$  in the Cox model, the Cox model  $f_{\text{esc}}$  predictions for the GLASS galaxies deviate from those of the Mascia et al. (2023) model. (c) For the CEERS galaxies, the  $f_{\text{esc}}$  predictions from the ELG-O32 Cox model disagree with the Mascia et al. (2024) predictions. (d) A Cox model using  $M_{1500}$  in addition to the Mascia et al. (2023) variables begins to agree more closely with the ELG-O32 Cox model predictions. Adding a measurement of galaxy luminosity in the form of  $\Sigma_{\text{SFR}}$  or  $M_{1500}$  can significantly affect the  $f_{\text{esc}}$  predictions.

we run a Cox model with the same input variables ( $\beta_{1550}$ ,  $r_{50, \text{NUV}}$ , and O32), our predictions agree closely with the Mascia et al. (2024) predictions (Figure 9(a)). Adding one additional variable,  $M_{1500}$ , as a measure of luminosity begins to bring the predicted  $f_{\text{esc}}$  into agreement with the ELG-O32 predictions (Figure 9(d)). The remaining disagreement between the Mascia et al. (2024) and ELG-O32 Cox model predictions is due to the use of radius as a variable in the Mascia et al. (2023) model and different estimates of dust content and ionization from different publications (Mascia et al. 2024; Fujimoto et al. 2023; Tang et al. 2023).

Observational and theoretical studies agree that  $f_{\text{esc}}$  depends on multiple physical parameters. However, studies have not yet reached a consensus as to which parameters matter and how. The predictions for  $f_{\text{esc}}$  in  $z \gtrsim 6$  galaxies from this work, Mascia et al. (2023), Choustikov et al. (2024), and Mascia et al. (2024) disagree because of different adopted scalings with O32 and with luminosity. Observationally testing and distinguishing

between these predictions with larger samples at  $z < 6$  will be critical to reliably predict  $f_{\text{esc}}$  in the epoch of reionization.

## 6. Implications for Reionization

As demonstrated in Section 3, the empirical Cox models derived from the LzLCS+ show promise as diagnostics of  $f_{\text{esc}}$  at high redshift. However, with the limited measurements at  $z > 6$  available so far, the models do not decisively show which galaxy populations dominate the reionization (Section 4.2). Because a combination of properties regulates a galaxy’s optical depth, we need estimates of many factors to accurately predict  $f_{\text{esc}}$ . Dust attenuation, galaxy morphology, ionization, and UV luminosity all play a role in promoting LyC escape. Based primarily on dust attenuation, we would expect faint galaxies to be strong LCEs (e.g., Chisholm et al. 2022; Atek et al. 2023). Nevertheless, at a fixed value of  $\beta_{1550}$ ,  $f_{\text{esc}}$  still shows considerable spread in the LzLCS+ and in multivariate predictions (Figure 6), spanning a range of up to  $\sim 0.7$  in  $f_{\text{esc}}$ . Without estimates of other properties

(e.g.,  $\Sigma_{\text{SFR}}$ , O32), we therefore cannot determine whether faint galaxies have  $f_{\text{esc}}$  above or below the average LzLCS+ value for the same UV color.

At the same time, we cannot yet rule out the contribution of more luminous galaxies. Several galaxies with  $M_{1500} \leq -20$  (e.g., CEERS 498, 1027, 698, 1019; JADES 18846; GLASS-JWST 100003, 10021) have the low dust content and high ionization suggestive of extreme  $f_{\text{esc}} \geq 0.5$ . The LzLCS+ data seem to suggest that, for the same O32 ratio and  $\beta_{1500}$ , a more luminous galaxy will have higher  $f_{\text{esc}}$  (see also Lin et al. 2024). This effect could possibly be an observational bias, as the LyC flux should be easier to detect for more luminous objects. However, we note that, at fixed O32, brighter LzLCS+ galaxies show higher ratios of 900–1100 Å flux, which suggests that their high  $f_{\text{esc}}$  is genuine. Moreover, the trend of higher  $f_{\text{esc}}$  with luminosity only appears in galaxies that have properties associated with LyC escape, such as high O32. Overall, the brighter galaxies in the LzLCS+ have fewer LyC detections (Flury et al. 2022b), which suggests that the detections are not biased toward the brightest galaxies. Furthermore, including UV luminosity in the multivariate models appears necessary to reproduce the  $f_{\text{esc}}$  of Ion2 (Sections 3 and 5), which suggests that the increase of  $f_{\text{esc}}$  with luminosity may be a real phenomenon.

In the multivariate models,  $M_{1500}$  might help to break the degeneracy between ionization parameter and optical depth in galaxies with high O32. For instance, low-metallicity galaxies tend to have both low luminosities and inherently high-ionization parameters (e.g., Dopita et al. 2006; Nagao et al. 2006), such that they may be more likely to have high O32 even at high optical depth. Alternatively, LCEs with higher UV luminosities could be able to ionize gas over a wider opening angle, so that we are statistically more likely to observe a high  $f_{\text{esc}}$  line of sight.

Of course, to evaluate a galaxy population’s influence on reionization, we must also consider how many ionizing photons are produced and the total LyC input into the IGM, not just  $f_{\text{esc}}$ . Unfortunately, the Cox models do not do as well at predicting the ionizing to nonionizing UV flux ratio or the total LyC luminosity (Paper I). Galaxies with high O32 and high nebular EWs may possess elevated  $\xi_{\text{ion}}$  values (e.g., Schaerer et al. 2016; Tang et al. 2019; Maseda et al. 2020; Naidu et al. 2022), such that, at fixed  $f_{\text{esc}}$ , these galaxies will provide more LyC photons to the IGM. These same properties scale with  $f_{\text{esc}}$  in the multivariate Cox models, and our identified candidate strong LCEs at  $z > 6$  generally have O32  $> 10$  and/or EW([O III] + H $\beta$ )  $> 1500$  Å. High  $f_{\text{esc}}$  may therefore be coupled with high  $\xi_{\text{ion}}$  (e.g., Schaerer et al. 2016; Naidu et al. 2022).

However, ionizing photon production may vary with other parameters as well, such as galaxy luminosity (e.g., Bouwens et al. 2016; Finkelstein et al. 2019). Fujimoto et al. (2023) estimate  $\xi_{\text{ion}}$  for spectroscopically confirmed galaxies at  $z \sim 8$ –9 and find that  $\xi_{\text{ion}}$  is 2 times higher at  $M_{\text{UV}} \sim -19.5$  compared to  $M_{\text{UV}} \sim -21.5$ . Although this enhanced efficiency is not enough to compensate for the 6 times fainter UV luminosity at  $M_{\text{UV}} \sim -19.5$  versus  $-21.5$ , these fainter galaxies are also approximately 40 times more numerous (Bouwens et al. 2015). Atek et al. (2023) infer similarly high  $\xi_{\text{ion}}$  for even fainter ( $M_{\text{UV}} \sim -17$  to  $-15$ ) and even more numerous galaxies, which only need  $f_{\text{esc}} < 0.05$  to drive reionization. For the faintest galaxies we consider ( $M_{1500} > -17.5$ ), the median of the Cox model predicted  $f_{\text{esc}}$  values is near this threshold, with the median  $f_{\text{esc}} \sim 0.04$ . However, these  $f_{\text{esc}}$

estimates require refinement with additional parameters (e.g., O32,  $\Sigma_{\text{SFR}}$ ). If  $f_{\text{esc}}$  does not vary strongly with luminosity, as suggested by our preliminary, limited models, the fainter population would dominate the reionization due to their higher  $\xi_{\text{ion}}$  and greater numbers.

Assessing the main contributors to reionization requires progress on several fronts. At  $z > 6$ , we need estimates of nebular and morphological parameters in addition to  $\beta_{1550}$  and  $M_{1500}$  for galaxy samples spanning a wide range of luminosities. Parameters such as O32 and  $\Sigma_{\text{SFR}}$  have some of the greatest effects on the  $f_{\text{esc}}$  prediction accuracy for both the LzLCS+ (Paper I) and galaxies at  $z \sim 3$  (Section 3). The limited magnitude range of the LzLCS+ reference sample ( $M_{1500} \sim -18.5$  to  $\sim -21.5$ ) also introduces uncertainty. Our  $f_{\text{esc}}$  estimates for spectroscopic samples within this magnitude range are likely reasonable and suggest a moderately high median  $f_{\text{esc}} \sim 0.04$ – $0.05$ , with some galaxies reaching  $f_{\text{esc}}$  as high as 0.6–0.7 (see Figure 4 and predictions in Table 6 for the ELG-O32 model and Table 7 for the ELG-O32- $\beta$  model). However, our  $f_{\text{esc}}$  estimates for fainter and brighter galaxy populations rely on extrapolation and are less trustworthy. To confirm these estimates, we need to explore  $f_{\text{esc}}$  and its dependence on galaxy properties across a wider magnitude range at  $z < 6$ . Lastly, observations of the IGM and galaxy population at  $z > 6$  will provide a further test of  $f_{\text{esc}}$  predictions. An accurate model of  $f_{\text{esc}}$  should reproduce both the timing and topology of reionization with the  $z > 6$  galaxy population. While much progress is needed to both improve and confirm  $f_{\text{esc}}$  estimates, our current  $f_{\text{esc}}$  predictions imply that plausible contributors to reionization appear at all magnitude ranges, and star-forming galaxies with the required levels of LyC escape do indeed exist at  $z > 6$ .

## 7. Summary

Quantifying the LyC escape fraction ( $f_{\text{esc}}$ ) of galaxies is critical to understand the sources of cosmic reionization. We have developed a flexible tool for predicting  $f_{\text{esc}}$  empirically using combinations of observable variables available in the  $z \sim 0.3$  LzLCS+ reference sample (Paper I). We generate multivariate diagnostics for  $f_{\text{esc}}$  with the Cox proportional hazards model (Cox 1972), a survival analysis technique that appropriately treats data with upper limits. We test Cox models developed from the  $z \sim 0.3$  galaxies on  $f_{\text{esc}}$  observations at  $z \sim 3$ , and we apply the models to several samples of  $z \gtrsim 6$  galaxies to predict  $f_{\text{esc}}$  for galaxies in the epoch of reionization. The Appendix provides all the information necessary to use the multivariate Cox models to predict  $f_{\text{esc}}$  for other samples of galaxies, including samples at  $z \sim 3$  and samples at  $z \gtrsim 6$ , where the LyC is unobservable.<sup>20</sup>

We summarize our main findings below.

1. The models successfully reproduce the observed  $f_{\text{esc}}$  values for the high-redshift  $z \sim 3$  galaxies and often have a lower rms for the  $z \sim 3$  samples than for the  $z \sim 0.3$  galaxies. The success of these models suggests that low-redshift and high-redshift LCEs may share similar properties (Section 3).
2. The best-performing models for the  $z \sim 3$  galaxies include the dust attenuation inferred from the UV SED

<sup>20</sup> We have also developed a Python script, available at <https://github.com/sflury/LyCsurv>, that allows the user to generate new Cox proportional hazards models to predict  $f_{\text{esc}}$  using custom combinations of variables for any given galaxy population. Version 0.1.0 of the code is archived in Zenodo (Flury et al. 2024).

- $(E(B - V)_{\text{UV}})$  or the UV slope  $\beta_{1550}$ , plus either O32 or a morphological measurement ( $r_{50,\text{NUV}}$  or  $\Sigma_{\text{SFR}}$ ). However, this result is tentative; each model is tested on a different subset of high-redshift galaxies, which makes comparing the models difficult. To determine which variable combinations best predict  $f_{\text{esc}}$  at high redshift, we require larger  $z \sim 3$  samples of LyC-emitters with the full suite of input variables (Section 3).
3. We generate new Cox models based on the LzLCS+ observations, which incorporate variables measured for  $z \gtrsim 6$  samples. One model, using  $M_{1500}$ ,  $M_*$ ,  $E(B - V)_{\text{UV}}$ , and  $\text{EW}([\text{O III}] + \text{H}\beta)$  as input variables, applies to 180 galaxies from Endsley et al. (2021), Saxena et al. (2024), Bouwens et al. (2023), Endsley et al. (2023), Fujimoto et al. (2023), and Tang et al. (2023) and predicts a median  $f_{\text{esc}} = 0.047$  for this combined sample. Of these 180 galaxies, 27% have  $f_{\text{esc}} > 0.1$ , and only 6% have  $f_{\text{esc}} > 0.2$ . However, this set of variables results in predictions that tend to underestimate  $f_{\text{esc}}$ , and the galaxy samples mostly consist of photometric measurements. A second model, using  $M_{1500}$ ,  $M_*$ ,  $\beta_{1550}$ , and  $r_{50,\text{NUV}}$  as input variables, applies to 278 galaxies, mostly with photometric redshifts, from Morishita et al. (2024) and Mascia et al. (2023). This model predicts a higher median  $f_{\text{esc}} = 0.14$ , with 56% and 39% of the galaxies having  $f_{\text{esc}} > 0.1$  and  $f_{\text{esc}} > 0.2$ , respectively. Many of these galaxies are more compact than the LzLCS+ sample, and this model therefore requires extrapolating beyond the LzLCS+ parameter space (Section 4.1).
  4. Smaller samples of spectroscopically confirmed  $z \gtrsim 6$  galaxies have higher predicted  $f_{\text{esc}}$ , likely because they tend to include stronger emission-line galaxies. We use a model with  $\beta_{1550}$ ,  $M_{1500}$ , and O32 to predict  $f_{\text{esc}}$  for 27 spectroscopically confirmed galaxies (Schaerer et al. 2022; Mascia et al. 2023; Saxena et al. 2023, 2024; Williams et al. 2023) and find that 33%–41% of these galaxies have  $f_{\text{esc}} > 0.2$ . For a smaller sample of 17 galaxies (Saxena et al. 2023; Fujimoto et al. 2023; Tang et al. 2023; Williams et al. 2023), using  $M_*$  as an additional input variable and  $E(B - V)_{\text{UV}}$  instead of  $\beta_{1550}$ , we find that 46%–85% of the galaxies have  $f_{\text{esc}} > 0.2$ . These models identify five galaxies at  $z > 5.9$  that may have a line-of-sight  $f_{\text{esc}}$  as high as 0.5–1: CEERS-698, CEERS-1019, CEERS-1027 from Tang et al. (2023), JADES 18846 and JADES 9422 from Saxena et al. (2024; Sections 4.1, 4.3).
  5. Currently, the predicted  $f_{\text{esc}}$  values for different galaxy samples show no strong trend with UV magnitude across the range of  $M_{1500} = -16$  to  $-22$ . However, many of these models are limited by a lack of spectroscopic information. The models suggest that low-luminosity galaxies ( $M_{1500} > -18$ ) have at least moderate  $f_{\text{esc}} \sim 0.05$ . If this  $f_{\text{esc}}$  is coupled with a high ionizing photon production efficiency, such faint galaxies could substantially contribute to reionization (e.g., Atek et al. 2023; Simmonds et al. 2024). Additional measurements of variables such as O32 and  $\Sigma_{\text{SFR}}$  for more  $z \gtrsim 6$  galaxies could enable more accurate  $f_{\text{esc}}$  predictions for both bright and faint galaxies and might increase the numbers of suspected strong LCEs (Section 4.2).
  6. The multivariate predictions for  $z \gtrsim 6$  galaxies can differ strongly from  $f_{\text{esc}}$  predictions based on a single variable. We predict  $f_{\text{esc}}$  for the strong Ly $\alpha$  Emitter, JADES-GS-z7-LA, discovered at  $z = 7.278$  (Saxena et al. 2023). Despite its exceedingly high  $f_{\text{esc,Ly}\alpha}$ , which would seem to imply a high  $f_{\text{esc}}$ , the two Cox models that include the most available information (dust, Ly $\alpha$ , O32, and luminosity) predict that JADES-GS-z7-LA is a moderate LCE with  $f_{\text{esc}} = 0.017$ – $0.085$ . Another example is the galaxy 11027 from Williams et al. (2023), which has O32 = 12, similar to the  $z \gtrsim 6$  galaxies with predicted  $f_{\text{esc}} \geq 0.6$ . Yet, models incorporating other information about the galaxy, such as its mass ( $\log_{10}(M_*/M_{\odot}) = 7.7$ ), UV magnitude ( $M_{1500} = -17.4$ ), and dust attenuation ( $\beta_{1550} = -2$ ), predict  $f_{\text{esc}} \leq 0.02$ . However, we caution that some  $z \gtrsim 6$  galaxies, including these ones, fall outside the parameter space covered by the  $z \sim 0.3$  galaxy sample, which make our  $f_{\text{esc}}$  predictions less reliable (Section 4.3).
  7. We compare our model predictions to single-variable predictions (Chisholm et al. 2022) and to other multivariate  $f_{\text{esc}}$  predictions derived from simulations (Choustikov et al. 2024) or observations (Mascia et al. 2023, 2024). The Cox models more accurately predict  $f_{\text{esc}}$  in nondetections than the  $\beta_{1550}$  model from Chisholm et al. (2022), and the inclusion of feedback-sensitive variables such as O32 and  $\Sigma_{\text{SFR}}$  may better trace the effect of absorption from HI in LCEs with low dust content. Turning to multivariate models, the LzLCS+ sample implies a correlation between O32 and  $f_{\text{esc}}$ , rather than the anticorrelation adopted in the Choustikov et al. (2024) model. As a result, we find different predicted  $f_{\text{esc}}$  for high-redshift galaxies, and the Choustikov et al. (2024) model fails to reproduce the  $f_{\text{esc}}$  of the LzLCS+ galaxies. In contrast, our predictions agree with the Mascia et al. (2023) model, when we use their same set of input variables. However, our high-redshift  $f_{\text{esc}}$  predictions differ strongly when we add UV luminosity or  $\Sigma_{\text{SFR}}$  as input variables.
- LyC-emitting galaxies show similar physical characteristics at both low and high redshift, which suggests that the same processes may govern LyC escape across cosmic time. Because a variety of factors can influence LyC escape, accurately predicting  $f_{\text{esc}}$  requires information about galaxy luminosities, dust, nebular properties, and ionization. Direct measurements of LyC at low redshift, combined with multivariate statistical models for  $f_{\text{esc}}$ , can give insight into the possible  $f_{\text{esc}}$  of galaxies in the epoch of reionization. Nevertheless, additional observations are necessary before these techniques can reach their full potential. At low redshift, LyC observations that push to a new parameter space, especially fainter and brighter UV magnitudes, will aid the application of low-redshift results to high-redshift samples. Measurements of relevant galaxy properties, especially in the rest-frame optical, for larger samples of  $z \sim 3$  LCEs will better test the relative performance of these models at high redshift. Based on our current results, models using O32 and  $\Sigma_{\text{SFR}}$  appear promising but are not yet applicable to large samples at  $z > 6$ . By measuring these properties for larger, representative galaxy samples at  $z > 6$ , high-redshift surveys can connect galaxies in the epoch of reionization with their LyC-emitting counterparts at low redshift.



## Acknowledgments

We thank Aayush Saxena for providing information about the dust attenuation in JADES-GS-z7-LA, and we thank the anonymous referee for feedback that improved the paper’s readability. A.E.J. and S.R.F. acknowledge support from NASA through grant HST-GO-15626. STScI is operated by AURA under NASA contract NAS-5-26555. A.S.L. acknowledges support from Knut and Alice Wallenberg Foundation. M.T. acknowledges support from the NWO grant 0.16.VIDI.189.162 (“ODIN”).

## Appendix

In Tables 11–23, we provide the information necessary to predict  $f_{\text{esc}}$  using the models discussed in this paper, and we give an example of how to use these models below. We have also developed a Python script, available on GitHub,<sup>21</sup> that allows the user to generate and apply new Cox proportional hazards models to predict  $f_{\text{esc}}$  using a desired set of observed variables (version 0.1.0 of the code is archived in Zenodo; Flury et al. 2024).

Here, we provide the parameters for the main models presented in Paper I and in this work. The data for the fiducial model from Paper I appear in Table 11. The best-performing model for the LzLCS+ sample is the fiducial model modified to use EW(HI,abs), and these model parameters appear in Table 12. In Table 13, we list the parameters for the Paper I JWST model, which only uses variables accessible at  $z > 6$ . Two alternative JWST models use only the top-ranked accessible variables:  $\beta_{1550}$ ,  $\Sigma_{\text{SFR}}$ , and either O32 or [Ne III]/[O II]. We provide the data for these models in Tables 14 and 15. Finally, Tables 16–25 present the parameters for the models used for the  $z \sim 3$  and  $z \gtrsim 6$   $f_{\text{esc}}$  predictions discussed in Sections 3–4.

Each table first lists the goodness-of-fit statistics for the LzLCS+ galaxies as a measure of the model performance. It then provides the fitted coefficients ( $b_i$ ) for each included variable and the reference values  $\bar{x}_i$ , which are the mean of the LzLCS+  $x_i$  values, where  $x_i$  is the input variable ( $\beta_{1550}$ ,  $\log_{10}(\text{O32})$ , etc.). Finally, the table lists the baseline cumulative hazard function,  $\text{HF}_0$ , calculated by the `lifelines` Cox model fitting routine for each of the observed  $f_{\text{esc}}$  values for LCEs in the LzLCS+. Together, these parameters predict  $f_{\text{esc}}$  for a given galaxy with a set of input variables  $x$  as follows (see Section 2.2 for more detail).

To predict  $f_{\text{esc}}$  for a galaxy, we first calculate the partial hazards function  $\text{ph}(x)$  for its set of variables:

$$\text{ph}(x) = \exp \left[ \sum_{i=1}^n b_i (x_i - \bar{x}_i) \right]. \quad (\text{A1})$$

We use the coefficients  $b_i$  given in the model table, the galaxy’s observed values  $x_i$  for each variable, and the mean values  $\bar{x}_i$  also given in the model table. For example, for the galaxy Ion2, we have observed values of  $E(B - V)_{\text{UV}} = 0$ ,  $\log_{10}(\text{O32}) = 0.851$ , and  $\log_{10}(\Sigma_{\text{SFR}}/(M_{\odot} \text{ yr}^{-1} \text{ kpc}^{-2})) = 2.03$ . Using  $b_i$  ( $-10.083$ ,  $1.240$ ,  $1.369$ ) and  $\bar{x}_i$  ( $0.158$ ,  $0.521$ ,  $0.705$ ) from

<sup>21</sup> LyCsurv codebase: <https://github.com/sflury/LyCsurv>.

**Table 11**  
Fiducial Model (Paper I)

Statistics			
$R^2$	$R_{\text{adj}}^2$	rms	$C$
0.60	0.49	0.36	0.89
Model Parameters			
Variable	$b_i$	$\bar{x}_i$	
$\log_{10}(M_*/M_{\odot})$	0.626	8.932	
$M_{1500}$	−0.462	−19.84	
$\log_{10}(\text{EW}(\text{H}\beta)/\text{\AA})$	−3.682e-02	1.954	
$E(B - V)_{\text{neb}}$	6.820	0.145	
$12 + \log_{10}(\text{O}/\text{H})$	1.462	8.112	
$\log_{10}([\text{O III}] \lambda 5007/[\text{O II}] \lambda 3727)$	2.651	0.521	
$\log_{10}(\Sigma_{\text{SFR}}/(M_{\odot} \text{ yr}^{-1} \text{ kpc}^{-2}))$	1.834	0.705	
$E(B - V)_{\text{UV}}$	−14.897	0.158	
$f_{\text{esc,Ly}\alpha}$	7.002	0.220	
$\text{EW}(\text{LIS})(\text{\AA})^a$	−0.264	1.060	
Baseline Cumulative Hazard			
$f_{\text{esc}}$	$H_0(f_{\text{esc}})$	$f_{\text{esc}}(\text{cont.})$	$H_0(\text{cont.})$
0.8890	0.001038	0.0376	0.099723
0.6247	0.002243	0.0333	0.110434
0.5838	0.003495	0.0309	0.121567
0.4911	0.004907	0.0307	0.132833
0.4333	0.006444	0.0280	0.146030
0.3845	0.008002	0.0268	0.160176
0.3052	0.009680	0.0264	0.175696
0.2659	0.011478	0.0258	0.191559
0.1921	0.013460	0.0257	0.208518
0.1777	0.015602	0.0232	0.226165
0.1767	0.017962	0.0231	0.244148
0.1607	0.020514	0.0220	0.264128
0.1197	0.023136	0.0219	0.284535
0.1185	0.025995	0.0198	0.310698
0.1053	0.028889	0.0188	0.349019
0.0917	0.031854	0.0163	0.390900
0.0898	0.035544	0.0149	0.436669
0.0658	0.041090	0.0132	0.489950
0.0600	0.046719	0.0129	0.552766
0.0519	0.052798	0.0124	0.618132
0.0493	0.059108	0.0070	0.765657
0.0473	0.065551	0.0055	1.052480
0.0431	0.073313	0.0052	1.396596
0.0421	0.081962	0.0044	1.962125
0.0405	0.090790	...	...

### Note.

<sup>a</sup> For this variable, a positive EW denotes net absorption.

Table 16, we calculate

$$\begin{aligned} \text{ph}(x) = & \exp[-10.083*(0 - 0.158) + 1.240*(0.851 - 0.521) \\ & + 1.369*(2.03 - 0.705)] = 45.43. \end{aligned} \quad (\text{A2})$$

We then use this value of  $\text{ph}(x)$  to scale the baseline cumulative hazard function  $\text{HF}_0(f_{\text{esc}})$  and calculate the survival function  $S(f_{\text{esc}})$ , which represents the probability that the true escape fraction,  $f_{\text{esc,true}}$ , is lower than each tabulated  $f_{\text{esc}}$ :

$$S(f_{\text{esc}}) = \exp[-\text{HF}_0(f_{\text{esc}})*\text{ph}(x)]. \quad (\text{A3})$$

From Table 16, for  $f_{\text{esc}} = 0.8890$ ,  $\text{HF}_0$  is 0.003949, and  $S(0.8890)$  is therefore  $\exp[-0.003949*45.43] = 0.8358$ . For  $f_{\text{esc}} = 0.6247$ ,

**Table 12**  
Fiducial Model with EW(H I,abs) (Paper I)

Statistics			
$R^2$	$R_{\text{adj}}^2$	rms	$C$
0.69	0.60	0.31	0.91
Model Parameters			
Variable	$b_i$	$\bar{x}_i$	
$\log_{10}(M_*/M_\odot)$	0.454	8.932	
$M_{1500}$	-1.370	-19.84	
$\log_{10}(\text{EW}(\text{H}\beta)/\text{\AA})$	-2.756	1.954	
$E(B - V)_{\text{neb}}$	7.127	0.145	
$12 + \log_{10}(\text{O}/\text{H})$	1.729	8.112	
$\log_{10}([\text{O III}] \lambda 5007 / [\text{O II}] \lambda 3727)$	5.170	0.521	
$\log_{10}(\Sigma_{\text{SFR}} / (M_\odot \text{ yr}^{-1} \text{ kpc}^{-2}))$	0.475	0.705	
$E(B - V)_{\text{UV}}$	-21.358	0.158	
$f_{\text{esc,Ly}\alpha}$	3.988	0.220	
$\text{EW}(\text{H I,abs})(\text{\AA})^{\text{a}}$	-1.487	2.468	
Baseline Cumulative Hazard			
$f_{\text{esc}}$	$H_0(f_{\text{esc}})$	$f_{\text{esc}}(\text{cont.})$	$H_0(\text{cont.})$
0.8890	0.000270	0.0376	0.084116
0.6247	0.000750	0.0333	0.093819
0.5838	0.001276	0.0309	0.104034
0.4911	0.001888	0.0307	0.114298
0.4333	0.002637	0.0280	0.126069
0.3845	0.003402	0.0268	0.139434
0.3052	0.004460	0.0264	0.154484
0.2659	0.005616	0.0258	0.170455
0.1921	0.006786	0.0257	0.186968
0.1777	0.008170	0.0232	0.203982
0.1767	0.009731	0.0231	0.221288
0.1607	0.011393	0.0220	0.240703
0.1197	0.013207	0.0219	0.260881
0.1185	0.015226	0.0198	0.285998
0.1053	0.017322	0.0188	0.319845
0.0917	0.019459	0.0163	0.367457
0.0898	0.023429	0.0149	0.419286
0.0658	0.028468	0.0132	0.479264
0.0600	0.033620	0.0129	0.554142
0.0519	0.039120	0.0124	0.633334
0.0493	0.045164	0.0070	0.793057
0.0473	0.051386	0.0055	1.175799
0.0431	0.058404	0.0052	1.623712
0.0421	0.066728	0.0044	2.096537
0.0405	0.075377	...	...

**Note.**<sup>a</sup> For this variable, a positive EW denotes net absorption.

$HF_0 = 0.008253$ , and  $S(0.6247) = \exp[-0.008253 * 45.43] = 0.6873$ , and so on, for each value of  $f_{\text{esc}}$ . Thus, according to the TopThree model, there is a 68.73% probability that Ion2 has a measured  $f_{\text{esc}} < 0.6247$  and an 83.58% probability that Ion2's  $f_{\text{esc}} < 0.8890$ . To find the median predicted  $f_{\text{esc}}$ , we calculate  $S$  for each value of  $f_{\text{esc}}$  in the table and find the value where  $S$  reaches 0.5;  $f_{\text{esc}}$  is predicted to be above this value 50% of the time and below this

**Table 13**  
Full JWST Model (Paper I)

Statistics			
$R^2$	$R_{\text{adj}}^2$	rms	$C$
0.29	0.14	0.47	0.83
Model Parameters			
Variable	$b_i$	$\bar{x}_i$	
$\log_{10}(M_*/M_\odot)$	6.877e-02	8.932	
$M_{1500}$	-0.397	-19.84	
$\log_{10}(\text{EW}(\text{H}\beta)/\text{\AA})$	-1.181	1.954	
$E(B - V)_{\text{neb}}$	0.710	0.145	
$12 + \log_{10}(\text{O}/\text{H})$	-2.750e-02	8.112	
$\log_{10}([\text{O III}] \lambda 5007 / [\text{O II}] \lambda 3727)$	2.740	0.521	
$\log_{10}(\Sigma_{\text{SFR}} / (M_\odot \text{ yr}^{-1} \text{ kpc}^{-2}))$	1.082	0.705	
$\beta_{1550}$	-2.166	-1.810	
Baseline Cumulative Hazard			
$f_{\text{esc}}$	$HF_0(f_{\text{esc}})$	$f_{\text{esc}}(\text{cont.})$	$HF_0(\text{cont.})$
0.8890	0.003071	0.0376	0.196676
0.6247	0.006718	0.0333	0.211196
0.5838	0.010477	0.0309	0.226775
0.4911	0.014362	0.0307	0.242624
0.4333	0.018463	0.0280	0.261375
0.3845	0.022643	0.0268	0.280335
0.3052	0.027504	0.0264	0.301007
0.2659	0.032501	0.0258	0.322308
0.1921	0.037624	0.0257	0.343655
0.1777	0.043075	0.0232	0.365982
0.1767	0.049150	0.0231	0.388662
0.1607	0.055522	0.0220	0.413036
0.1197	0.062102	0.0219	0.437812
0.1185	0.069015	0.0198	0.466136
0.1053	0.076045	0.0188	0.496786
0.0917	0.083300	0.0163	0.529638
0.0898	0.090759	0.0149	0.565906
0.0658	0.100730	0.0132	0.611820
0.0600	0.110801	0.0129	0.672871
0.0519	0.121542	0.0124	0.739342
0.0493	0.132930	0.0070	0.872244
0.0473	0.144577	0.0055	1.060466
0.0431	0.156904	0.0052	1.258225
0.0421	0.169902	0.0044	1.535025
0.0405	0.183203	...	...

value 50% of the time. For Ion2, we see that  $S(0.5838) = 0.5629$ , and  $S(0.4911) = 0.4577$ . We linearly interpolate to find where  $S(f_{\text{esc}}) = 0.5$ , finding that  $S(0.53) \sim 0.5$ . The predicted  $f_{\text{esc}}$  of Ion2 is then  $f_{\text{esc}} = 0.53$ . Similarly, to determine the uncertainties in the predicted  $f_{\text{esc}}$ , we find where  $S(f_{\text{esc}})$  reaches 0.159 and 0.841. According to the model,  $f_{\text{esc}}$  will be between these  $f_{\text{esc}}$  values 68% of the time. For some galaxies,  $S(f_{\text{esc}})$  is always  $> 0.5$  for the tabulated  $f_{\text{esc}}$  values, indicating a  $> 50\%$  probability that  $f_{\text{esc}}$  is smaller than the smallest tabulated  $f_{\text{esc}}$ . This situation corresponds to an arbitrarily small predicted  $f_{\text{esc}}$ ,  $f_{\text{esc}} \sim 0$ . Conversely, if  $S(f_{\text{esc}})$  is always  $< 0.5$ ,  $f_{\text{esc}}$  is arbitrarily large, and  $f_{\text{esc}} \sim 1$ .

**Table 14**  
Limited JWST Model (Paper I):  $\beta_{1550}$ ,  $\log_{10}(\Sigma_{\text{SFR}})$ ,  $\log_{10}(\text{O32})$

Statistics			
$R^2$	$R_{\text{adj}}^2$	rms	$C$
0.34	0.29	0.46	0.83
Model Parameters			
Variable	$b_i$	$\bar{x}_i$	
$\log_{10}([\text{O III}] \lambda 5007 / [\text{O II}] \lambda 3727)$	0.996	0.521	
$\log_{10}(\Sigma_{\text{SFR}} / (M_{\odot} \text{ yr}^{-1} \text{ kpc}^{-2}))$	1.404	0.705	
$\beta_{1550}$	-2.274	-1.810	
Baseline Cumulative Hazard			
$f_{\text{esc}}$	$\text{HF}_0 (f_{\text{esc}})$	$f_{\text{esc}} (\text{cont.})$	$\text{HF}_0 (\text{cont.})$
0.8890	0.003311	0.0376	0.196361
0.6247	0.007118	0.0333	0.210168
0.5838	0.011019	0.0309	0.225189
0.4911	0.015019	0.0307	0.240530
0.4333	0.019271	0.0280	0.259278
0.3845	0.023685	0.0268	0.278291
0.3052	0.028643	0.0264	0.298441
0.2659	0.033702	0.0258	0.319131
0.1921	0.038913	0.0257	0.339855
0.1777	0.044432	0.0232	0.361703
0.1767	0.050682	0.0231	0.383929
0.1607	0.057192	0.0220	0.407993
0.1197	0.063925	0.0219	0.432411
0.1185	0.070990	0.0198	0.460300
0.1053	0.078219	0.0188	0.491690
0.0917	0.085704	0.0163	0.525259
0.0898	0.093495	0.0149	0.563415
0.0658	0.103034	0.0132	0.615517
0.0600	0.112666	0.0129	0.678401
0.0519	0.123204	0.0124	0.747133
0.0493	0.134221	0.0070	0.893528
0.0473	0.145450	0.0055	1.105639
0.0431	0.157431	0.0052	1.324779
0.0421	0.170192	0.0044	1.623314
0.0405	0.183168	...	...

**Table 15**  
Limited JWST Model (Paper I):  $\beta_{1550}$ ,  $\log_{10}(\Sigma_{\text{SFR}})$ ,  $\log_{10}([\text{Ne III}] / [\text{O II}])$

Statistics			
$R^2$	$R_{\text{adj}}^2$	rms	$C$
0.40	0.35	0.44	0.83
Model Parameters			
Variable	$b_i$	$\bar{x}_i$	
$\log_{10}([\text{Ne III}] \lambda 3869 / [\text{O II}] \lambda 3727)$	1.315	-0.532	
$\log_{10}(\Sigma_{\text{SFR}} / (M_{\odot} \text{ yr}^{-1} \text{ kpc}^{-2}))$	1.343	0.705	
$\beta_{1550}$	-2.192	-1.810	
Baseline Cumulative Hazard			
$f_{\text{esc}}$	$\text{HF}_0 (f_{\text{esc}})$	$f_{\text{esc}} (\text{cont.})$	$\text{HF}_0 (\text{cont.})$
0.8890	0.003276	0.0376	0.196231
0.6247	0.007022	0.0333	0.210160
0.5838	0.010866	0.0309	0.225310
0.4911	0.014827	0.0307	0.240797
0.4333	0.019033	0.0280	0.260131
0.3845	0.023395	0.0268	0.279810
0.3052	0.028307	0.0264	0.300595
0.2659	0.033303	0.0258	0.322009
0.1921	0.038458	0.0257	0.343462
0.1777	0.043933	0.0232	0.366090
0.1767	0.050163	0.0231	0.389078
0.1607	0.056638	0.0220	0.413937
0.1197	0.063307	0.0219	0.439215
0.1185	0.070313	0.0198	0.467915
0.1053	0.077462	0.0188	0.500143
0.0917	0.084871	0.0163	0.535109
0.0898	0.092566	0.0149	0.574662
0.0658	0.102235	0.0132	0.628551
0.0600	0.111988	0.0129	0.693361
0.0519	0.122604	0.0124	0.764890
0.0493	0.133634	0.0070	0.918966
0.0473	0.144908	0.0055	1.139392
0.0431	0.156907	0.0052	1.368115
0.0421	0.169810	0.0044	1.677980
0.0405	0.182933	...	...

**Table 16**  
TopThree Model

Statistics			
$R^2$	$R_{\text{adj}}^2$	rms	$C$
0.38	0.34	0.44	0.82
Model Parameters			
Variable	$b_i$	$\bar{x}_i$	
$E(B - V)_{\text{UV}}$	-10.083	0.158	
$\log_{10}([\text{O III}] \lambda 5007 / [\text{O II}] \lambda 3727)$	1.240	0.521	
$\log_{10}(\Sigma_{\text{SFR}} / (M_{\odot} \text{ yr}^{-1} \text{ kpc}^{-2}))$	1.369	0.705	
Baseline Cumulative Hazard			
$f_{\text{esc}}$	$\text{HF}_0 (f_{\text{esc}})$	$f_{\text{esc}} (\text{cont.})$	$\text{HF}_0 (\text{cont.})$
0.8890	0.003949	0.0376	0.206394
0.6247	0.008253	0.0333	0.220201
0.5838	0.012651	0.0309	0.234665
0.4911	0.017203	0.0307	0.249373
0.4333	0.022038	0.0280	0.268514
0.3845	0.027025	0.0268	0.287870
0.3052	0.032320	0.0264	0.308337
0.2659	0.037741	0.0258	0.329388
0.1921	0.043347	0.0257	0.350502
0.1777	0.049274	0.0232	0.372931
0.1767	0.055904	0.0231	0.395743
0.1607	0.062835	0.0220	0.420512
0.1197	0.069951	0.0219	0.445571
0.1185	0.077408	0.0198	0.473320
0.1053	0.085092	0.0188	0.505102
0.0917	0.093180	0.0163	0.539307
0.0898	0.101614	0.0149	0.578156
0.0658	0.111233	0.0132	0.630685
0.0600	0.120940	0.0129	0.688622
0.0519	0.131675	0.0124	0.754340
0.0493	0.142953	0.0070	0.895800
0.0473	0.154429	0.0055	1.107382
0.0431	0.166817	0.0052	1.327823
0.0421	0.179768	0.0044	1.631787
0.0405	0.192981	...	...

**Table 17**  
LAE Model

Statistics			
$R^2$	$R_{\text{adj}}^2$	rms	$C$
0.21	0.15	0.50	0.81
Model Parameters			
Variable	$b_i$	$\bar{x}_i$	
$M_{1500}$	-1.022	-19.84	
$E(B - V)_{\text{UV}}$	-13.943	0.157	
$\text{EW}(\text{Ly}\alpha)(\text{\AA})^{\text{a}}$	1.716e-02	68.16	
Baseline Cumulative Hazard			
$f_{\text{esc}}$	$H_0 (f_{\text{esc}})$	$f_{\text{esc}} (\text{cont.})$	$H_0 (\text{cont.})$
0.8890	0.004221	0.0376	0.217054
0.6247	0.008679	0.0333	0.232258
0.5838	0.013292	0.0309	0.247919
0.4911	0.018280	0.0307	0.264020
0.4333	0.023421	0.0280	0.281165
0.3845	0.028705	0.0268	0.298428
0.3052	0.034055	0.0264	0.316930
0.2659	0.039708	0.0258	0.335720
0.1921	0.046128	0.0257	0.355020
0.1777	0.053219	0.0232	0.375092
0.1767	0.060496	0.0231	0.395379
0.1607	0.068146	0.0220	0.416328
0.1197	0.075984	0.0219	0.437797
0.1185	0.084457	0.0198	0.461351
0.1053	0.093132	0.0188	0.489491
0.0917	0.102126	0.0163	0.518083
0.0898	0.111330	0.0149	0.549562
0.0658	0.120646	0.0132	0.584802
0.0600	0.130097	0.0129	0.623000
0.0519	0.140548	0.0124	0.671296
0.0493	0.151240	0.0070	0.788813
0.0473	0.162539	0.0055	0.952357
0.0431	0.175252	0.0052	1.126629
0.0421	0.188749	0.0044	1.368313
0.0405	0.202829	...	...

**Note.**<sup>a</sup> Positive EW(Ly $\alpha$ ) denotes net emission.

**Table 18**  
LAE-O32 Model

Statistics			
$R^2$	$R_{\text{adj}}^2$	rms	$C$
0.37	0.29	0.45	0.82
Model Parameters			
Variable	$b_i$	$\bar{x}_i$	
$M_{1500}$	-0.935	-19.84	
$\log_{10}(M_*/M_{\odot})$	0.961	8.920	
$E(B - V)_{\text{UV}}$	-15.987	0.157	
$\log_{10}([\text{O III}] \lambda 5007 / [\text{O II}] \lambda 3727)$	1.863	0.523	
$\text{EW}(\text{Ly}\alpha)(\text{\AA})^{\text{a}}$	1.506e-02	68.16	
Baseline Cumulative Hazard			
$f_{\text{esc}}$	$H_0(f_{\text{esc}})$	$f_{\text{esc}}(\text{cont.})$	$H_0(\text{cont.})$
0.8890	0.002782	0.0376	0.186577
0.6247	0.005695	0.0333	0.200373
0.5838	0.008706	0.0309	0.214414
0.4911	0.012297	0.0307	0.228667
0.4333	0.016140	0.0280	0.245277
0.3845	0.020018	0.0268	0.261970
0.3052	0.024057	0.0264	0.280210
0.2659	0.028335	0.0258	0.298707
0.1921	0.033372	0.0257	0.317460
0.1777	0.038847	0.0232	0.337641
0.1767	0.044733	0.0231	0.358145
0.1607	0.051327	0.0220	0.379255
0.1197	0.058081	0.0219	0.400762
0.1185	0.065680	0.0198	0.424827
0.1053	0.073441	0.0188	0.451867
0.0917	0.081748	0.0163	0.479495
0.0898	0.090210	0.0149	0.510421
0.0658	0.099141	0.0132	0.545046
0.0600	0.108225	0.0129	0.583697
0.0519	0.117912	0.0124	0.633505
0.0493	0.127906	0.0070	0.745027
0.0473	0.138188	0.0055	0.916397
0.0431	0.149428	0.0052	1.095363
0.0421	0.161445	0.0044	1.343768
0.0405	0.173974	...	...

**Note.**<sup>a</sup> Positive EW(Ly $\alpha$ ) denotes net emission.**Table 19**  
LAE-O32-nodust Model

Statistics			
$R^2$	$R_{\text{adj}}^2$	rms	$C$
0.02	-0.06	0.59	0.77
Model Parameters			
Variable	$b_i$	$\bar{x}_i$	
$M_{1500}$	-0.913	-19.84	
$\log_{10}([\text{O III}] \lambda 5007 / [\text{O II}] \lambda 3727)$	3.475	0.523	
$\text{EW}(\text{Ly}\alpha)(\text{\AA})^{\text{a}}$	2.219e-03	68.16	
Baseline Cumulative Hazard			
$f_{\text{esc}}$	$H_0(f_{\text{esc}})$	$f_{\text{esc}}(\text{cont.})$	$H_0(\text{cont.})$
0.8890	0.006247	0.0376	0.266926
0.6247	0.012954	0.0333	0.282683
0.5838	0.019834	0.0309	0.299106
0.4911	0.027167	0.0307	0.315676
0.4333	0.034710	0.0280	0.333150
0.3845	0.042320	0.0268	0.350663
0.3052	0.050342	0.0264	0.369016
0.2659	0.058559	0.0258	0.388124
0.1921	0.067137	0.0257	0.407508
0.1777	0.076178	0.0232	0.427799
0.1767	0.085515	0.0231	0.448402
0.1607	0.095217	0.0220	0.469654
0.1197	0.104993	0.0219	0.491134
0.1185	0.115165	0.0198	0.513681
0.1053	0.125425	0.0188	0.539040
0.0917	0.135985	0.0163	0.565361
0.0898	0.146749	0.0149	0.593704
0.0658	0.158378	0.0132	0.626154
0.0600	0.170137	0.0129	0.661094
0.0519	0.182291	0.0124	0.708036
0.0493	0.194771	0.0070	0.783818
0.0473	0.208242	0.0055	0.882875
0.0431	0.222481	0.0052	0.989315
0.0421	0.237057	0.0044	1.116652
0.0405	0.251873	...	...

**Note.**<sup>a</sup> Positive EW(Ly $\alpha$ ) denotes net emission.

**Table 20**  
ELG-EW Model

Statistics			
$R^2$	$R_{\text{adj}}^2$	rms	$C$
0.14	0.05	0.53	0.79
Model Parameters			
Variable	$b_i$	$\bar{x}_i$	
$M_{1500}$	-0.248	-19.84	
$\log_{10}(M_*/M_{\odot})$	0.771	8.920	
$E(B - V)_{\text{UV}}$	-12.149	0.157	
$\log_{10}(\text{EW}([\text{O III}]\lambda\lambda 5007, 4959 + \text{H}\beta)/\text{\AA})$	1.893	2.725	
Baseline Cumulative Hazard			
$f_{\text{esc}}$	$\text{HF}_0(f_{\text{esc}})$	$f_{\text{esc}}(\text{cont.})$	$\text{HF}_0(\text{cont.})$
0.8890	0.005687	0.0376	0.240170
0.6247	0.011541	0.0333	0.254842
0.5838	0.017533	0.0309	0.269815
0.4911	0.023684	0.0307	0.285298
0.4333	0.030206	0.0280	0.302262
0.3845	0.036819	0.0268	0.319275
0.3052	0.043720	0.0264	0.337274
0.2659	0.050851	0.0258	0.355343
0.1921	0.058194	0.0257	0.373607
0.1777	0.065762	0.0232	0.392994
0.1767	0.073806	0.0231	0.412665
0.1607	0.082393	0.0220	0.433109
0.1197	0.091141	0.0219	0.453932
0.1185	0.100320	0.0198	0.477037
0.1053	0.109807	0.0188	0.502072
0.0917	0.119757	0.0163	0.527919
0.0898	0.129963	0.0149	0.559078
0.0658	0.140592	0.0132	0.596941
0.0600	0.151376	0.0129	0.637311
0.0519	0.162692	0.0124	0.679393
0.0493	0.174373	0.0070	0.783432
0.0473	0.186132	0.0055	0.931661
0.0431	0.198855	0.0052	1.084933
0.0421	0.212176	0.0044	1.283606
0.0405	0.226118	...	...

**Table 21**  
ELG-O32 Model

Statistics			
$R^2$	$R_{\text{adj}}^2$	rms	$C$
0.42	0.36	0.44	0.79
Model Parameters			
Variable	$b_i$	$\bar{x}_i$	
$M_{1500}$	-0.727	-19.84	
$\log_{10}(M_*/M_{\odot})$	0.693	8.920	
$E(B - V)_{\text{UV}}$	-10.230	0.157	
$\log_{10}([\text{O III}]\lambda 5007/[\text{O II}]\lambda 3727)$	3.810	0.523	
Baseline Cumulative Hazard			
$f_{\text{esc}}$	$\text{HF}_0(f_{\text{esc}})$	$f_{\text{esc}}(\text{cont.})$	$\text{HF}_0(\text{cont.})$
0.8890	0.003619	0.0376	0.205717
0.6247	0.007534	0.0333	0.219975
0.5838	0.011580	0.0309	0.234524
0.4911	0.015999	0.0307	0.249271
0.4333	0.020713	0.0280	0.266818
0.3845	0.025461	0.0268	0.284396
0.3052	0.030795	0.0264	0.303606
0.2659	0.036338	0.0258	0.323166
0.1921	0.042111	0.0257	0.342907
0.1777	0.048239	0.0232	0.363696
0.1767	0.055076	0.0231	0.384742
0.1607	0.062454	0.0220	0.406670
0.1197	0.069944	0.0219	0.428908
0.1185	0.077866	0.0198	0.453903
0.1053	0.085926	0.0188	0.481290
0.0917	0.094440	0.0163	0.509903
0.0898	0.103111	0.0149	0.541931
0.0658	0.112911	0.0132	0.579921
0.0600	0.122814	0.0129	0.622891
0.0519	0.133160	0.0124	0.670930
0.0493	0.144055	0.0070	0.779178
0.0473	0.155155	0.0055	0.930883
0.0431	0.167143	0.0052	1.089376
0.0421	0.179615	0.0044	1.293047
0.0405	0.192628	...	...

**Table 22**  
ELG-O32- $\beta$  Model

Statistics			
$R^2$	$R_{\text{adj}}^2$	rms	$C$
0.30	0.25	0.48	0.79
Model Parameters			
Variable	$b_i$	$\bar{x}_i$	
$M_{1500}$	-0.852	-19.84	
$\beta_{1550}$	-1.479	-1.819	
$\log_{10}([\text{O III}] \lambda 5007 / [\text{O II}] \lambda 3727)$	3.307	0.523	
Baseline Cumulative Hazard			
$f_{\text{esc}}$	$H_0(f_{\text{esc}})$	$f_{\text{esc}}(\text{cont.})$	$H_0(\text{cont.})$
0.8890	0.004039	0.0376	0.219993
0.6247	0.008739	0.0333	0.234945
0.5838	0.013629	0.0309	0.250650
0.4911	0.018793	0.0307	0.266648
0.4333	0.024186	0.0280	0.284011
0.3845	0.029664	0.0268	0.301400
0.3052	0.035829	0.0264	0.320107
0.2659	0.042177	0.0258	0.339289
0.1921	0.048750	0.0257	0.358640
0.1777	0.055873	0.0232	0.378715
0.1767	0.063520	0.0231	0.398969
0.1607	0.071496	0.0220	0.420108
0.1197	0.079577	0.0219	0.441571
0.1185	0.088029	0.0198	0.465104
0.1053	0.096590	0.0188	0.491366
0.0917	0.105423	0.0163	0.518661
0.0898	0.114447	0.0149	0.548480
0.0658	0.124389	0.0132	0.584009
0.0600	0.134414	0.0129	0.625072
0.0519	0.144961	0.0124	0.671057
0.0493	0.155928	0.0070	0.780254
0.0473	0.167396	0.0055	0.920758
0.0431	0.179884	0.0052	1.068231
0.0421	0.192936	0.0044	1.252412
0.0405	0.206403	...	...

**Table 23**  
ELG-O32- $\beta$ -Ly $\alpha$  Model

Statistics			
$R^2$	$R_{\text{adj}}^2$	rms	$C$
0.40	0.32	0.45	0.81
Model Parameters			
Variable	$b_i$	$\bar{x}_i$	
$M_{1500}$	-0.488	-19.84	
$\log_{10}(M_*/M_{\odot})$	0.969	8.920	
$\beta_{1550}$	-1.863	-1.819	
$\log_{10}([\text{O III}] \lambda 5007 / [\text{O II}] \lambda 3727)$	3.138	0.523	
$f_{\text{esc,Ly}\alpha}$	2.323	0.221	
Baseline Cumulative Hazard			
$f_{\text{esc}}$	$H_0(f_{\text{esc}})$	$f_{\text{esc}}(\text{cont.})$	$H_0(\text{cont.})$
0.8890	0.003503	0.0376	0.204488
0.6247	0.007231	0.0333	0.219681
0.5838	0.011119	0.0309	0.235210
0.4911	0.015539	0.0307	0.250950
0.4333	0.020235	0.0280	0.268220
0.3845	0.024968	0.0268	0.285541
0.3052	0.030584	0.0264	0.303822
0.2659	0.036388	0.0258	0.322431
0.1921	0.042636	0.0257	0.341630
0.1777	0.049220	0.0232	0.361799
0.1767	0.056445	0.0231	0.382292
0.1607	0.064118	0.0220	0.403271
0.1197	0.071902	0.0219	0.424872
0.1185	0.080101	0.0198	0.450274
0.1053	0.088375	0.0188	0.481809
0.0917	0.096898	0.0163	0.514436
0.0898	0.105662	0.0149	0.550082
0.0658	0.115446	0.0132	0.591342
0.0600	0.125364	0.0129	0.640608
0.0519	0.135529	0.0124	0.694320
0.0493	0.145923	0.0070	0.809258
0.0473	0.156598	0.0055	0.964455
0.0431	0.167861	0.0052	1.129894
0.0421	0.179836	0.0044	1.333051
0.0405	0.192127	...	...

**Table 24**  
R50- $\beta$  Model

Statistics			
$R^2$	$R_{\text{adj}}^2$	rms	$C$
0.36	0.30	0.44	0.84
Model Parameters			
Variable	$b_i$	$\bar{x}_i$	
$M_{1500}$	-0.196	-19.84	
$\log_{10}(M_*/M_{\odot})$	0.361	8.932	
$\beta_{1550}$	-2.766	-1.810	
$\beta_{1550}$	-4.472	-0.210	
Baseline Cumulative Hazard			
$f_{\text{esc}}$	$H_0(f_{\text{esc}})$	$f_{\text{esc}}(\text{cont.})$	$H_0(\text{cont.})$
0.8890	0.003630	0.0376	0.196594
0.6247	0.007542	0.0333	0.209834
0.5838	0.011518	0.0309	0.223769
0.4911	0.015572	0.0307	0.237960
0.4333	0.019859	0.0280	0.253960
0.3845	0.024345	0.0268	0.270600
0.3052	0.029476	0.0264	0.288151
0.2659	0.034731	0.0258	0.306004
0.1921	0.040111	0.0257	0.323881
0.1777	0.045787	0.0232	0.342762
0.1767	0.052227	0.0231	0.362072
0.1607	0.059035	0.0220	0.382566
0.1197	0.066366	0.0219	0.403879
0.1185	0.074039	0.0198	0.431991
0.1053	0.081864	0.0188	0.463839
0.0917	0.089852	0.0163	0.497097
0.0898	0.098085	0.0149	0.533738
0.0658	0.107216	0.0132	0.583785
0.0600	0.116484	0.0129	0.645270
0.0519	0.126573	0.0124	0.709809
0.0493	0.137188	0.0070	0.848036
0.0473	0.147942	0.0055	1.054349
0.0431	0.159479	0.0052	1.269787
0.0421	0.171621	0.0044	1.562856
0.0405	0.184018	...	...

**Table 25**  
 $\beta$ -Metals Model

Statistics			
$R^2$	$R_{\text{adj}}^2$	rms	$C$
0.11	0.02	0.55	0.77
Model Parameters			
Variable	$b_i$	$\bar{x}_i$	
$M_{1500}$	-7.760e-02	-19.84	
$\log_{10}(M_*/M_{\odot})$	0.397	8.920	
$\beta_{1550}$	-2.227	-1.819	
$12+\log_{10}(\text{O}/\text{H})$	-1.225	8.111	
Baseline Cumulative Hazard			
$f_{\text{esc}}$	$\text{HF}_0(f_{\text{esc}})$	$f_{\text{esc}}(\text{cont.})$	$\text{HF}_0(\text{cont.})$
0.8890	0.006818	0.0376	0.272610
0.6247	0.013970	0.0333	0.288649
0.5838	0.021388	0.0309	0.305054
0.4911	0.029052	0.0307	0.322206
0.4333	0.037085	0.0280	0.341201
0.3845	0.045286	0.0268	0.360368
0.3052	0.054005	0.0264	0.380230
0.2659	0.062894	0.0258	0.400381
0.1921	0.071991	0.0257	0.420683
0.1777	0.081252	0.0232	0.441694
0.1767	0.091036	0.0231	0.463044
0.1607	0.101098	0.0220	0.485047
0.1197	0.111387	0.0219	0.507680
0.1185	0.121976	0.0198	0.533235
0.1053	0.132764	0.0188	0.560019
0.0917	0.143767	0.0163	0.587757
0.0898	0.154960	0.0149	0.618128
0.0658	0.166527	0.0132	0.652861
0.0600	0.178210	0.0129	0.691653
0.0519	0.190433	0.0124	0.732358
0.0493	0.202887	0.0070	0.824911
0.0473	0.215452	0.0055	0.936173
0.0431	0.228888	0.0052	1.051480
0.0421	0.243133	0.0044	1.195328
0.0405	0.257825	...	...



## ORCID iDs

Anne E. Jaskot  <https://orcid.org/0000-0002-6790-5125>  
 Anna Plantinga  <https://orcid.org/0000-0001-6281-7727>  
 Sophia R. Flury  <https://orcid.org/0000-0002-0159-2613>  
 Matthew Hayes  <https://orcid.org/0000-0001-8587-218X>  
 John Chisholm  <https://orcid.org/0000-0002-0302-2577>  
 Timothy Heckman  <https://orcid.org/0000-0001-6670-6370>  
 Laura Pentericci  <https://orcid.org/0000-0001-8940-6768>  
 Daniel Schaerer  <https://orcid.org/0000-0001-7144-7182>  
 Maxime Trebitsch  <https://orcid.org/0000-0002-6849-5375>  
 Anne Verhamme  <https://orcid.org/0000-0002-2201-1865>  
 Cody Carr  <https://orcid.org/0000-0003-4166-2855>  
 Henry C. Ferguson  <https://orcid.org/0000-0001-7113-2738>  
 Zhiyuan Ji  <https://orcid.org/0000-0001-7673-2257>  
 Mauro Giavalisco  <https://orcid.org/0000-0002-7831-8751>  
 Alaina Henry  <https://orcid.org/0000-0002-6586-4446>  
 Rui Marques-Chaves  <https://orcid.org/0000-0001-8442-1846>  
 Göran Östlin  <https://orcid.org/0000-0002-3005-1349>  
 Alberto Saldana-Lopez  <https://orcid.org/0000-0001-8419-3062>  
 Claudia Scarlata  <https://orcid.org/0000-0002-9136-8876>  
 Gábor Worseck  <https://orcid.org/0000-0003-0960-3580>  
 Xinfeng Xu  <https://orcid.org/0000-0002-9217-7051>

## References

- Atek, H., Labbé, I., Furtak, L. J., et al. 2024, *Natur*, 626, 975  
 Backhaus, B. E., Trump, J. R., Pirzkal, N., et al. 2024, *ApJ*, 962, 195  
 Barrow, K. S. S., Robertson, B. E., Ellis, R. S., et al. 2020, *ApJL*, 902, L39  
 Bassett, R., Ryan-Weber, E. V., Cooke, J., et al. 2019, *MNRAS*, 483, 5223  
 Becker, G. D., D'Aloisio, A., Christenson, H. M., et al. 2021, *MNRAS*, 508, 1853  
 Bezanson, R., Labbe, I., Whitaker, K. E., et al. 2022, arXiv:2212.04026  
 Bian, F., & Fan, X. 2020, *MNRAS*, 493, L65  
 Blanton, M. R., Bershady, M. A., Abolfathi, B., et al. 2017, *AJ*, 154, 28  
 Borthakur, S., Heckman, T. M., Leitherer, C., & Overzier, R. A. 2014, *Sci*, 346, 216  
 Bouwens, R. J., Illingworth, G. D., Oesch, P. A., et al. 2015, *ApJ*, 811, 140  
 Bouwens, R. J., Smit, R., Labbé, I., et al. 2016, *ApJ*, 831, 176  
 Bouwens, R. J., Stefanon, M., Brammer, G., et al. 2023, *MNRAS*, 523, 1036  
 Bradburn, M. J., Clark, T. G., Love, S. B., & Altman, D. G. 2003, *Brit. J. Cancer*, 89, 431  
 Bremer, J., & Dayal, P. 2023, *MNRAS*, 524, 118  
 Castellano, M., Pentericci, L., Cupani, G., et al. 2022, *A&A*, 662, A115  
 Cen, R., & Kimm, T. 2015, *ApJL*, 801, L25  
 Chisholm, J., Rigby, J. R., Bayliss, M., et al. 2019, *ApJ*, 882, 182  
 Chisholm, J., Gazagnes, S., Schaerer, D., et al. 2018, *A&A*, 616, A30  
 Chisholm, J., Saldana-Lopez, A., Flury, S., et al. 2022, *MNRAS*, 517, 5104  
 Choustikov, N., Katz, H., Saxena, A., et al. 2024, *MNRAS*, 529, 3751  
 Clark, T. G., Bradburn, M. J., Love, S. B., & Altman, D. G. 2003, *Brit. J. Cancer*, 89, 232  
 Cox, D. R. 1972, *J. R. Stat. Soc. B*, 34, 187  
 Curti, M., D'Eugenio, F., Carniani, S., et al. 2023, *MNRAS*, 518, 425  
 Davidson-Pilon, C. 2019, *JOSS*, 4, 1317  
 de Barros, S., Vanzella, E., Amorín, R., et al. 2016, *A&A*, 585, A51  
 Donnan, C. T., McLeod, D. J., Dunlop, J. S., et al. 2023, *MNRAS*, 518, 6011  
 Dopita, M. A., Fischera, J., Sutherland, R. S., et al. 2006, *ApJS*, 167, 177  
 Duan, Q., Conselice, C. J., Li, Q., et al. 2024, *MNRAS*, 529, 4728  
 Eilers, A.-C., Davies, F. B., & Hennawi, J. F. 2018, *ApJ*, 864, 53  
 Endsley, R., Stark, D. P., Chevallard, J., & Charlot, S. 2021, *MNRAS*, 500, 5229  
 Endsley, R., Stark, D. P., Whitler, L., et al. 2023, *MNRAS*, 524, 2312  
 Fan, X., Carilli, C. L., & Keating, B. 2006, *ARA&A*, 44, 415  
 Feigelson, E. D., & Nelson, P. I. 1985, *ApJ*, 293, 192  
 Finkelstein, S. L., Ryan, R. E., Jr., Papovich, C., et al. 2015, *ApJ*, 810, 71  
 Finkelstein, S. L., D'Aloisio, A., Paardekooper, J.-P., et al. 2019, *ApJ*, 879, 36  
 Finkelstein, S. L., Bagley, M. B., Ferguson, H. C., et al. 2023, *ApJL*, 946, L13  
 Fitzpatrick, E. L. 1999, *PASP*, 111, 63  
 Fletcher, T. J., Tang, M., Robertson, B. E., et al. 2019, *ApJ*, 878, 87  
 Flury, S., Jaskot, A., & Silveyra, A., 2024 LyCsurv, v0.1.0, Zenodo, doi:10.5281/zenodo.11392442  
 Flury, S. R., Jaskot, A. E., Ferguson, H. C., et al. 2022a, *ApJS*, 260, 1  
 Flury, S. R., Jaskot, A. E., Ferguson, H. C., et al. 2022b, *ApJ*, 930, 126  
 Fujimoto, S., Arrabal Haro, P., Dickinson, M., et al. 2023, *ApJL*, 949, L25  
 Gazagnes, S., Chisholm, J., Schaerer, D., et al. 2018, *A&A*, 616, A29  
 Green, G. M., Schlafly, E. F., Finkbeiner, D., et al. 2018, *MNRAS*, 478, 651  
 Hainline, K. N., Shapley, A. E., Kornei, K. A., et al. 2009, *ApJ*, 701, 52  
 Harikane, Y., Ouchi, M., Oguri, M., et al. 2023, *ApJS*, 265, 5  
 Inoue, A. K., Shimizu, I., Iwata, I., & Tanaka, M. 2014, *MNRAS*, 442, 1805  
 Isobe, T., Feigelson, E. D., & Nelson, P. I. 1986, *ApJ*, 306, 490  
 Izotov, Y. I., Orlitová, I., Schaerer, D., et al. 2016a, *Natur*, 529, 178  
 Izotov, Y. I., Schaerer, D., Thuan, T. X., et al. 2016b, *MNRAS*, 461, 3683  
 Izotov, Y. I., Schaerer, D., Worseck, G., et al. 2018a, *MNRAS*, 474, 4514  
 Izotov, Y. I., Worseck, G., Schaerer, D., et al. 2018b, *MNRAS*, 478, 4851  
 Izotov, Y. I., Worseck, G., Schaerer, D., et al. 2021, *MNRAS*, 503, 1734  
 James, B. L., Auger, M., Pettini, M., et al. 2018, *MNRAS*, 476, 1726  
 Jaskot, A. E., Silveyra, A. C., Plantinga, A., et al. 2024, *ApJ*, 972, 92  
 Ji, Z., Giavalisco, M., Vanzella, E., et al. 2020, *ApJ*, 888, 109  
 Johnson, B. D., Leja, J. L., Conroy, C., & Speagle, J. S., 2019 Prospector: Stellar population inference from spectra and SEDs, Astrophysics Source Code Library, ascl:1905.025  
 Jung, I., Finkelstein, S. L., Dickinson, M., et al. 2020, *ApJ*, 904, 144  
 Kakiichi, K., & Gronke, M. 2021, *ApJ*, 908, 30  
 Katz, H., Saxena, A., Rosdahl, J., et al. 2023, *MNRAS*, 518, 270  
 Kennicutt, R. C., & Evans, N. J. 2012, *ARA&A*, 50, 531  
 Kerutt, J., Oesch, P. A., Wisotzki, L., et al. 2024, *A&A*, 684, A42  
 Kimm, T., Blaizot, J., Garel, T., et al. 2019, *MNRAS*, 486, 2215  
 Larson, R. L., Finkelstein, S. L., Kocevski, D. D., et al. 2023, *ApJL*, 953, L29  
 Leitet, E., Bergvall, N., Hayes, M., Linné, S., & Zackrisson, E. 2013, *A&A*, 553, A106  
 Leitet, E., Bergvall, N., Piskunov, N., & Andersson, B. G. 2011, *A&A*, 532, A107  
 Leitherer, C., Ekström, S., Meynet, G., et al. 2014, *ApJS*, 212, 14  
 Leitherer, C., Hernandez, S., Lee, J. C., & Oey, M. S. 2016, *ApJ*, 823, 64  
 Leitherer, C., Schaerer, D., Goldader, J., et al., 2011 Starburst99: Synthesis Models for Galaxies with Active Star Formation, Astrophysics Source Code Library, ascl:1104.003  
 Leja, J., Johnson, B. D., Conroy, C., van Dokkum, P. G., & Byler, N. 2017, *ApJ*, 837, 170  
 Lin, Y.-H., Scarlata, C., Williams, H., et al. 2024, *MNRAS*, 527, 4173  
 Liu, Y., Jiang, L., Windhorst, R. A., Guo, Y., & Zheng, Z.-Y. 2023, *ApJ*, 958, 22  
 Luridiana, V., Morisset, C., & Shaw, R. A. 2015, *A&A*, 573, A42  
 Ma, X., Kasen, D., Hopkins, P. F., et al. 2015, *MNRAS*, 453, 960  
 Maji, M., Verhamme, A., Rosdahl, J., et al. 2022, *A&A*, 663, A66  
 Makan, K., Worseck, G., Davies, F. B., et al. 2021, *ApJ*, 912, 38  
 Marchi, F., Pentericci, L., Guaita, L., et al. 2018, *A&A*, 614, A11  
 Marques-Chaves, R., Schaerer, D., Álvarez-Márquez, J., et al. 2021, *MNRAS*, 507, 524  
 Marques-Chaves, R., Schaerer, D., Álvarez-Márquez, J., et al. 2022, *MNRAS*, 517, 2972  
 Marques-Chaves, R., Schaerer, D., Kuruvanathi, A., et al. 2024, *A&A*, 681, A30  
 Martin, C., Barlow, T., Barnhart, W., et al. 2003, *Proc. SPIE*, 4854, 336  
 Mascia, S., Pentericci, L., Calabrò, A., et al. 2023, *A&A*, 672, A155  
 Mascia, S., Pentericci, L., Calabrò, A., et al. 2024, *A&A*, 685, A3  
 Maseda, M. V., Bacon, R., Lam, D., et al. 2020, *MNRAS*, 493, 5120  
 McGreer, I. D., Mesinger, A., & D'Odorico, V. 2015, *MNRAS*, 447, 499  
 McLeod, D. J., Donnan, C. T., McLure, R. J., et al. 2024, *MNRAS*, 527, 5004  
 McLernon, D. J., Giardiello, D., Van Calster, B., et al. 2023, *Ann. Intern. Med.*, 176, 105  
 Morishita, T., Stiavelli, M., Chary, R.-R., et al. 2024, *ApJ*, 963, 9  
 Mostardi, R. E., Shapley, A. E., Steidel, C. C., et al. 2015, *ApJ*, 810, 107  
 Nagao, T., Maiolino, R., & Marconi, A. 2006, *A&A*, 459, 85  
 Naidu, R. P., Tacchella, S., Mason, C. A., et al. 2020, *ApJ*, 892, 109  
 Naidu, R. P., Matthee, J., Oesch, P. A., et al. 2022, *MNRAS*, 510, 4582  
 Nakajima, K., Ellis, R. S., Robertson, B. E., Tang, M., & Stark, D. P. 2020, *ApJ*, 889, 161  
 Nakajima, K., & Ouchi, M. 2014, *MNRAS*, 442, 900  
 Ormerod, K., Conselice, C. J., Adams, N. J., et al. 2024, *MNRAS*, 527, 6110  
 Paardekooper, J.-P., Khochfar, S., & Dalla Vecchia, C. 2015, *MNRAS*, 451, 2544  
 Pahl, A. J., Shapley, A., Steidel, C. C., Chen, Y., & Reddy, N. A. 2021, *MNRAS*, 505, 2447

- Pilyugin, L. S., Thuan, T. X., & Vílchez, J. M. 2006, *MNRAS*, 367, 1139
- Planck Collaboration, Aghanim, N., Akrami, Y., et al. 2020, *A&A*, 641, A6
- Rivera-Thorsen, T. E., Dahle, H., Chisholm, J., et al. 2019, *Sci*, 366, 738
- Robertson, B. E. 2022, *ARA&A*, 60, 121
- Robertson, B. E., Ellis, R. S., Furlanetto, S. R., & Dunlop, J. S. 2015, *ApJL*, 802, L19
- Robertson, B. E., Tacchella, S., Johnson, B. D., et al. 2023, *NatAs*, 7, 611
- Rudie, G. C., Steidel, C. C., Shapley, A. E., & Pettini, M. 2013, *ApJ*, 769, 146
- Saldana-Lopez, A., Schaerer, D., Chisholm, J., et al. 2022, *A&A*, 663, A59
- Saldana-Lopez, A., Schaerer, D., Chisholm, J., et al. 2023, *MNRAS*, 522, 6295
- Sanders, R. L., Shapley, A. E., Topping, M. W., Reddy, N. A., & Brammer, G. B. 2023, *ApJ*, 955, 54
- Saxena, A., Pentericci, L., Ellis, R. S., et al. 2022, *MNRAS*, 511, 120
- Saxena, A., Bunker, A. J., Jones, G. C., et al. 2024, *A&A*, 684, A84
- Saxena, A., Robertson, B. E., Bunker, A. J., et al. 2023, *A&A*, 678, A68
- Schaerer, D., Izotov, Y. I., Verhamme, A., et al. 2016, *A&A*, 591, L8
- Schaerer, D., Marques-Chaves, R., Barrufet, L., et al. 2022, *A&A*, 665, L4
- Shapley, A. E., Steidel, C. C., Strom, A. L., et al. 2016, *ApJL*, 826, L24
- Simmonds, C., Tacchella, S., Hainline, K., et al. 2024, *MNRAS*, 527, 6139
- Steidel, C. C., Bogosavljević, M., Shapley, A. E., et al. 2018, *ApJ*, 869, 123
- Storey, P. J., & Hummer, D. G. 1995, *MNRAS*, 272, 41
- Tang, M., Stark, D. P., Chevallard, J., & Charlot, S. 2019, *MNRAS*, 489, 2572
- Tang, M., Stark, D. P., Chen, Z., et al. 2023, *MNRAS*, 526, 1657
- Trebitsch, M., Blaizot, J., Rosdahl, J., Devriendt, J., & Slyz, A. 2017, *MNRAS*, 470, 224
- Treu, T., Roberts-Borsani, G., Bradac, M., et al. 2022, *ApJ*, 935, 110
- Vanzella, E., de Barros, S., Vasei, K., et al. 2016, *ApJ*, 825, 41
- Vanzella, E., Nonino, M., Cupani, G., et al. 2018, *MNRAS*, 476, L15
- Vanzella, E., Caminha, G. B., Calura, F., et al. 2020, *MNRAS*, 491, 1093
- Vasei, K., Siana, B., Shapley, A. E., et al. 2016, *ApJ*, 831, 38
- Verhamme, A., Orlitová, I., Schaerer, D., et al. 2017, *A&A*, 597, A13
- Wang, B., Heckman, T. M., Leitherer, C., et al. 2019, *ApJ*, 885, 57
- Wang, B., Heckman, T. M., Amorín, R., et al. 2021, *ApJ*, 916, 3
- Weaver, J. R., Cutler, S. E., Pan, R., et al. 2024, *ApJS*, 270, 7
- Williams, H., Kelly, P. L., Chen, W., et al. 2023, *Sci*, 380, 416
- Wise, J. H., & Cen, R. 2009, *ApJ*, 693, 984
- Witstok, J., Smit, R., Saxena, A., et al. 2024, *A&A*, 682, A40
- Worseck, G., Prochaska, J. X., Hennawi, J. F., & McQuinn, M. 2016, *ApJ*, 825, 144
- Xu, X., Henry, A., Heckman, T., et al. 2023, *ApJ*, 943, 94
- Zackrisson, E., Inoue, A. K., & Jensen, H. 2013, *ApJ*, 777, 39