Semi-analytic forecasts for JWST -- VI. Simulated lightcones and galaxy clustering predictions

Semi-analytic forecasts for $JWST$ – VI. Simulated lightcones and galaxy clustering predictions

L. Y. Aaron Yung,[1][*][†] Rachel S. Somerville,[2] Henry C. Ferguson,[3]
Steven L. Finkelstein,[4] Jonathan P. Gardner,[1] Romeel Davé,[5,6,7] Micaela B. Bagley,[4]
Gergö Popping[8] and Peter Behroozi[9]

[1] *Astrophysics Science Division, NASA Goddard Space Flight Center, 8800 Greenbelt Rd, Greenbelt, MD 20771, USA*
[2] *Center for Computational Astrophysics, Flatiron Institute, 162 5th Ave, New York, NY 10010, USA*
[3] *Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218, USA*
[4] *Department of Astronomy, The University of Texas at Austin, Austin, TX 78712, USA*
[5] *Institute for Astronomy, University of Edinburgh, Edinburgh EH9 3HJ, UK*
[6] *Department of Physics and Astronomy, University of the Western Cape, Cape Town 7535, South Africa*
[7] *South African Astronomical Observatory, Cape Town 7925, South Africa*
[8] *European Southern Observatory, Karl-Schwarzschild-Strasse 2, D-85748 Garching, Germany*
[9] *Department of Astronomy, University of Arizona, 933 N Cherry Ave, Tucson, AZ 85721, USA*

ABSTRACT

In anticipation of the new era of high-redshift exploration marked by the commissioning of the *James Webb Space Telescope* (*JWST*), we present two sets of galaxy catalogues that are designed to aid the planning and interpretation of observing programs. We provide a set of 40 wide-field lightcones with footprints spanning approximately $\sim 1,000$ arcmin^2 containing galaxies up to $z = 10$, and a new set of 8 ultra-deep lightcones with 132 arcmin^2 footprints, containing galaxies up to $z \sim 12$ down to the magnitudes expected to be reached in the deepest *JWST* surveys. These mock lightcones are extracted from dissipationless N-body simulations and populated with galaxies using the well-established, computationally efficient Santa Cruz semi-analytic model for galaxy formation. We provide a wide range of predicted physical properties, and simulated photometry from NIRCam and many other instruments. We explore the predicted counts and luminosity functions and angular two-point correlation functions for galaxies in these simulated lightcones. We also explore the predicted field-to-field variance using multiple lightcone realizations. We find that these lightcones reproduce the available measurements of observed clustering from $0.2 \lesssim z \lesssim 7.5$ very well. We provide predictions for galaxy clustering at high redshift that may be obtained from future *JWST* observations. All of the lightcones presented here are made available through a web-based, interactive data release portal https://flathub.flatironinstitute.org/group/sam-forecasts.

Key words: galaxies: evolution – galaxies: formation – galaxies: high-redshifts – galaxies: star formation – astronomical data base: surveys

1 INTRODUCTION

The launch, deployment, and commissioning of the *James Webb Space Telescope* (*JWST*; Gardner et al. 2006) marks the beginning of a new era of exploration of the physical origin of our Universe. Its onboard Near-Infrared Camera (NIRCam) possesses unprecedented infrared sensitivity and spatial resolution, which will enable it to image distant, faint objects with great efficiency. It is expected to detect distant galaxies with luminosities many magnitudes fainter than those discovered by its predecessors the *Hubble* and *Spitzer Space Telescopes*. The Mid-Infrared Instrument (MIRI) has both imaging and spectroscopic capabilities at longer wavelength mid-IR surpassing *Spitzer* and *Herschel* in both spatial resolution and spectral resolution. The Near-Infrared Imager and Slitless Spectrograph (NIRISS) is capable of effectively obtaining low-resolution spectra of multiple objects at a time, while the Near-Infrared Spectrograph (NIRSpec)

[*] E-mail: aaron.yung@nasa.gov
[†] NPP Fellow

will enable high spectral resolution diagnostics with greater sensitivity. These instruments are able to work in parallel to maximize the science yield.

The combination of the large primary mirror and unprecedented infrared sensitivity of *JWST's* scientific instruments will enable the detection of galaxies forming during or prior to the Epoch of Reionization (EoR), providing important information that will help complete our picture of the origins of galaxies and the sources of reionization. The detection of galaxies of faint apparent magnitude across all epochs of cosmic history will deliver new constraints on the evolution of galaxy physical properties and number densities across cosmic time, as well as insights on the physical processes that shaped these objects. We also anticipate that *JWST* will, for the first time in history, definitively detect and deliver direct constraints on populations of galaxies forming at $z > 10$, which will provide new tests of galaxy formation models and theories.

Although *JWST* was not designed to survey large areas to deep levels, with the anticipated potential long lifetime that the facility may enjoy, we also anticipate that *JWST* will be able to make important contributions to measuring the spatial clustering of these faint, high redshift galaxies. This can provide an important counterpart to the wider, shallower clustering measurements that will be enabled by ground based facilities and the Euclid and Nancy Grace Roman space telescopes. Galaxy clustering can provide important constraints on the distribution of the underlying dark matter density field, as well as on how galaxies of different types populate halos of different mass (halo occupation).

In order to leverage the rich data accumulated from past observations, *JWST* is expected to revisit the five relatively well-surveyed legacy fields from the Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey (CANDELS; Grogin et al. 2011; Koekemoer et al. 2011). These legacy fields include the Extended Groth Strip (EGS) from the All-wavelength Extended Groth strip International Survey (AEGIS; Davis et al. 2007), the north and south fields of the Great Observatories Origins Deep Survey (GOODS-N and GOODS-S; Giavalisco et al. 2004), and the Cosmic Evolution Survey (COSMOS; Scoville et al. 2007), and the UKIDSS Ultra Deep Survey (UDS; Cirasuolo et al. 2007). In addition, there is also the *Hubble* Ultra Deep Field (HUDF; Ellis et al. 2013; Koekemoer et al. 2013) and the eXtreme Deep Field (XDF; Illingworth et al. 2013), located within GOODS-S.

Within the first year of operation, *JWST* will execute a number of Early Release Science and Guaranteed Time (GTO) deep extragalactic survey programs. The Cosmic Evolution Early Release Science (CEERS[1]) Survey will cover ~ 100 square arcminutes within the EGS field (Finkelstein et al. 2017) with *JWST* imaging and spectroscopy using NIRCam, MIRI, and NIRSpec. The joint NIRCam–NIRSpec GTO *JWST* Advanced Deep Extragalactic Survey (JADES; Williams et al. 2018) will observe the GOODS-S and HUDF field. In addition, the Cycle 1 General Observer (GO) program Next Generation Deep Extragalactic Exploratory Public (NGDEEP) will dedicate 120 hours to observing the HUDF with NIRISS and imaging the HUDF Parallel 2 field

with NIRCam (Finkelstein et al. 2021). The Public Release IMaging for Extragalactic Research (PRIMER; Dunlop et al. 2021) and COSMOS–Web (Kartaltepe et al. 2021) will cover over 390 square arcminutes (over COSMOS and UDS) and 2100 square arcminutes (over COSMOS), respectively. We refer the reader to Robertson (2022) for a comprehensive review of *JWST* approved galaxy evolution programs and the science that is expected to be enabled by them.

Given *JWST's* limited lifespan and high oversubscription rates for time allocation, optimization of observational strategies including exposure time, pointings, survey strategy, as well as parallel observing, are crucial to maximize the science return. Detailed mock catalogues and synthetic data can help support the planning and optimization of these observational strategies.

Lightcones are a form of mock galaxy catalogues that mimic the way we observe the real Universe, providing galaxy locations on the plane of the sky (right ascension and declination) as well as redshifts along the line of sight (including both cosmological redshifting and peculiar velocities) along a past lightcone. They more realistically include the continuous sampling of galaxies at different redshifts along the line of sight in a real survey, along with the accompanying impact on the k-corrections of the SED, etc, than using a single redshift slice from a box, as is commonly done for simulations. The output from lightcones is also useful as an input to create synthetic images or spectroscopy. It is straightforward to include observational effects such as photometric noise, a point spread function, and observationally based selection criteria (Torrey et al. 2015; Rodriguez-Gomez et al. 2019; Snyder et al. 2019; Marshall et al. 2022). Thus simulated lightcones are an important tool for both *forecasting* and interpreting high-redshift galaxy surveys. In the past, mock lightcones have been created to aid the theoretical interpretation of the CANDELS surveys (Behroozi et al. 2019; Somerville et al. 2021). The lightcones presented in this work have been used in the interpretation of galaxies observed at high redshifts (Finkelstein et al. 2022; Tacchella et al. 2022) with *Hubble* and *Spitzer*, as well as in the planning for upcoming CEERS and NGDEEP surveys. Other *JWST* mock catalogues have also been made available, such as those based on UniverseMachine (Behroozi et al. 2019, 2020), the *JWST* Extragalactic Mock Catalogue (Williams et al. 2018), and the Deep Realistic Extragalactic Model (DREaM) (Drakos et al. 2022).

There are several different methods used in the literature for creating mock catalogues and lightcones. In purely empirical methods such as the JAGUAR models used to create mock catalogues in support of the JADES survey (Williams et al. 2018), observed galaxy properties are interpolated or extrapolated, and there is no underlying physics model nor setting within a ΛCDM context. In what are sometimes called 'semi-empirical' methods (also called subhalo abundance matching (SHAM) or halo occupation distribution (HOD) models; see Wechsler & Tinker 2018), galaxy properties are mapped onto the properties of dark matter halos such that a set of observational quantities is reproduced (Behroozi et al. 2010, 2019; Moster et al. 2013, 2018; Wechsler et al. 2022). Both of these methods have the advantage that they are computationally efficient, are not dependent on a specific model for galaxy formation, and are guaranteed to match the observations that were used to calibrate them.

However, they have the disadvantage that using them for *forecasts* for new observations is highly uncertain, and they are of limited use for interpretation. Semi-empirical models are typically calibrated using derived physical properties such as stellar masses and star formation rates, which are highly uncertain at high redshifts, leading to models that are nominally calibrated on the same observations, but which have very different predictions for the link between galaxy and dark matter halo properties (see e.g. Yung et al. 2019b). We note that this is only a general overview for the semi-empirical modelling approach. These models are designed with different purposes in mind and adopt different calibration criteria. For example, UNIVERSEMACHINE used high-redshift UV luminosity functions to calibrate galaxy growth at the highest redshifts (Behroozi et al. 2019, 2020) and the semi-empirical model presented in Behroozi & Silk (2015) allows forecasts under the assumption that galaxy-halo growth relationships are given by power laws.

Lightcones can also be extracted from numerical hydrodynamic simulations (e.g. Snyder et al. 2017), but due to the computational expense of the underlying simulations, these tend to have limited volume and are able to provide only a small number of realizations. Moreover, all cosmological hydrodynamic simulations must adopt uncertain 'sub-grid' recipes to describe processes that cannot be explicitly resolved, such as star formation, stellar feedback, and black hole seeding, accretion, and feedback (Somerville & Davé 2015).

In this work we adopt the well-established 'middle way' for modelling galaxy formation using semi-analytic models. Semi-analytic models are built on the backbone of merger trees based on the ΛCDM structure formation paradigm, either extracted from dissipationless N-body simulations or created using Monte Carlo methods. They include simplified but physically motivated treatments of the main processes that are thought to shape galaxy evolution, namely gas accretion from the intergalactic medium (IGM) into the circumgalactic medium (CGM), gas cooling from the CGM into the interstellar medium (ISM), formation of stars from dense gas in the ISM, stellar feedback and chemical enrichment from massive stars and supernovae, and black hole growth and feedback. These recipes contain free parameters, that are calibrated to match a subset of observations at $z = 0$. Although SAMs can be run in a fraction of the computational cost of numerical hydrodynamic simulations, their predictions for many observable galaxy properties have been shown to be in excellent agreement with those of the numerical simulations (Somerville & Davé 2015; Pandya et al. 2020; Gabrielpillai et al. 2021). In particular, the predictions of the Santa Cruz SAM used here have been shown to be in excellent agreement with the predicted rest-UV luminosity functions, stellar mass functions and SFR functions from numerical hydrodynamic simulations over the redshift range $6 \lesssim z \lesssim 10$ (Yung et al. 2019b; Lovell et al. 2020; Vogelsberger et al. 2020).

This is the concluding paper in the *Semi-analytic forecasts for JWST* series. This series is a collection of papers that provides in-depth, comprehensive predictions for the properties and demographics of galaxies and AGN forming in the ultrahigh-redshift Universe. The modelling pipeline used for the series is based on the versatile Santa Cruz semi-analytic models, with additional model components inserted

to expand the model's capabilities in making predictions during the EoR. In Yung et al. (2019a,b, hereafter Paper I and Paper II), we presented predicted photometric and physical properties of high-redshift galaxy populations, which are compared extensively against existing observations and other simulations. In Yung et al. (2020a, hereafter Paper III), we calculated the ionizing photon production rate based on the star formation and chemical enrichment history of individual galaxies, using both single star and binary stellar population models. In Yung et al. (2020b, hereafter Paper IV), we combined the SAM-based source model with an analytic reionization model to compute the cosmic reionization history of intergalactic hydrogen and compare with available observational constraints. In Yung et al. (2021, hereafter Paper V), we computed the contribution of AGN to hydrogen and helium reionization. In this final paper of the series (Paper VI), we present mock lightcones with area and resolution chosen to be similar to the footprint and depth of upcoming *JWST* ERS, GTO, and Cycle 1 observing programs, which should be of general utility for planning and interpretation of future *JWST* programs as well. In addition to galaxy physical properties, we include an extensive suite of photometry for *JWST* instruments as well as legacy *HST*, *Spitzer*, and ground based filters. These include lightcones with the same mass resolution and characteristics as the mock CANDELS fields presented in Somerville et al. (2021), as well as a set of new ultra-deep lightcones with smaller footprints, comparable to deeper, smaller area surveys such as NGDEEP. This work is complementary to the companion paper *Semi-analytic forecasts for Roman* paper (Yung et al. 2022), which presents a suite of 2-deg^2 lightcones with depths comparable to the wide-field lightcones from this work.

We show the predicted luminosity functions and counts from these lightcones in comparison with the predictions from our previous work. We also show a significant new science result enabled by the lightcones. We compute projected two point correlation functions in broad redshift bins, and compare with observational measurements from $0.2 \lesssim z \lesssim 7.5$. Encouraged by the excellent agreement that we find, we present predictions for clustering measurements that may be obtained with future *JWST* observations.

All mock catalogues presented in this work series are accessible through the data release portal https://www.simonsfoundation.org/semi-analytic-forecasts-for-jwst/ and the Flatiron Institute Data Exploration and Comparison Hub (Flathub; http://flathub.flatironinstitute.org/SAM_Forecasts/).

The structure of this paper is summarized as follows: the galaxy formation model and lightcone construction pipeline are summarized briefly in Section 2. An overview and description of the simulated lightcones are presented in Section 3. We present the clustering statistics and other main scientific results in Section 4. We discuss our findings in Section 5, and a summary and conclusions follow in Section 6.

2 LIGHTCONE CONSTRUCTION PIPELINE WITH PHYSICAL MODELS

In this section, we provide a concise summary of the semi-analytic model (SAM) for galaxy formation developed by the Santa Cruz group (Somerville & Primack 1999; Somerville
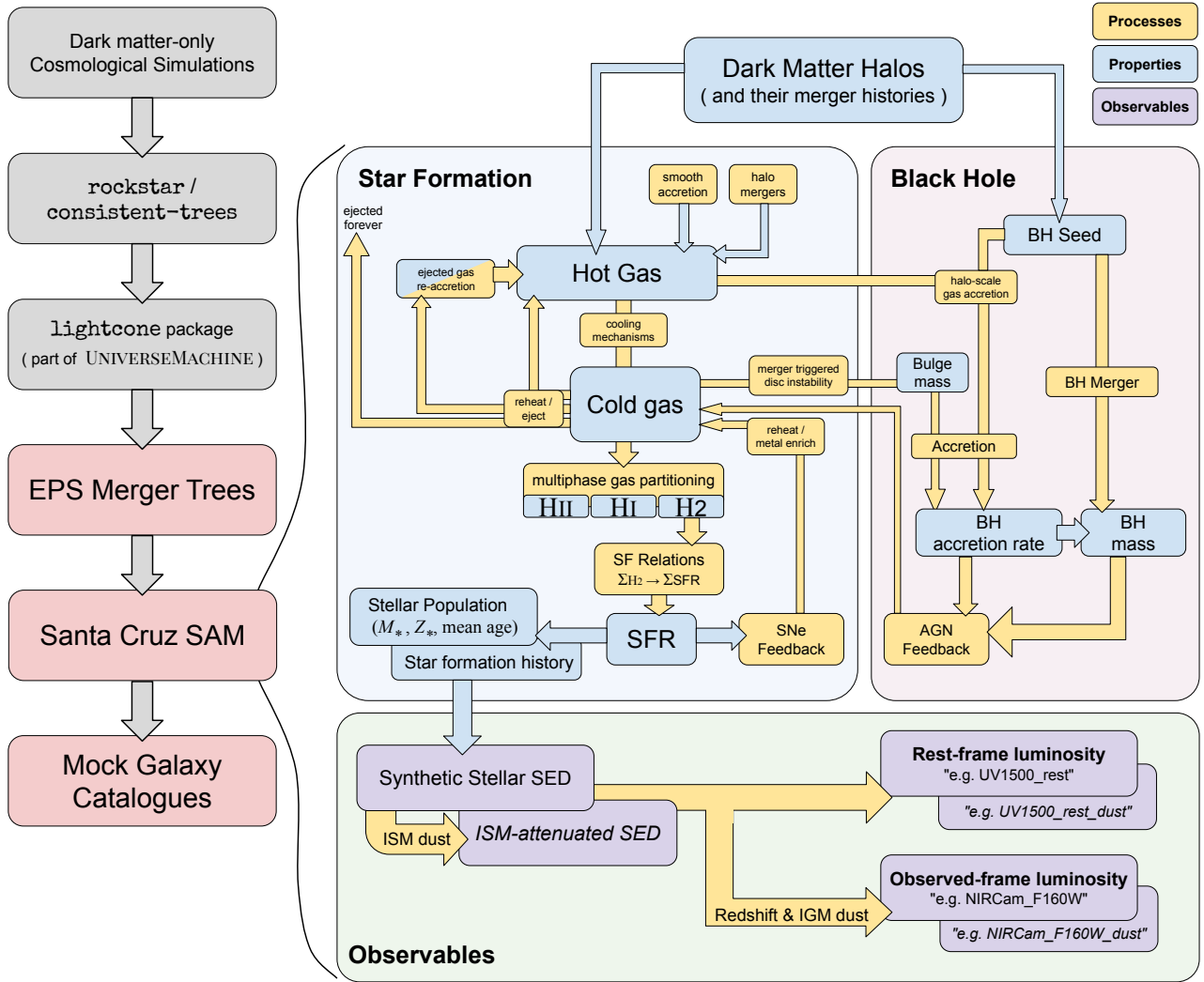
Figure 1. A schematic diagram illustrating the full pipeline of the creation of mock galaxy catalogues (*left column*) and the internal workflow of the Santa Cruz SAM (*right*). On the left, we show the steps for lightcone construction from dark matter-only cosmological simulations to dark matter halo catalogues, labelled in grey. The semi-analytic modelling components that are responsible for creating the galaxy catalogues are labelled in red. On the right, we show a subset of internal processes in the Santa Cruz SAM that are most relevant to the predictions we present in this work. These processes are broken down into three groups: star formation related, black hole related, and observables. The physical processes are labelled with yellow boxes, blue boxes are for physical properties, and purple boxes are observable properties. The quantities with blue and purple labels are available in the mock galaxy catalogues.

et al. 2008, 2012, 2015; Popping et al. 2014). We refer the reader to these papers for a full description of the model components. The specific modelling and parameter choices for galaxy and AGN formation are documented in Paper I and Paper V. Free parameters in these models are calibrated as described in Paper I and Somerville et al. (2021). Throughout this work, as well as the rest of the paper series, we adopt cosmological parameters $\Omega_m = 0.308$, $\Omega_\Lambda = 0.692$, $H_0 = 67.8$ km s^{-1}Mpc^{-1}, $\sigma_8 = 0.831$, and $n_s = 0.9665$; which are broadly consistent with the ones reported by the Planck Collaboration in 2015 (Planck Collaboration XIII 2016). Throughout this work, all magnitudes are expressed in the AB system (Oke & Gunn 1983) and all uses of log are base 10 unless otherwise specified. The calculations presented in this work make use of ASTROPY (Robitaille et al. 2013; Price-Whelan et al. 2018), NUMPY (van der Walt et al.

2011), SCIPY (Virtanen et al. 2020), and pandas (Reback et al. 2022).

2.1 Physical components of the galaxy formation model

The SAM includes a fairly comprehensive set of physical processes, including gas cooling and accretion, star formation, stellar feedback, chemical evolution, black hole growth, and AGN feedback. In this work, we adopt the fiducial 'GK–Big2' model configuration that is consistent with the rest of the *Semi-analytic forecasts for JWST* series papers. This configuration includes a multi-phase gas partitioning recipe motivated by numerical simulation results from Gnedin & Kravtsov (2011, hereafter GK), which partitions the galactic disc into a neutral, ionized, and molecular component, and an observationally-motivated H_2-based star formation recipe from Bigiel et al. (2008, hereafter Big), where the sur-
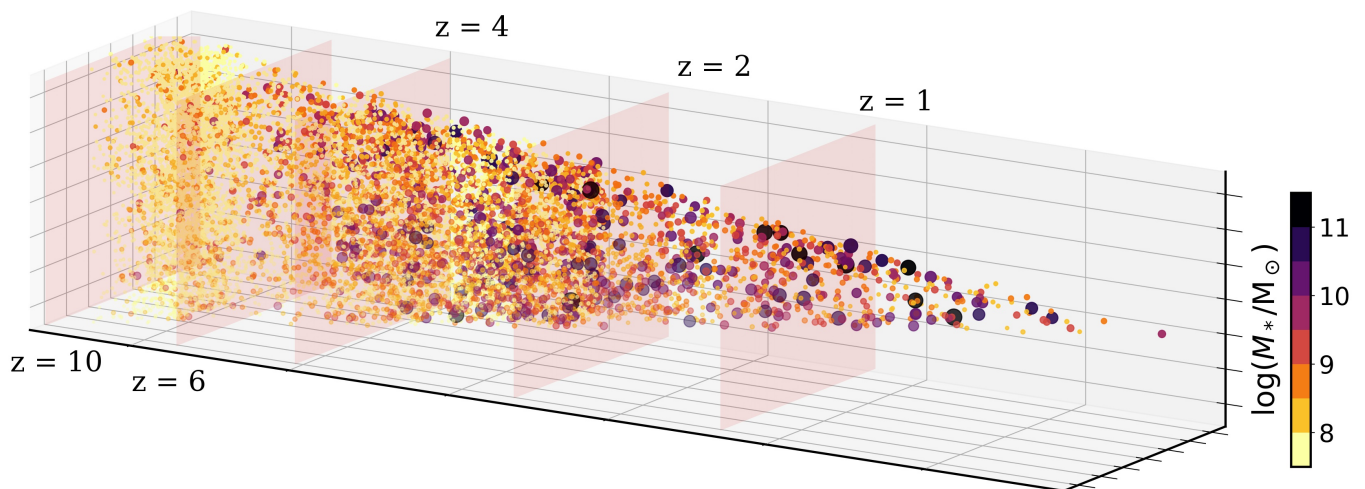
Figure 2. This is a three-dimensional visualization of a simulated lightcone of the COSMOS field. The cone-shaped feature is a manifestation of the predicted physical positions and distances for objects expected within the survey area along our line-of-sight (viewing from the right edge of the figure). Simulated lightcones of this sort offer a new perspective to understand and visualize the galaxies and quasars detected in deep-field surveys. This plot only includes central galaxies. Galaxies in this plot are both colour-coded and size-coded by their stellar mass. To aid visibility in this plot, this plot only contains 1% of objects with $\log(M_*/M_\odot) \geq 8.0$ at $z \leq 2$, 2% of objects with $\log(M_*/M_\odot) \geq 7.5$ between $2 < z < 7$, and 25% of objects with $\log(M_*/M_\odot) \geq 7.0$ at $z \geq 7$ are uniformly sampled from the simulated lightcone.

face density of SFR, Σ_{SFR}, scales with the surface density of molecular hydrogen, Σ_{H_2}. Here we adopt a double power law SF relation, where the slope of the relation steepens above a critical density (see Somerville et al. 2015, for details). Stellar feedback is modelled by ejecting cold gas from the Interstellar Medium (ISM) at a rate proportional to the SFR, according to a phenomenological power-law function of the central halo circular velocity. Explorations of the dependence of the model predictions on the parameters of the star formation and stellar feedback recipes are presented in Yung et al. (2019a) and Yung et al. (2019b). In Fig. 1, we show a schematic diagram that summarizes the internal workflow of the Santa Cruz SAM.

The free parameters in the galaxy model are calibrated to reproduce a variety of $z \sim 0$ observations, including stellar-to-halo mass ratio, stellar mass function, M_{BH}–M_{bulge} relation, stellar metallicity, and cold gas fraction (Paper I and S21). The model outputs of the Santa Cruz SAM at low and intermediate redshift (e.g. $z \lesssim 6$) have been tested against observations in many works, including SPT15 and S21, and at high redshift (e.g. $z \gtrsim 4$) in Paper I and Paper II. Note that in this paper series, as in all previous work with the Santa Cruz SAMs, the models have not been tuned to match observations at high redshift.

When a halo merges with another halo, the galaxy within the smaller halo becomes a 'satellite' galaxy. The Santa Cruz SAM includes a semi-analytic treatment of the decay of satellite orbits due to dynamical friction, and of tidal stripping and destruction of satellites (see Somerville et al. 2008, for details). The radius of the satellite galaxy relative to the centre of the halo is tracked as the orbit decays; when a satellite reaches the centre it is merged with the central galaxy. In some cases, satellite galaxies can become completely tidally disrupted and destroyed. In this case their

stars are added to a diffuse stellar halo, and their cold gas is added to the hot gas reservoir of the central halo. This model was calibrated to match the sub-halo conditional mass function from the Bolshoi Planck N-body simulation (Klypin et al. 2016), but it does not predict the correct radial distribution of satellites within their host halos compared with Bolshoi Planck. Moreover, we found in this work that, when using the 'native' satellite positions that are included in the published version of the SC SAM, we do not reproduce the observed clustering on small scales. We therefore reassign the satellite positions in post-processing as follows.

The positions of satellite galaxies are assigned in post-processing with an approach similar to the one detailed in Kakos et al. (2022), which assumes that the satellites follow the same radial profile as the underlying dark matter halo, as described by an NFW model (Navarro, Frenk & White 1997). The halo profile can be expressed in terms of the halo's virial mass as $M_h = M_{vir} \times u_{vir}(r)$, where

$$u_{vir} = \frac{f(c_{NFW}(r/R_{vir}))}{f(c_{NFW})} \tag{1}$$

and $f \equiv \ln(1+x) - x/1+x$. Here c_{NFW} and R_{vir} are the NFW halo concentration parameter and virial radius of the (host) halo, respectively, which are obtained from the fits provided by ROCKSTAR (see 2.3 for details). For each satellite galaxy, a random number U_r, which takes values between 0 and 1, is assigned, and the distance r from the centre of the host halo is determined by solving $U_r - u_{vir}(r) = 0$. Similarly, every satellite galaxy is assigned a random polar angle, $\theta \in [0, \pi)$, and azimuthal angle, $\phi \in [0, 2\pi)$. The position of a satellite relative to the halo centre in Cartesian coordinates,

X'_i, is given by

$$X'_1 = r \sin(\theta) \, \cos(\phi) \tag{2}$$

$$X'_2 = r \sin(\theta) \, \sin(\phi) \tag{3}$$

$$X'_3 = r \cos(\theta). \tag{4}$$

These relative coordinates are added to their host halo position relative to lightcone axes X_i to obtain their positions in the lightcone $X_{i,\mathrm{lc}} = X_i + X'_i$. The physical positions on the sky $X_{\mathrm{sky}} = R(X_{\mathrm{lc}} - 0)$, where R is the rotation matrix specific to the lightcone, are then converted to sky coordinates with

$$\mathrm{RA} = \tan^{-1}\left(\frac{-X_{2,\mathrm{sky}}}{X_{1,\mathrm{sky}}}\right)\frac{180\,\mathrm{deg}}{\pi} \tag{5}$$

$$\mathrm{Dec} = \sin^{-1}\left(\frac{X_{3,\mathrm{sky}}}{||X||}\right)\frac{180\,\mathrm{deg}}{\pi}, \tag{6}$$

where $||X|| = (X_{1,\mathrm{sky}}^2 + X_{2,\mathrm{sky}}^2 + X_{3,\mathrm{sky}}^2)^{-2}$.

2.2 Forward modelling to *JWST* observables

Based on the predicted star formation and chemical enrichment histories (SFHs, stored mass in bins of stellar age and metallicity), galaxies are assigned spectral energy distributions (SEDs) generated based on the results from the stellar population synthesis (SPS) model by Bruzual & Charlot (2003). The rest-frame SEDs are used to calculate rest-frame luminosities in filter bands as presented in the mock catalogue. In addition, quantities labelled with dust are calculated accounting for the effect of dust in the ISM. We assume a dust attenuation curve for starburst galaxies by Calzetti et al. (2000). The V-band dust attenuation is calculated based on the surface density of cold gas and metallicity, based on a 'slab' model as described in Somerville et al. (2012), but adopting the latest recalibration of ISM dust optical depth presented in Paper V. This update improves the agreement between UV LF predictions and observations at $z > 6$ compared to CANDELS DR1.

The rest-frame SEDs are then redshifted according to their redshift in the lightcone, and observed-frame magnitudes are computed, accounting for attenuation effects in the intervening IGM (Madau et al. 1996). Mid- to far-infrared bands also include additional contributions from dust emission, modelled as described in Somerville et al. (2012), except that here we use the dust emission templates of Chary & Elbaz (2001). A new aspect of the lightcone catalogues described in this work is the addition of *JWST* NIRCam[2] broad- and medium-band photometry.

This dataset also include predicted photometry for the *Roman Space Telescope* Wide Field Instrument (WFI), *Euclid* visible imager (VIS) Near Infrared Spectrometer and Photometer (NISP-P), and *Rubin Observatory*. We refer the reader to the companion *Semi-analytic forecasts for Roman* paper for full descriptions (Yung et al. 2022).

2.3 Dark matter halos catalogues and merger histories

Dark matter halos extracted from dark matter only cosmological N-body simulations serve as the the foundation

of our simulated lightcones. In this work, we present two sets of lightcones. The wide-field lightcones, which each span ~ 1000 arcmin^2, is based on halos extracted from the Bolshoi-Planck simulation from the MultiDark suite (Klypin et al. 2016), which has a simulated volume of (250 Mpc h^{-1})^3 and has a dark matter particle mass of $M_{\mathrm{DM}} \sim 1.5 \times 10^8\,\mathrm{M}_\odot\,h^{-1}$. This set of lightcones are the same as the ones from Somerville et al. (2021), which have coordinates encompassing the five CANDELS legacy fields, with eight realizations of each field. We note that these wide-field lightcones are much larger than the actual observed CANDELS Fields. Future *JWST* wide-field programs, such as COSMOS-Web in cycle 1, will be able to match, or even exceed, the area of these simulated wide-field lightcones. In addition, we provide a set of eight realizations of deep-fled lightcones, each spanning 132 arcmin^2, constructed with halos extracted from the IllustrisTNG-100 dark matter-only simulation (Nelson et al. 2019), which has a simulated volume of (75 Mpc h^{-1})^3 and a dark matter particle mass of $M_{\mathrm{DM}} \sim 8.9 \times 10^6\,\mathrm{M}_\odot$. Halos in both cosmological simulations are identified using ROCKSTAR and CONSISTENT TREES (Behroozi et al. 2013a,b). For more details on the halo catalogues, readers should refer to Rodríguez-Puebla et al. (2016) and Gabrielpillai et al. (2021). As in those work, we adopt the virial mass definition from Bryan & Norman (1998). Mock lightcones are then created using the `lightcone` package from UNIVERSEMACHINE (Behroozi et al. 2020).

For each lightcone realization, the `lightcone` tool picks a random origin and viewing angle within the base dark matter-only simulation (see Table 1), and includes all halos that fall within the specified survey area. The tool makes use of the periodic boundary conditions when halos lie beyond the boundary of the simulated volume. The distance along the lightcone axis determines the redshift of the simulation snapshot from which halo properties are taken. While the lightcones in this paper were allowed to pass through the same region of the simulation volume multiple times, since the halos are sampled at a random angle, it is unlikely that a slice of the lightcone will be repeated in the same redshift slice (which happens only if halos are sampled in a slice that is perpendicular to the boundary of the simulation).

Using the virial mass of each halo in the lightcone as the 'root mass', a Monte Carlo realization of the merger history of each halo were created using an extended Press-Schechter (EPS) based method (Somerville & Kolatt 1999; Somerville et al. 2008). Merger trees are resolved down to progenitors with a mass of at least a 100th of the root mass for all halos. Merger trees are only generated for halos with $M_{\mathrm{h}} > 10^{10}\,\mathrm{M}_\odot$ and $M_{\mathrm{h}} > 5.75 \times 10^8\,\mathrm{M}_\odot$ for the wide-field and ultra-deep lightcones, respectively, which is equivalent to requiring at least 64 dark-matter particles for a halo to be resolved. We note that at $z > 10$ for the ultra-deep lightcones, we only include halos with $M_{\mathrm{h}} > 3 \times 10^9\,\mathrm{M}_\odot$ to avoid processing low-mass halos in regimes where the merger tree algorithm become highly uncertain, and galaxies are likely to be too faint to be observed with *JWST*. The steps involved in constructing halo lightcones and merger trees are summarized in Fig. 1.

In Fig. 2, we show an *illustration* of the comoving spatial distribution of a subset of galaxies in one of the simulated lightcones for the COSMOS field. In order to avoid saturation, this plot only contains 1% of objects

---

[2] https://jwst-docs.stsci.edu/near-infrared-camera/nircam-instrumentation/nircam-filters
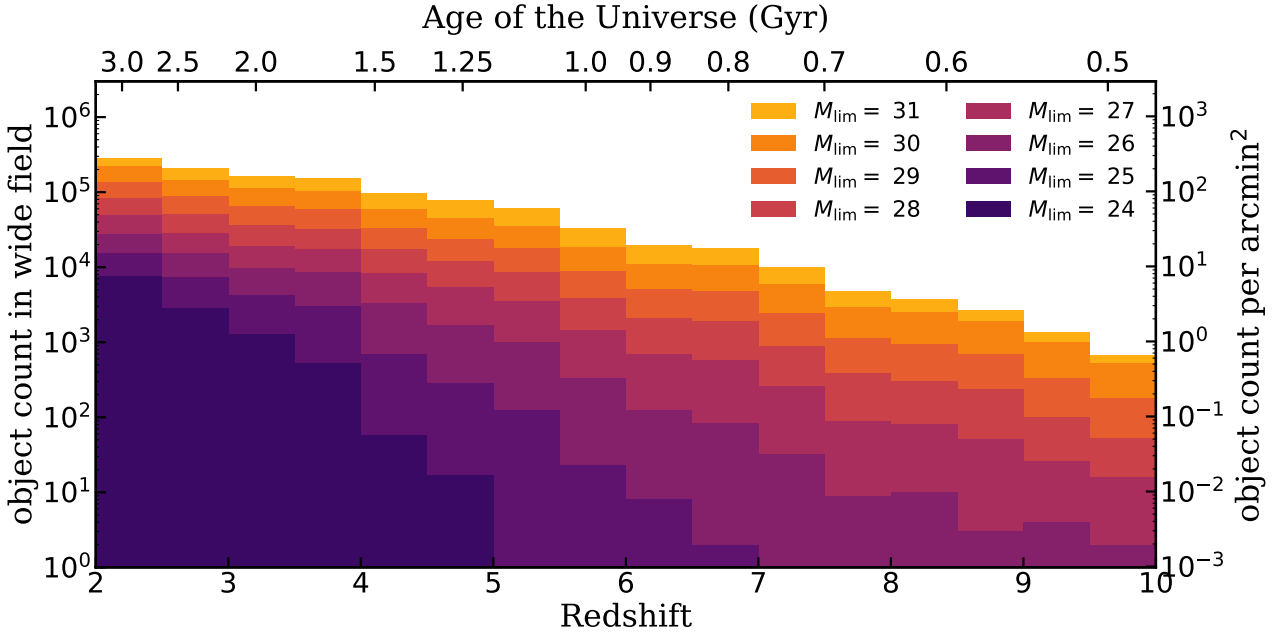
Figure 3. This histogram shows the number of objects predicted in a simulated wide-field lightcone (GOODS-N, realization 0) binned by redshift. This field spans 1024 arcmin². The number counts are colour-coded for a given limiting observed magnitude. On the right axis, we mark the equivalent number density of objects expected per arcmin², and on the top axis, we plot the age of the Universe equivalent to the marked redshift.

with $\log(M_*/M_\odot) \geq 8.0$ at $z \leq 2$, 2% of objects with $\log(M_*/M_\odot) \geq 7.5$ between $2 < z < 7$, and 25% of objects with $\log(M_*/M_\odot) \geq 7.0$ at $z \geq 7$ that are uniformly sampled from the simulated lightcone. Note that in this illustration, the data points are colour-coded and size-coded according to the predicted stellar mass of the simulated galaxies. The size of the data points does not reflect the galaxies' angular size or physical size. Satellite galaxies are omitted from this illustration.

In previous papers in the *Semi-analytic forecasts for JWST* series, instead of a halo catalogue from an N-body simulation, we used a grid of halo masses, where each mass bin contained the same number of halo realizations. These results are then weighted by the halo mass function to obtain cosmologically representative quantities. This 'grid mode' approach allows more efficient sampling of halos across an extremely wide mass range. For instance, in Paper I through Paper III, we adopted a grid of one hundred halo masses equally spaced spanning $V_{\rm vir} = 20$–500 km s^{-1} at output redshifts between $z = 4$ to 10 at integer increments. In Paper IV, we extended these outputs with identical configurations up to $z = 15$. In order to better capture AGN powered by supermassive black holes found in massive, rare halos, in Paper V we adopted a new mass grid of two hundred masses spanning $V_{\rm vir} = 100$–1400 km s^{-1} at discrete output redshifts between $z = 2$ to 7 at half-integer increments. For each halo mass in the grid, one hundred Monte Carlo realizations of merger histories were created using the same EPS-based tree algorithm used here. The volume-averaged number density of these grid halos are weighted using halo mass functions from the Bolshoi-Planck simulations (Klypin et al. 2016; Rodríguez-Puebla et al. 2016, 2017).

The full object catalogues from these previous papers

are also released via the same data portal as the lightcone data released with this work.

3 SIMULATED DATA PRODUCTS

Based on the physical models that have been examined extensively and shown to reproduce existing observational constraints up to $z \sim 10$, we present predictions for large populations of galaxies across wide ranges of redshift and halo masses, providing a comprehensive compilation of photometric and physical properties. In addition, we also provide full high-resolution spectra and star formation histories.

3.1 *JWST* wide-field lightcones

The five CANDELS fields are some of the most extensively surveyed patches of the sky and are expected to be frequently revisited by future surveys with *JWST* and other instruments. Expanding on the same framework that generated the SC-SAM simulated lightcones presented in S21, we provide additional *JWST*-specific predictions, including infrared photometry for NIRCam broad- and medium-band filters, and include fainter galaxies that are expected to be detected by *JWST* by relaxing the luminosity threshold from HST/WFC3 $m_{\rm F160W} = 29$ to 42. These lightcones include galaxies spanning $z = 0$ to 10. The simulated COSMOS, EGS, GOODS-N, GOODS-S, and UDS fields span 697, 782, 1024, 1599, and 1260 arcmin², respectively. The modelling pipeline includes a hard stellar mass limit of $M_{*,\rm lim} = 10^7\,M_\odot$, such that galaxies with masses below this limit are not recorded. This threshold is set relative to $M_{\rm h,res}$. These wide-field lightcones contain galaxy samples that are complete across the range of rest-frame
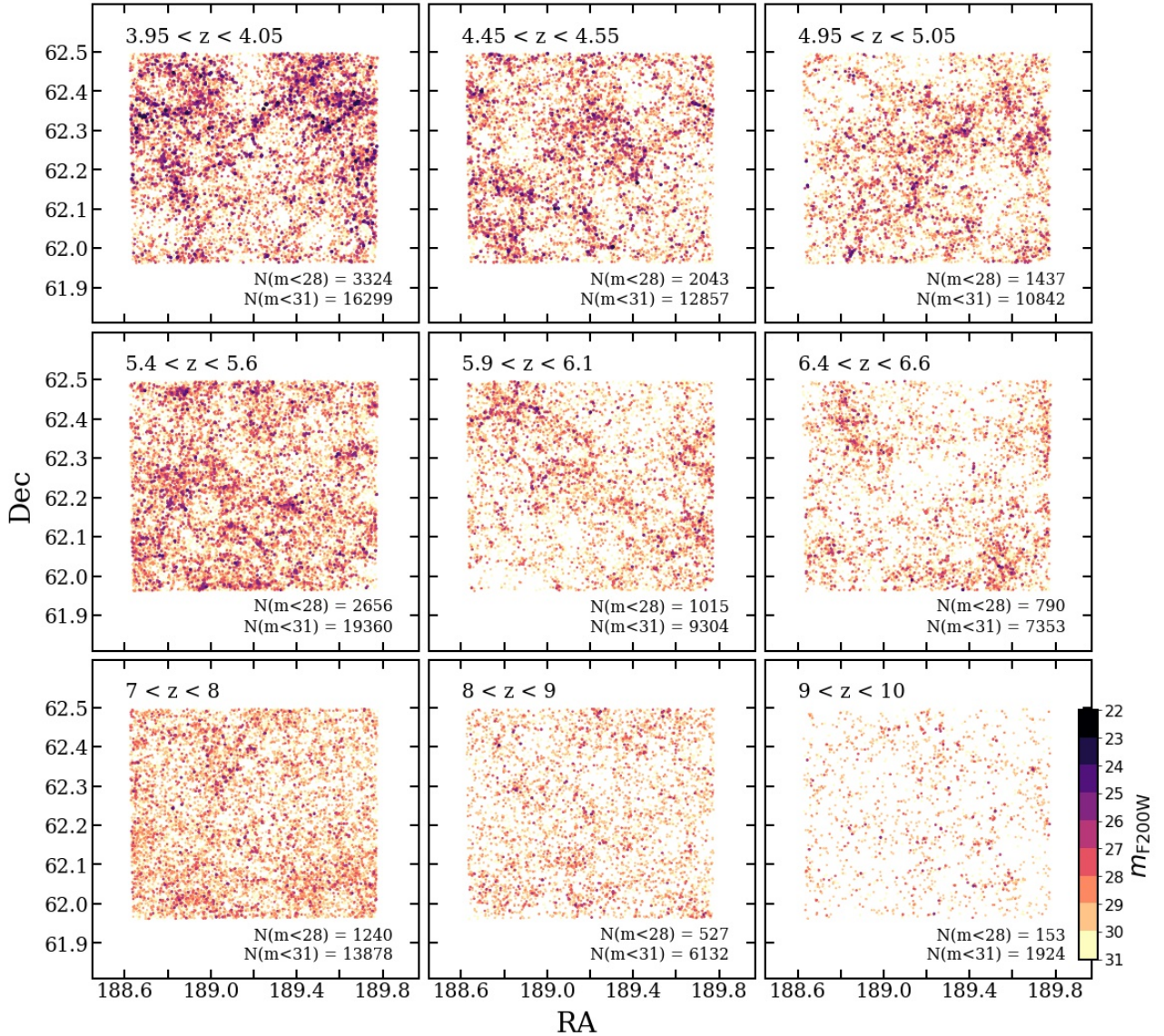
Figure 4. A summary of the footprint and galaxy populations of the first realization of the simulated GOODS-N lightcones (1024 arcmin^2) at various redshift slices between $z \sim 4$ to 10. The data points are colour-coded by the predicted observed-frame IR magnitudes in the NIRCam F200W band. The sizes of the data points are also scaled to emphasize brighter objects and do not reflect their predicted angular sizes. In addition, we show the counts of bright and faint objects within each panel with $m_{F200W} < 28$ and < 31, respectively.

UV luminosity $-16 \lesssim M_{UV} \lesssim -22$. We summarize some of the key specifications of these lightcones in Table 1. For the rest of this work, we present results from the simulated GOODS-N field to represent the wide-field lightcone results unless specified otherwise. This field is chosen because its area is close to the average among the five wide fields and the aspect ratio is close to a square.

In Fig. 3, we present the object counts as a function of redshift using the first realization of the simulated GOODS-N lightcone. The number of galaxies in this lightcone is broken down by the observed-IR magnitude in the NIRCam F200W broad-band filter m_{F200W}. We also show the object counts normalized to the survey area on the vertical axis, and the corresponding age of the universe on the horizontal axis. On one hand, this figure illustrates the number of objects predicted in the simulated field as a function of redshift

(left axis). On the other hand, this figure can be used as an easy look-up table for the number of objects expected as a function of redshift, survey depth (or equivalently exposure time), and survey area (using the normalized object counts).

Fig. 4 shows several cross-sections of the same lightcone. These slices are taken perpendicular to the line of sight, spanning $z \sim 4$ to 10. The object counts above two expected detection limits $m_{lim} = 28$ and 31 are also shown. The predicted galaxies in these figures are both size-coded and colour-coded by their observed-frame IR magnitude in the NIRCam broad-band F200W filter, m_{F200W}. This figure demonstrates the evolution of the number of objects and their spatial distribution as a function of redshift.

Figs. 3 and 4 together provide a comprehensive view of the simulated lightcone, illustrating how the predicted galaxies are distributed on the sky and in the redshift direction.
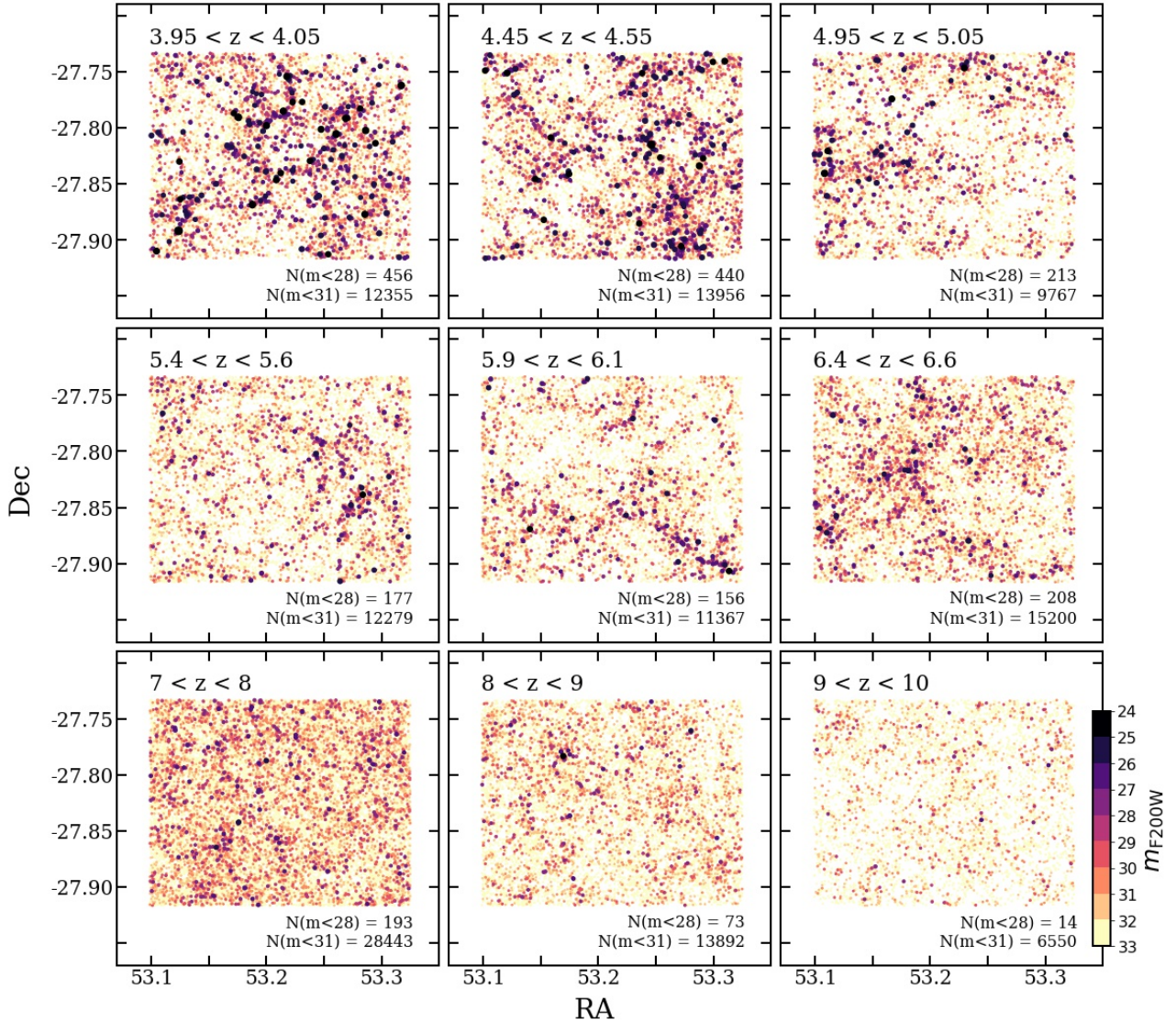
Figure 5. A summary of the footprint and galaxy populations of the first realization of the ultra-deep lightcones (132 arcmin^2) at various redshift slices between $z \sim 4$ to 10. The data points are colour-coded by the predicted observed-frame IR magnitudes in the NIRCam F200W band. The sizes of the data points are also scaled to emphasize brighter objects and do not reflect their predicted angular sizes. In addition, we show the counts of bright and faint objects within each panel with $m_{\rm F200W} < 28$ and < 31, respectively.

The lightcone construct of these mock galaxy catalogues enables a wide variety of predictions for observations, such as predictions that are sensitive to the survey geometry and quantities that are measured with non-negligible Δz, which are not achievable with a grid of halos (e.g. previous papers in this series) or with a simulated snapshot at a single redshift or a coarse grid of redshifts (e.g. most state-of-the-art cosmological simulations). Furthermore, the geometries and depths of these lightcones can be further customized to match real surveys, which enables direct comparison with an observed field, making them an ideal tool for survey planning and interpretation.

These wide-field lightcones reach a comparable depth to that expected for planned wide-field *JWST* surveys, such as CEERS, PRIMER, and COSMOS-Web. For example, CEERS has a five-sigma magnitude limit of $m_{5\sigma} \sim 29.0$ We note that the CEERS survey is expected to observe the EGS field with *JWST* (Finkelstein et al. 2017). The simulated EGS field has been used to aid the planning of CEERS and has been released as part of the project's pre-launch data products[3], these catalogues are further processed into simulated NIRCam, MIRI, and NIRISS images (Bagley et al. in preparation).

Some model adjustments made in this work, particularly the recalibration of dust optical depth parameters based on the results from Paper V, are applied to the galaxies and predicted quantities that were previously presented in S21. Therefore, this work also doubles as the second data release (DR2) for the mock CANDELS lightcones.

[3] CEERS Simulated Data Release v1: https://ceers.github.io/sdr1.html#catalogs

Table 1. This table summarizes the dimension, area, and key configurations for the lightcones release with this work and the sets of available data.

| Specification	Wide (GOODS-N)	Ultra-deep
Dimension (arcmin)	32×32	12×11
Area (arcmin^2)	1024	132
Base simulation	Bolshoi-Planck	TNG-100 DM
$\log M_{\rm h,res}/\,{\rm M}_\odot$	10.00	8.76
$\log M_{*,\rm lim}/\,{\rm M}_\odot$	7.00	5.76
$M_{\rm UV}$ range	-16 to -22	-12 to -21
redshift range	$0 < z \leq 10$	$0 < z \lesssim 12$

3.2　*JWST* ultra-deep lightcones

In addition to the set of wide-field lightcones enhanced with *JWST* observables, we introduce a brand-new set of ultra-deep lightcones with superior mass resolution but for a smaller survey area. These lightcones cover an area of 132 arcmin^2, containing galaxies spanning $0 < z \lesssim 12$. At $z \leq 10$, our model is configured to output all galaxies with stellar mass above a limit of $M_{*,\rm lim} = 5.7 \times 10^5\,{\rm M}_\odot$, which is set relative to the mass at which the halos are resolved in the halo catalogue $M_{\rm h,res} = 5.7 \times 10^8\,{\rm M}_\odot$. These ultra-deep lightcones contain galaxy samples that are complete across the range of rest-frame UV luminosity $-12 \lesssim M_{\rm UV} \lesssim -21$, or down to $m_{\rm F200W} \sim 34$ at $z \lesssim 10$. For dark matter halos at $z > 10$, only halos above $M_{\rm h,res} = 3 \times 10^9\,{\rm M}_\odot$ are processed, which give a complete galaxy sample down to $m_{\rm F200W} \sim 32$, with the assumption that galaxies forming in halos below this mass are too faint to be detected by *JWST*. We also deliver eight realizations of this simulated field, consistent with the wide-field lightcones. The key specifications of these lightcones are summarized in Table 1.

　　Fig. 5 shows the survey footprint for the first realization of the ultra-deep lightcones. Data points are both colour-coded and size-coded by their observed IR magnitude. These new lightcones are geared towards predictions for *JWST* deep-field surveys, such as NGDEEP, which is expected to reach a limit of $m_{5\sigma} \sim 30.7$. This lightcone was used in the planning of the NGDEEP survey (Finkelstein et al. 2021). Even though the performance of the SAM has been thoroughly examined and shown to reproduce a wide variety of existing observational constraints in Paper I and Paper II, we note that the physical properties for the predicted galaxy populations at $z \gtrsim 10$ are poorly constrained due to the current lack of direct observations. We refer the reader to Paper IV, where the rest-frame UV luminosity functions for $z = 11$ to 15 were first presented, for a thorough discussion of the limitations and uncertainties of the SAM in this regime. These predictions will be tested more stringently as high-redshift observational constraints from *JWST* become available.

4　RESULTS

In this section, we present key predictions at high redshift that are derived from the simulated lightcones. Specifically, we present several one-point distribution functions and two-point correlation functions. These specific results are se-
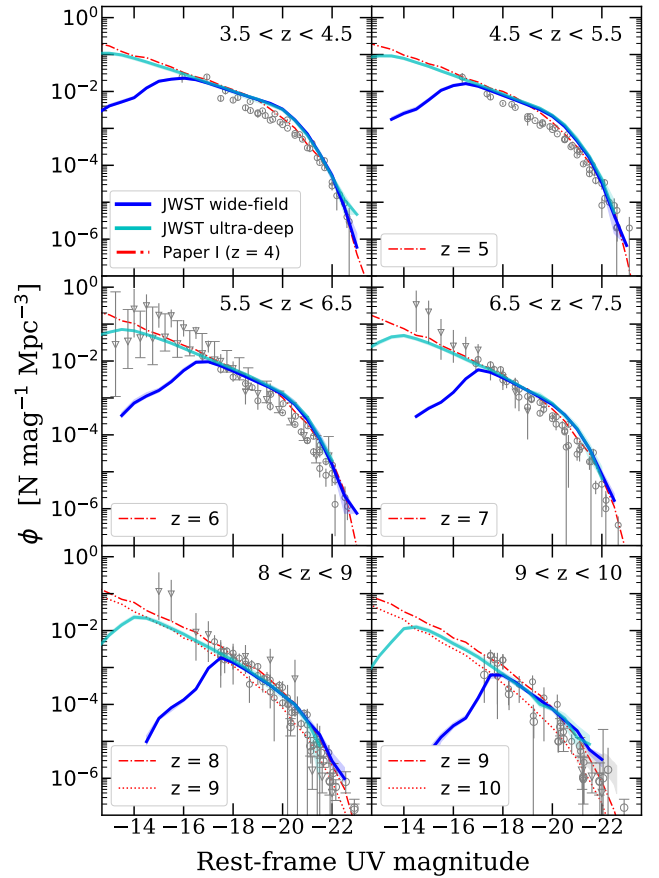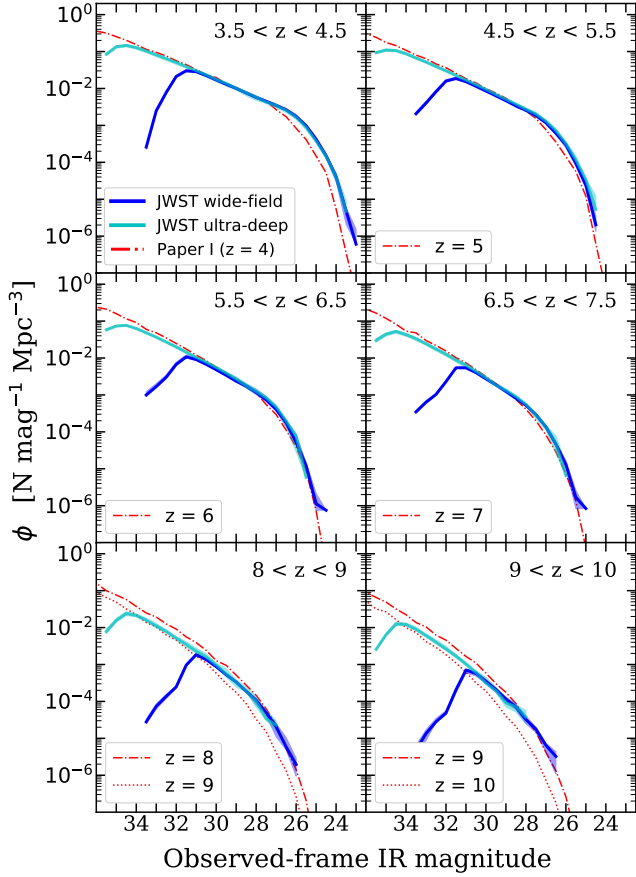

Figure 6. Predicted rest-frame UV LFs (including the effects of ISM dust attenuation and IGM extinction) at various redshift slices, together spanning the range between $z \sim 4$ to 10. The blue and cyan solid lines show results from the simulated *JWST* wide-field (GOODS-N) and ultra-deep lightcones, respectively. The line shows the median of the eight realizations, and the shaded area marks the 16th and 84th percentile. We note that in most cases the shaded area is smaller than the line thickness. The red lines show results from a grid that is uniformly sampled in halo mass, as presented in Paper I. We also show a compilation of observational constraints for comparison, where blank-field observations are shown by circle markers (Finkelstein et al. 2015, 2022; Bouwens et al. 2017, 2019; Atek et al. 2018; Oesch et al. 2018; Stefanon et al. 2019; Bowler et al. 2020) and results from lensed fields are shown with upside-down triangles (Bouwens et al. 2017; Livermore et al. 2017).

lected to demonstrate the advantages of the mock lightcones, which include the spatial distribution and redshift distribution along the line of sight, with survey area and depth that mimics real high-redshift galaxy surveys. We also take advantage of the multiple realizations available to explore field-to-field variance.

4.1　Luminosity functions

One-point distribution functions are an effective way to summarize the statistical characteristics of specific observable or physical properties over a population of detected galaxies. In this subsection, we show the one-point distribution func-

Figure 7. Predicted LFs for observed-frame magnitude in the NIRCam F200W filter (including the effects of ISM dust attenuation and IGM extinction) in various redshift slices, together spanning the range between $z \sim 4$ to 10. The blue and cyan solid lines show results from the simulated *JWST* wide-field (GOODS-N) and ultra-deep lightcones, respectively. The line shows the median of the eight realizations, and the shaded area marks the 16th and 84th percentile. We note that in most cases the shaded area is smaller than the line thickness. The red lines show results from a grid that is uniformly sampled in halo mass as presented in Paper I. The differences that can be seen between the lightcone and grid results are primarily due to evolution in galaxy properties and differences in the k-correction over the redshift range of the bin. The grid represents galaxy properties at a fixed 'snapshot' in time.

tions of luminosities and magnitudes for galaxies in various redshift slices in the two sets of simulated lightcones.

4.1.1 Rest-frame UV luminosity

The rest-frame UV magnitudes are computed for our model galaxies by integrating over synthetic stellar SEDs using a tophat filter centred at 1500 Å, with a full width at half maximum (FWHM) of 225 Å. These magnitudes include ISM dust attenuation as described in Section 2.2. Far-UV magnitudes calculated with a similar filter centred at 2300 Å are also available as part of the full object catalogues. In Fig. 6, we show the rest-frame UV luminosity functions (UV LFs) at 1500 Å for galaxies from eight selected redshift slices of the lightcones, spanning a wide redshift range from $z \sim 4$

to 10. We show results from the wide-field (GOODS-N) and ultra-deep lightcones. The UV LFs are first computed individually for each of the eight realizations. We show the median in solid lines and the shaded regions mark the 16th and 84th percentile.

The turn-over on the faint end of the UV LFs is indicative of the decrease in completeness in our galaxy sample due to the mass resolution limit of the underlying dark matter halo catalogue (see Table 1). On the other hand, the cut-off on the bright end of the luminosity functions reflects where the galaxy populations become too rare to produce statistically robust measurements for their volume-averaged number densities. This also shows where *JWST* wide surveys are not expected to constrain the bright end of the UV LFs as precisely as past large ground-based surveys. We note that while the wide-field and ultra-deep lightcones are made based on dark matter halo populations extracted from two different cosmological simulations, this does not have any noticeable effects on our predictions.

In addition, we show results based on a grid of haloes uniformly sampled in mass, as presented in Paper I, which covers halos spanning a mass range much wider than what can be resolved in a single typical numerical simulation (see Section 2.3). The *volume-averaged* number densities of the halos in the grid are weighted by halo mass functions extracted from cosmological simulations. We note the excellent agreement between the two methods.

We compare these predictions to a compilation of observational constraints, including Finkelstein et al. (2015, 2022); Livermore et al. (2017); Atek et al. (2018); Oesch et al. (2018); Bouwens et al. (2019); Stefanon et al. (2019); Bowler et al. (2020). Blank-field observations are shown by circle markers and the lensed results are shown with upside-down triangles. We note that the cut-off in the bright end of the UV LFs is indicative of the limit where bright, rare objects are expected to be detected in these different types of surveys. These high-redshift bright (massive) objects, which are often found by large surveys with ground-based instruments, are too rare for the expected survey volume of *JWST*. Here we highlight the objects that are expected to be detected in upcoming *JWST* surveys, and contrast them with all objects predicted by our model. We also note that lensed surveys will enable *JWST* to detect objects with even fainter intrinsic magnitudes. See Paper I and Bouwens et al. (2022) for detailed comparison of the UV LF faint-end predictions and existing lensed survey constraints.

4.1.2 Observed-frame near-IR magnitude

In Fig. 7, we show the distribution functions for observed-frame near-IR magnitude in the NIRCam broadband F200W filter, m_{F200W}, for all galaxies in the wide-field (GOODS-N) and ultra-deep lightcones. The calculation of observed-frame magnitudes include both ISM dust attenuation and IGM extinction (see Section 2.2). In addition, we show results based on a grid of halo mass as presented in Paper I, which covers haloes spanning a mass range much wider than a single typical numerical simulations (see Section 2.3). Similar to the previous figure, this figure provides a quick assessment for the observed-frame IR luminosity range the galaxies span in these wide-field and deep-deep lightcones. We note that here, the agreement between the grid mode and lightcone

predictions is not as good in some cases. This is due to the k-correction effects, which are more accurately modelled in the lightcone predictions presented here because the lightcones contain a continuous distribution of redshifts sampling the backwards past lightcone, while in our previous work, all galaxies were assumed to be at a set of discrete redshifts.

4.2 Galaxy clustering statistics

The two-point correlation function (2PCF) is a ubiquitous tool to characterize the spatial clustering of galaxies across different length scales (Peebles 1980). In particular, the angular correlation function, $w(\theta)$, represents the angular separation, θ, between random pairs of objects on the sky. A related quantity is the projected correlation function, $w_p(r_p)$, which is the correlation function in terms of the projected distance on the plane of the sky, integrated over a specific range in line-of-sight velocities.

In this subsection, we present selected 2PCFs for galaxies in our simulated lightcones. Lightcones are well suited for this application because, especially for high redshift clustering studies, it is often necessary to include galaxies from a relatively large bin in redshift in the 2PCF analysis. Galaxy properties may evolve over the time represented by the highest and lowest redshift edge of the bin. This effect is not included in theoretical predictions of the 2PCF from a single redshift slice. We make our clustering predictions accounting for the field geometry and spatial distribution of galaxies in our simulated lightcones, as well as the redshift distribution along the line of sight in a way similar to the three-dimensional distribution of objects in an *observed* field. The spatial locations of galaxies are determined by the physical coordinates of their host halos, which are inherited from the underlying N-body simulations. The contribution from peculiar velocities along the line of sight is included in the redshifts. The positions of satellite galaxies within the host halos are assigned as described in Section 2. We calculate the 2PCFs using CORRFUNC[4] (Sinha & Garrison 2020), which utilizes the Landy-Szalay estimator (Landy & Szalay 1993).

4.2.1 Low redshift

In Fig. 8, we compare the projected 2PCFs from the wide-field (GOODS-N) and ultra-deep lightcones to measurements from the PRIMUS (Coil et al. 2011; Cool et al. 2013) and DEEP2 (Newman et al. 2013) galaxy surveys presented by (Skibba et al. 2015). The comparison is done in two redshift bins, $0.2 < z < 0.7$ and $0.7 < z < 1.2$, for galaxies with stellar mass $\log(M_*/M_\odot) \gtrsim 10.0$ and $\gtrsim 10.5$, respectively. These projected 2PCFs are measured by integrating $\xi(r_p, \pi)$ out to line-of-sight separations of $\pi_{\max} = 80h^{-1}$ Mpc and $20h^{-1}$ Mpc for PRIMUS and DEEP2, respectively.

We show results from the eight GOODS-N wide-field lightcones and the eight ultra-deep lightcones. We computed the 2PCFs independently for each realization of the fields. These predicted 2PCFs are integrated out to $\pi_{\max} = 40h^{-1}$ Mpc. The dashed line shows the median of the eight realizations and the shaded regions show the 16th and 84th percentile, characterizing the field-to-field variance. Note that

the shaded regions are *not* expected to be representative of the field-to-field variance in the observational surveys, as the areas and geometries of the mock and real surveys are very different.

In addition, we present new measurements of the 2PCF from the five observed CANDELS legacy fields, utilizing the theory-friendly catalogues from the CANDELS surveys curated by S21. The theory friendly catalogues are a set of catalogues for the five CANDELS field that provide a standardized set of photometric and physical properties for galaxies, including redshifts and stellar masses. We utilize the setup with CORRFUNC same as the one applied to the mock lightcones, and calculate $w_p(r_p)$ for observed galaxies. In this calculation, a random catalogue with matching object counts in generated within an area similar to the observed fields, which is approximated with a quadrilateral shape that traces the edges of the field. The projected auto-correlation functions are first calculated independently for each of the five observed CANDELS fields. In Fig. 8, the data points and error bars show the median and the 16th and 84th percentiles across the five fields. We applied matching stellar mass cuts to the observed galaxy samples based on the inferred median stellar mass at observation time included in the theory-friendly catalogue, which is based on Pacifici et al. (2012, 2016).

We note that the 2PCF estimate and the field-to-field variance can be sensitive to survey geometry, depth, and completeness, and redshift accuracy. Given that there are many differences between the PRIMUS and DEEP2 surveys and CANDELS, and that we have made the crude approximation that the depth of the CANDELS fields is uniform over their entire area, we find the agreement encouraging. For example, PRIMUS and DEEP2 utilize prism-based and spectroscopic redshifts, which are much more accurate than the CANDELS photometric redshifts that are all that is available for most galaxies. In addition, PRIMUS and DEEP2 have much larger areas than CANDELS. The 'up-tick' in $w_p(r_p)$ at large separations is likely due to the small area of the CANDELS fields.

Overall, the SC SAM model predictions are in good agreement with observational estimates of the 2PCF. This is a genuine prediction of the models, as no information about galaxy clustering was used to calibrate the models. These results also demonstrate the range of pair separations over which we expect to be able to obtain robust measurements. This depends on survey area and depth, and object selection criteria. The theoretical results shown here are not intended to be representative of a particular observational survey, but users may download our catalogues and create customized predictions for a desired set of survey characteristics. We note that we are unable to generate meaningful projected 2PCF for the ultra-deep cone in the lowest redshift bin due to both the small physical scale the cone spans and the small number of galaxies that fall within the selection criteria.
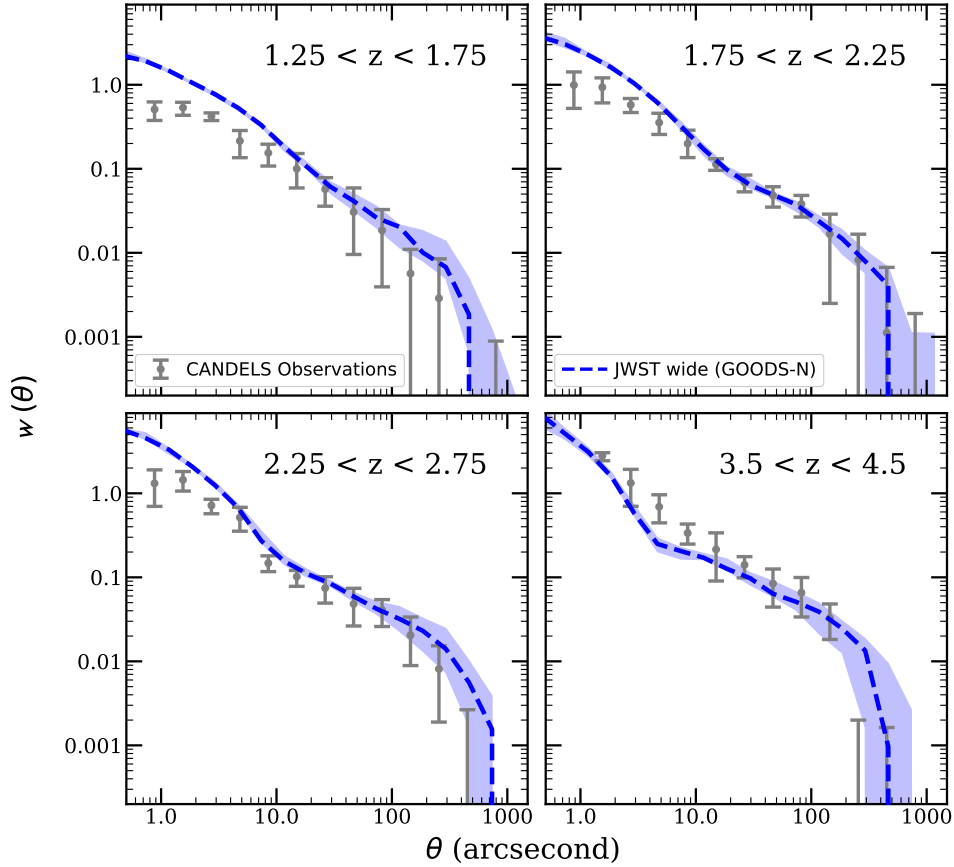
4.2.2 Intermediate redshift

In Fig. 9, we compare the angular 2PCFs from the wide-field lightcones with observations at intermediate redshift $1.25 \lesssim z \lesssim 4.5$. The observational constraints are computed from the galaxies in the CANDELS theory-friendly catalogues as described in the previous subsection. We select
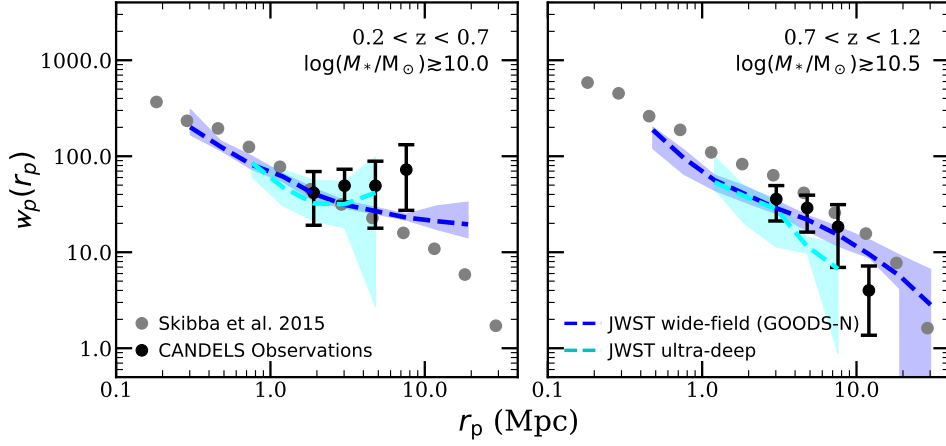
Figure 8. A comparison of the projected 2PCF, $w_p(r_p)$, between our calculations based on our mock lightcones and observations from PRIMUS and DEEP2 observations (Skibba et al. 2015) in two redshift bins, $0.2 < z < 0.7$ and $0.7 < z < 1.2$, for galaxies with stellar mass $\log(M_*/M_\odot) \gtrsim 10.0$ and $\gtrsim 10.5$, respectively. We also show results for the 2PCF which we have computed from the five CANDELS observed fields, for the same stellar mass and redshift bins. For comparison, we show results from our wide-field (GOODS-N) and ultra-deep lightcones in blue and cyan, respectively. The dashed lines show the median of the eight realizations, and the boundaries of the shaded regions mark the 16th and 84th percentile over the different realizations.

Figure 9. A comparison of the angular 2PCF, $w(\theta)$, at low and intermediate redshifts between the simulated lightcones and the CANDELS observational catalogues. A magnitude cutoff of $m_{F160W} \sim 26.5$ is applied to both sets of data. The dashed and shaded area show the median and the 16th and 84th percentile among the eight realizations of the simulated wide-field (GOODS-N) lightcones. The data point and error bars mark the median and 16th and 84th percentiles among the five CANDELS legacy fields.

galaxies in both the theory-friendly catalogues and our simulated lightcones to be brighter than a magnitude limit of $m_{F160W} = 26.5$. The CANDELS wide fields are highly complete at this limit, and the photometric redshift estimates are relatively robust (see Somerville et al. 2021, and references therein).

We find qualitatively good agreement between the predicted and observed angular correlation functions at intermediate redshifts, at least on large scales. We note that the wide-field lightcones are significantly larger than the observed fields. We refer the reader to S21 for detailed specifications for both datasets. We note that the 2PCF on small scales is dominated by galaxies that reside within the same host halo (sometimes called the 'one-halo term'; the positions of these satellite galaxies are particularly uncertain in the SAM, so these small scale measurements should be given a lower weight in assessing the success of the model. We also note that we have not attempted to add photometric noise or selection effects beyond a magnitude limit, or to simulate redshift uncertainties in the mock lightcones in the comparison presented here. The considerable uncertainties in the photometric redshifts that are predominantly available in CANDELS could have a significant effect on clustering measurements. However, this is intended to be a first order comparison to give us confidence that our predicted clustering properties are reasonable. We defer a more detailed comparison with observations to a future work.

4.2.3 High redshift

In Fig. 10, we compare the 2PCF from our simulated lightcones between $z \sim 4$–7 to observational measurements from *Hubble* legacy deep imaging and Subaru/Hyper Suprime-Cam data presented by Harikane et al. (2016). These observed results are based on over $\sim 10,000$ Lyman-break galaxies identified in 10 deep optical-near-IR imaging data sets over the five CANDELS fields and the Hubble Frontier Field parallel fields. For 2PCFs calculated with our simulation lightcones, we apply a matching magnitude limit $m_{UV} < 27.6$ at $z \sim 4$ and ~ 5, and $m_{UV} < 28.4$ at $z \sim 6$ and ~ 7. The observed-frame magnitude limit is applied to the filter band with central wavelength closest to rest-frame 1500 Å in the given redshift range. This is also individually marked in the panels of Fig. 10. The redshift bin of the observed sources is estimated based on the colour selection criteria detailed in Bouwens et al. (2015). In this comparison, we adopt coarse redshift bins for our sample to compensate for the large redshift uncertainties in the observed samples. Similar results are also reported by Barone-Nugent et al. (2014) and Qiu et al. (2018). However, 2PCF calculations are highly sensitive to sample selection criteria (e.g. by stellar mass, by rest-frame luminosity, or by observed-frame luminosity), so it is difficult to consolidate the results in a homogeneous way. We find strikingly good agreement between our model predictions and the observed clustering properties of these high redshift galaxies reported by Harikane et al. (2016), within the large observational error bars.

Fig. 11 shows *forecasts* for 2PCF that may be measured by *JWST*. The predictions for wide surveys utilise the full 1024 arcmin^2 and are marked with blue lines, with the solid line marking the expected survey depth of a typical *JWST* wide-field survey of $m_{F200W} < 29.0$. This prediction is accompanied by bracketing cases of a shallower limit $m_{F200W} < 27.2$ and an ultra-deep limit $m_{F200W} < 31.0$. Similarly, we show predictions for deep surveys over the 132 arcmin^2 field. The assumed deep survey depth $m_{F200W,lim} \sim 31$ is represented by the solid lines, and is accompanied by results with shallower $m_{F200W} < 29.0$ and $m_{F200W} < 33.0$ depths. We note that at $9 < z < 10$, the number of bright objects, $m_{F200W} < 27.2$ and < 29.0 in the wide and deep fields, respectively, are not sufficient for a meaningful 2PCF calculation. All lines shown in this comparison are the medians over eight realizations calculated as described in previous sections. We emphasize again that the area and volume of the mocks is significantly larger than the anticipated area/volume of *JWST* surveys, at least those that will be available in the first several years of operation. Over time, we may be able to expand the surveyed area to be comparable to that of our mock surveys.

4.3 Field to field variance

In this section, we leverage the combined set of 40 simulated wide-field lightcones (the ones overlapping with EGS, COSMOS, GOODS-N, GOODS-S, and UDS fields) to demonstrate the potential spread in the 2PCFs and cosmic SFR due to field-to-field variance expected in future *JWST* wide-field surveys of similar sizes. Fig. 12 shows the field-to-field variation in the predicted wide-field 2PCFs with $m_{F200W} < 29.0$ (same as the solid blue lines in Fig. 11). We show the 2PCF from individual fields in grey. We mark the median and 16th and 84th percentiles with solid and dashed red lines, respectively. Tabulated values of the predicted ACF are available in Table A1. Additional ACFs computed with galaxies from the ultra-deep lightcones down to $m_{F200W} \sim 31$ are available in Table A2.

Combining results in Figs. 11 and 12, we expect that *JWST* will be able to improve constraints on galaxy clustering statistics for separation scales up to ~ 10 arcseconds up to $z \sim 7$. At higher redshifts, the number of bright galaxies decreases significantly and the statistics are significantly impacted by the number of sources in the limited survey volume. We also show that the expected field-to-field variance in angular correlation function at separation scale ~ 100 arcseconds can be up to a factor of 2 for galaxies with $m_{F200W} < 29$ in fields with area ~ 1000 arcmin^2. The effect of survey area on galaxy clustering measurements will be explored more systematically and quantitatively in Yung et al. (2022).

Similar to the experiment shown in the previous figure, Fig. 13 shows the cosmic SFR density (SFRD) as a function of redshift between $z \sim 0$ to 10 predicted in all 40 wide-field lightcones. The cosmic SFRD is calculated by summing over the contributions from galaxies with $M_{UV} < -17$ and normalizing by the comoving volume in redshift bins with width of $\Delta z = 0.5$. This shows that the cosmic SFRD estimated from one wide field (of the order of ~ 1000 deg^2, which is a few times larger than an observed CANDELS field) will be able to constrain the cosmic SFRD down to $\sim 5\%$ with a field-to-field variance uncertainty of $\sim 10\%$ to 35% from $z = 4$ to 10 for galaxies with $M_{UV} < -17$. In addition, we show the calculation from Paper I ($z = 4$ to 10) and Paper IV ($z = 11$ to 15). These predictions are made with the more flexible 'grid mode' for halo across a wide mass range
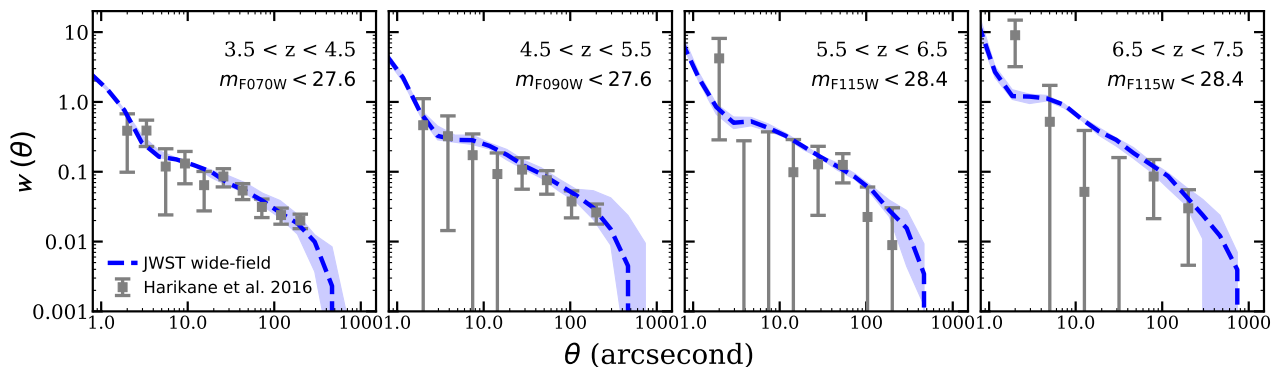
Figure 10. A comparison of the angular 2PCF, $w(\theta)$, at high redshifts between the simulated lightcones and observational constraints from Harikane et al. (2016). A magnitude cut-off is applied to the simulated data to match the observations. In this comparison, we adopt the observed-IR magnitude in the NIRCam filter that is closest to rest-frame UV 1500Å at the centre of the redshift bins, which is approximately equivalent to the selection criteria adopted for the observational constraints.
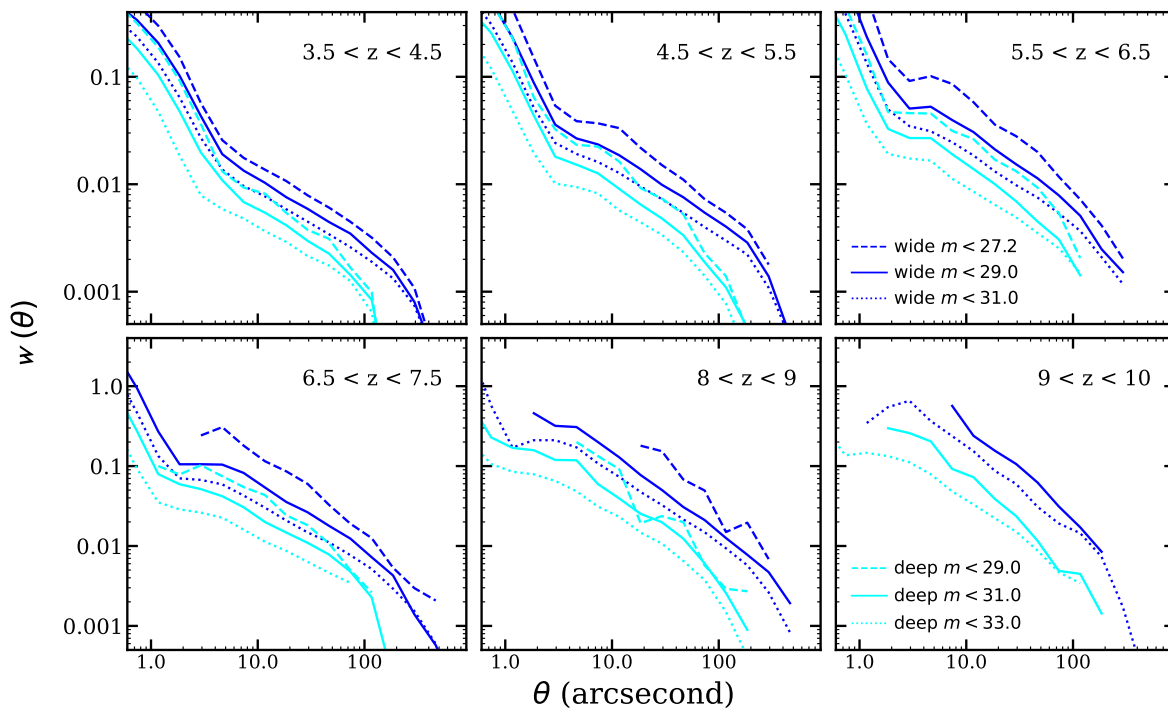
Figure 11. Predicted angular 2PCF, $w(\theta)$, at $z \sim 4$ to 10. The blue and cyan lines show predictions from the wide-field and ultra-deep lightcones, respectively. These predictions adopted an observed-frame magnitude limit of $m_{\rm F200W,lim} = 29$ and 31, which are comparable to anticipated *JWST* wide and deep surveys.

(see Section 2.3). Tabulated values of the predicted cosmic SFR are provided in Table B1.

Field-to-field variance can be one of the major sources of uncertainty in estimates of the number of objects from observational studies. Somerville et al. (2004) and Moster et al. (2011) have provided an analytic 'cookbook' for cosmic variance for stellar mass selected galaxies in CANDELS-sized deep pencil beam surveys (of the order of hundreds of arcmin^2) for $z \sim 0.5$ to 4, based on an empirically established halo occupation distribution (HOD) model. In this work, we calculate the field-to-field variance for magnitude or luminosity selected galaxies at high redshift utilizing the
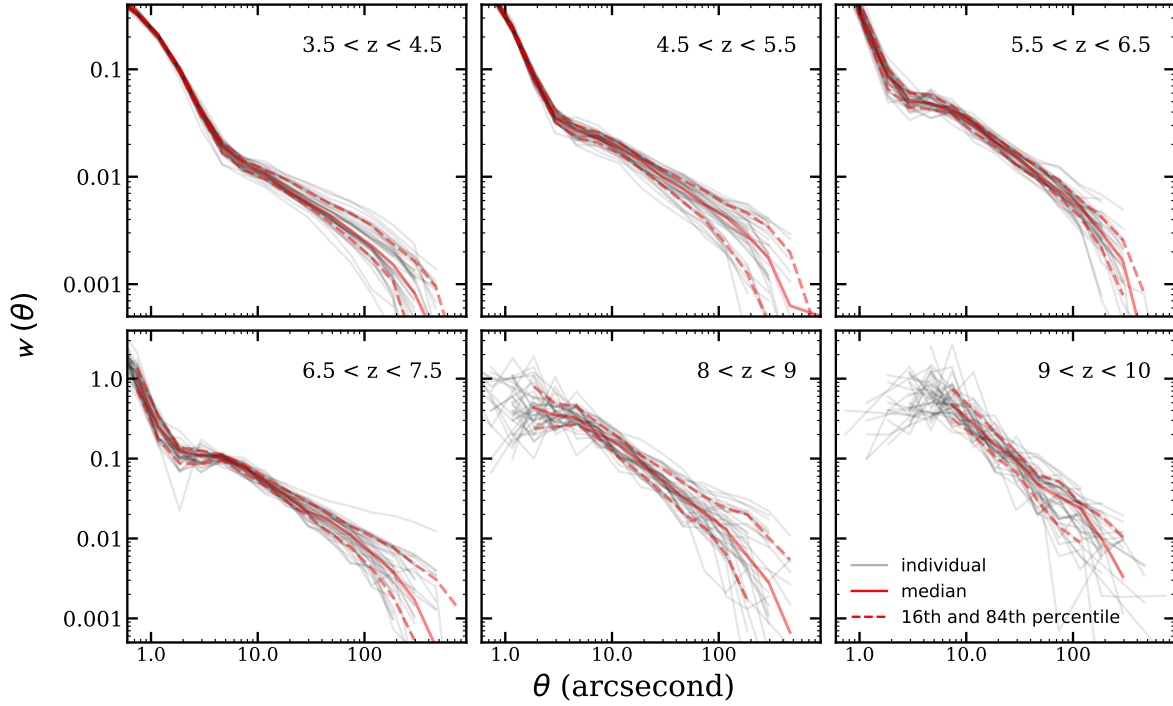
full set of 40 realizations of wide-field lightcones, which span $\sim 1000\,{\rm arcmin}^2$ each.

As detailed in Somerville et al. (2004), the relative cosmic variance (with shot noise removed) is defined as

$$\sigma_v^2 \equiv \frac{\langle N^2 \rangle - \langle N \rangle^2}{\langle N \rangle^2} - \frac{1}{\langle N \rangle}, \tag{7}$$

where $\langle N \rangle$ and $\langle N^2 \rangle$, which denote the mean and variance of object count N, respectively, are the first and second moments of the probability distribution function $P_N(V)$. Given that the variance is sensitive to volume (and hence also sur-

Figure 12. Predicted angular 2PCF, $w(\theta)$, at $z \sim 4$ to 10 for galaxies $m_{\rm F200W} < 29$ (same survey configurations and limits as blue solid lines in Fig. 11). The 2PCFs from the full set of 40 wide-field lightcones are shown individually in grey, and the median and 16th and 84th percentiles with solid and dashed red lines, respectively. Tabulated values of the predicted ACF are available in Table A1.
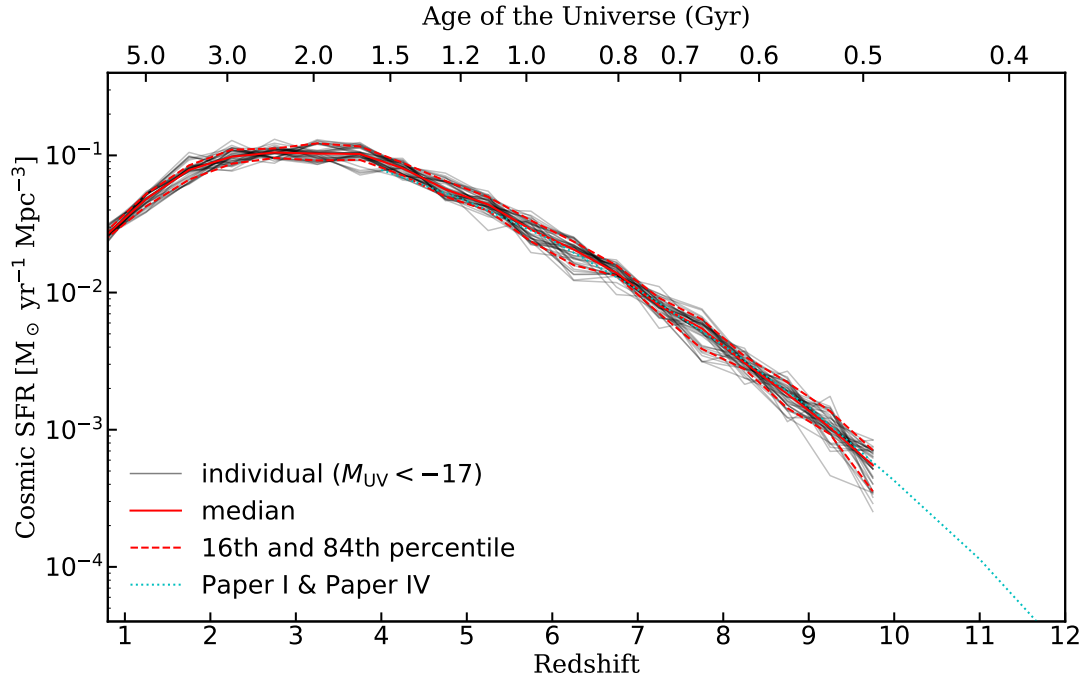
Figure 13. Predicted cosmic SFR density (SFRD) as a function of redshift at $z \sim 0$ to 10 from the wide-field lightcones for galaxies with rest-frame $M_{\rm UV} < -17$. We also show the age of the Universe equivalent to the marked redshifts. The cosmic SFRD from the full set of 40 lightcones are shown individually in grey, and the median and 16th and 84th percentiles with solid and dashed red lines, respectively. In addition, we show the calculation from Paper I ($z = 4$ to 10) and Paper IV ($z = 11$ to 15). Tabulated values of the predicted cosmic SFR are provided in Table B1.

vey area), we adopted the following modified calculation

$$\sigma_v^2 = \frac{\langle n^2 \rangle - \langle n \rangle^2}{\langle n \rangle^2} - \frac{1}{V\langle n \rangle}, \tag{8}$$

where n is the number density of galaxies within a redshift slice and V is the volume. In this exercise, we bin galaxies in each lightcone by redshift with bin width $\Delta z = 0.5$ over the range $4 \lesssim z \lesssim 10$. In this calculation, V is fixed at an assigned value such that $V\langle n \rangle > 1$ even in subregions with extremely low number density (e.g. $n < 1$ when $N > 0$). In Fig. 14, we show the calculated field-to-field variance as a function of redshift for galaxies above several thresholds in rest-frame UV magnitude or observed-frame magnitude in the NIRCam F200W band. Based on the full set of 40 realizations of wide-field lightcones, We show that the expected fractional root variance evolves from $\sigma_V \sim 0.04$ to ~ 0.27 from $z \sim 4$ to 10 for galaxies with $m_{\rm F200W} < 29$ in fields with area ~ 1000 arcmin^2.

We note that the full set of wide-field lightcones provides 40 example fields for our field-to-field variance calculation, but this would benefit from a larger number of realizations to provide better statistics. The impact on field-to-field variance from survey area and depth will be further explored in Yung et al. (2022), which leverages the robust statistics provided by a set of larger area 2-deg^2 lightcones.

5 DISCUSSION

These simulated lightcones enable a whole new set of applications which will facilitate both planning for future observational programs, and interpreting observational results. In the previous papers in this series, we presented predictions for rest- and observed-frame luminosity functions (Paper I); stellar mass functions, stellar-to-halo mass ratios, and other important physical properties (Paper II); ionizing photon production efficiency (Paper III) and their subsequent impact on cosmic hydrogen reionization (Paper IV); and luminosity functions for active galactic nuclei and their subsequent impact on cosmic helium reionization (Paper V). In this work we exploit the lightcones to add predictions of galaxy clustering and field-to-field variance. The lightcones are made publicly available, to enable explorations of the effect of survey size and depth for any desired survey configuration. Furthermore, this enables studies of the evolution of galaxy populations over a continuous redshift distribution, reflecting the evolutionary effects in real surveys with broad redshift bins.

5.1 Lightcones compared with snapshots

We first compared one-point distribution functions for rest-frame UV luminosity and observed frame magnitude as predicted by our new lightcones and by our previous approach of utilizing a grid of halos at a fixed output redshift. As expected, the distributions for the rest-UF luminosity agree quite well, although one can see that in some cases there is significant evolution over the redshift range spanned by the bin (e.g. $8 < z < 9$ and $9 < z < 10$ in Fig. 6). Larger differences are seen in the distributions of observed frame magnitude, especially in the lowest redshift bin that we show ($3.5 < z < 4.5$), which is due to the effects of k-corrections.

This comparison allows estimates of the magnitude of the effect of a continuous redshift distribution on observables compared to using a fixed snapshot in redshift, as is frequently done in comparisons with cosmological hydrodynamic simulations.

5.2 Angular Clustering at low, medium, and high redshift

Our main new result is the presentation of measurements of the projected (angular) two-point correlation function (2PCF) both for the CANDELS observations and for our mock lightcones, over a broad redshift range. At low redshift ($0.2 < z < 1.2$), we compare the projected correlation function $w_p(r_p)$ estimate from the PRIMUS and DEEP2 redshifts surveys published by Skibba et al. (2015) with our estimates based on the five observed CANDELS fields for the same stellar mass limit. To our knowledge, the angular correlation functions from CANDELS have not been published before. Our estimate does not attempt to account for the variable depth across the CANDELS fields; however, we select regions for which the depth is reasonably uniform (see S21 for details). Considering the very different areas of these surveys, the different methods used to estimate stellar masses, and the very different precision of the redshift estimates (prism-based or spectroscopic in PRIMUS and DEEP2 vs. predominantly photometric in CANDELS), we found the agreement very encouraging. Unsurprisingly, we find a fairly large field-to-field variance in $w_p(r_p)$ across the five CANDELS fields, particularly in the smaller volume lower redshift bin. We also compare with the predictions for $w_p(r_p)$ from our mock lightcones, again adopting the same stellar mass limit, but not attempting to mimic specific survey characteristics or to include observational effects such as detailed selection or redshift uncertainties. Again, we find remarkably good agreement between the estimates based on our mock lightcones and the observational estimates, mostly within the uncertainties of the observed CANDELS fields and within a factor of 2 in the small-scale end.

In addition, we compared the angular correlation function (ACF) that we measured from the observed CANDELS catalogues (again, from all five fields) with that measured from our mock lightcones for intermediate redshifts ($1.25 < z < 4.5$, in four redshift bins). Again, considering the relative crudeness of the comparison, we were encouraged by the agreement. Referring to Fig. 9, we clearly see the change in slope of the ACF at around 4-10 arcsecond which is thought to be due to the transition from domination by the 'one halo' term (galaxies that reside within the same dark matter halo) to the 'two-halo' term (galaxies residing in separate host halos). That this transition occurs at roughly the same scale in both the observational and theoretical estimates is particularly encouraging (the contribution of the one- and two-halo term to our predicted ACF will be quantified in more detail in Yung et al. (2022)). We note in this context that the modelling of satellite galaxies in SAMs (which produce the one-halo term) is particularly uncertain (see Section 5.4 for a discussion).

Lastly we compare the estimates of the angular correlation function at high redshift ($4 < z < 7$) from our mock lightcones with observational estimates from HST and Subaru (Harikane et al. 2016). Once again, considering the many uncertainties, predictions from our lightcones are well within
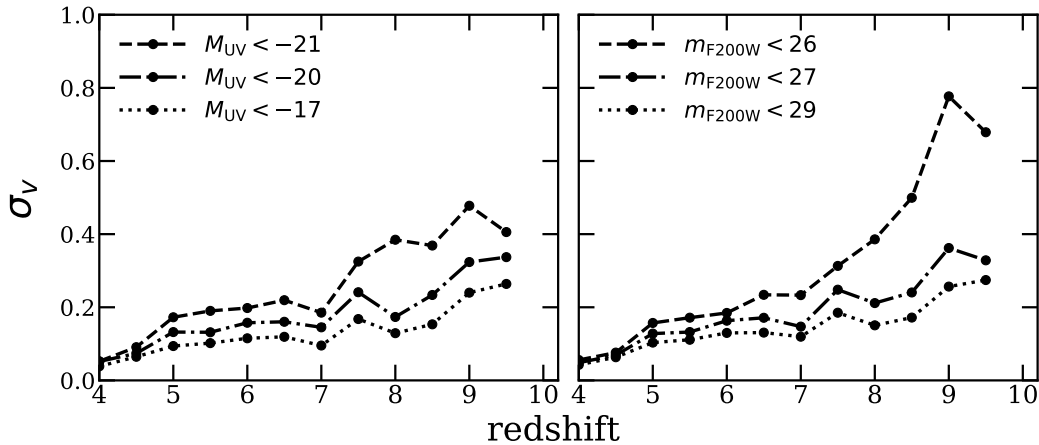
Figure 14. Square root of field to field variance, σ_V, as a function of redshift between $4 < z < 10$ calculated for galaxies across the full set of 40 realizations of wide-field lightcones, each spanning ~ 1000 arcmin^2. The difference in line style represents the variance expected for galaxies brighter than the specified rest-frame UV magnitudes (*left*) and observed-frame F200W magnitudes (*right*). See text for detailed descriptions of redshift binning and volume normalization across lightcones with different sizes.

the errors provided for these observational constraints, including a hint of the transition from the one- to two-halo term.

We emphasize that these are genuine predictions of our model, as no information about clustering is used to calibrate the model. It has been demonstrated before that when galaxy global properties (stellar mass or luminosity) are mapped to halo mass using abundance matching, good agreement with observed clustering is obtained (Conroy & Wechsler 2009; Behroozi et al. 2013c, 2019, 2020; Moster et al. 2013, 2018). These previous studies have generally focussed on low to intermediate redshift. We have shown previously that our model predictions for luminosity functions and stellar mass functions are in good agreement with existing observations up to $z \sim 8$–9 (Yung et al. 2019a,b). However, the good agreement of our model predictions with observed clustering measurements is an independent indication that the mapping between halo mass and galaxy stellar mass or luminosity assigned by the physical prescriptions in our models is consistent with that in the real Universe. Given the large uncertainties in dust corrections and stellar mass estimates at high redshift, this is an important confirmation.

Encouraged by these results, we present forecasts for the angular correlation function measurements that may be obtained with future *JWST* surveys with a range of areas and depths. Additionally, using the many realizations of our mock lightcones, we illustrate the field-to-field variance in measurements from surveys with an area of ~ 1000 arcmin^2. Our forecasts suggest that we may eventually be able to obtain robust measurements of angular clustering out to redshifts of 8 or 9 with *JWST*.

In principle, one can attempt to estimate the host halo mass of a galaxy population based on its observed clustering, using a halo occupation distribution (HOD) formalism (Berlind & Weinberg 2002; Zehavi et al. 2004, 2011; Zheng et al. 2007; Harikane et al. 2016). This can then be used as an alternative method for constraining the relationship between galaxy luminosity or stellar mass and dark matter halo mass. Our mocks can be used to forecast the expected

accuracy of such estimates based on specific survey characteristics. We plan to explore this in a future work.

5.3 A tool for detailed estimates of field-to-field variance

The multiple realizations of lightcones with a range of area and depth provide a powerful tool to the community for estimating field-to-field variance in a side variety of observable quantities in currently planned or future surveys. We have deliberately provided lightcones with larger areas than anticipated near-term *JWST* survey projects, so that they can be cut to any desired geometry. We emphasize that the quantitative results for field-to-field variance presented here do *not* correspond to specific observational surveys, but represent the expected variance over our suite of mock lightcones. Field-to-field variance can also be estimated using analytic prescriptions (Somerville et al. 2004; Trenti & Stiavelli 2008; Moster et al. 2011) and in hydrodynamic simulations (Bhowmick et al. 2020). Endsley et al. (2020) conducted a related study based on lightcones filled with galaxies from the empirical UNIVERSEMACHINE model and explored the impact on clustering predictions due to sample variance. Predictions for how field-to-field variance for the ACF and SFRD vary with survey area will be explored in more detail in Yung et al. (2022).

5.4 Caveats and limitations of current models

The limitations and caveats regarding the baryonic processes in the SAM and in galaxies at $z \gtrsim 10$ have been thoroughly discussed in previous works; we refer the reader to section 6.3 in Paper II and section 4.2 in Paper IV. This discussion will be focused mainly on the topics related to the construction of simulated lightcones and clustering predictions presented in this work.

We utilize Extended Press-Schechter based merger trees to populate halos in our lightcones with galaxies. The EPS-based merger trees adopted in this work series have been compared with trees extracted from numerical simulations and the results shown to be in good agreement. However,

the EPS algorithm has never been tested over the full halo mass and redshift ranges that are explored in this work, as there is currently no publicly available relevant suite of dark matter only simulations. Furthermore, the EPS-based merger tree approach reconstructs only an approximate reconstruction of the joint distribution of halo progenitor masses (Somerville et al. 2000).

Moreover, it is known that in N-body simulations, the merger histories of halos depend on their large scale environmental density (Avila-Reese et al. 2005; Maulbetsch et al. 2007; Wang et al. 2007; Fakhouri & Ma 2009, 2010). This second-order dependence is not captured by any of the current semi-analytic merger tree building algorithms. Because galaxy properties such as stellar mass are known to depend strongly on halo formation history in SAMs (Gabrielpillai et al. 2021), this implies that the mapping between stellar mass and halo mass predicted by the SAMs presented here will not properly reflect this secondary dependence on large scale environment, which has implications for the predicted clustering. Similarly, each halo is processed independently with no 'knowledge' of its larger scale environment. Other processes, such as metal enrichment or radiation fields could introduce an environmental dependence on galaxy physical properties, which will not be captured by our current models.

The properties of satellite galaxies are particularly uncertain in our SAM framework. The SC SAM includes a semi-analytic model that describes the decay of satellite orbits due to dynamical friction, and the tidal stripping and destruction of these objects as they orbit within the halo. However, this modelling is highly uncertain and does not include the effects of stripping and destruction due to strong encounters with the baryonic components of galaxies, which can have a large effect on tidal stripping and destruction (Dooley et al. 2016; Smith et al. 2016; Han et al. 2018). The satellite population is also particularly sensitive to inaccuracies in the EPS-based halo merger trees discussed above. Furthermore, we currently assume that when a galaxy becomes a satellite, its hot gas reservoir is instantaneously stripped off, robbing the galaxy of a supply of new fuel for star formation. The satellite galaxy continues to eject gas from its ISM via stellar feedback, but this gas is assumed to be added to the hot gas reservoir of the central galaxy. These assumptions may lead to satellite galaxies being prematurely starved of fresh gas and quenching too early (e.g. Kimm et al. 2009). Finally, we have found that the prescription for tidal stripping and destruction in our models does not yield a radial distribution of satellites that is in agreement with the predictions from N-body simulations, so we have reassigned satellite positions in post-processing, making the simple assumption that the radial profile of satellites traces the overall dark matter density profile as represented by an NFW profiles. This is known not to be the case in detail in hydrodynamic simulations (Bose et al. 2020, 2022; McDonough & Brainerd 2022).

Current photometry are computed based on composite stellar spectra from BC03. The photometry in the released mock lightcones include contributions from the stellar continuum only, and do not include the emission from nebular lines or continuum, nor radiation from AGN. In most cases, these contributions will be negligible. However, in a future work, we will implement emission line predictions from mod-els presented by Hirschmann et al. (2017, 2019) and explore their effects on broad-, medium-, and narrow band photometry (Yung, Hirschmann, Somerville et al., in prep). A larger source of uncertainty is the impact of attenuation from ISM dust. We have adopted a simple, empirical approach in which we adjust the normalization of the optical depth at the V band in order to match the observed UV luminosity function, and we adopt a fixed functional form for the attenuation as a function of wavelength (fixed attenuation curve). The finding that we need to adjust the 'fudge factor' for the dust optical depth as an ad hoc function of wavelength is a good indication that we do not yet have a good physical handle on the evolution of dust properties. We note that many other studies have similarly found that an ad hoc adjustment of the dust optical depth normalization is needed to match observations (e.g. Vogelsberger et al. 2020). We also plan on improving this component of our modelling in future work. Lastly, we have adopted the Madau et al. (1996) prescription for attenuation by the IGM throughout our lightcones, while it is only applicable for a uniform IGM, i.e. post-reionization. Improving on this would require full radiative transfer.

5.5 The role of *JWST forecasts* in survey planning

These lightcones have been used extensively in developing upcoming *JWST* surveys CEERS and NGDEEP surveys, where the simulated datasets are used to inform survey specifications, including number of objects expected for a given survey area and exposure time. The simulated catalogues are also further processed into mock images, which are useful for practising data reduction, source extraction, removal of dithering pattern, and optimizing colour selection criteria. At the time when *JWST* observations become available, the theory framework presented in the work series will be utilized for the physical interpretation of detected sources.

While *JWST* will be the powerhouse for high-redshift observations, its relatively narrow field of view restricts its ability to explore large-scale structures. Selected *JWST* Cycle 1 program COSMOS-Web (Kartaltepe et al. 2021) is expected to survey ~ 0.6 deg^2 of the sky, which is expected to be larger than any existing contiguous fields surveyed by *Hubble* while reaching a survey depth comparable to *HUDF*. Current generation of *JWST* wide-field surveys will set up for future deep-field with *Roman Space Telescope*, which IR sensitivity is comparable to *Hubble* and a single pointing of its on-board Wide-Field Instrument (WFI) will span ~ 0.28 deg^2. *JWST* observations will also shape future high-redshift surveys with ESA's *Euclid* Observatory. A companion paper *Semi-analytic forecasts for Roman* (Yung et al. 2022) will present a set of 2-deg^2 lightcones that are constructed with the same infrastructure that produces the lightcones presented in this work and will support future wide-field surveys and the potential synergy across multiple instruments.

6 SUMMARY AND CONCLUSIONS

In this work, we present mock galaxy catalogues that are optimized to support the planning and interpretation of *JWST* surveys. These mock catalogues come in two flavours: wide-field and deep-field. The set of wide-field lightcones consists

of five independent fields with coordinates and geometries overlapping the five CANDELS legacy fields, spanning $z \sim 0$ to 10 and resolving down to observed-frame $m_{\rm F200W} \sim 30$ or rest-frame $M_{\rm UV} \sim -18$. The set of ultra-deep lightcones represents a 132 arcmin^2 field overlapping HUDF, spanning range $z \sim 0$ to ~ 12 and resolving down to observed-frame $m_{\rm F200W} \sim 34$ or rest-frame $M_{\rm UV} \sim -14$. Eight realizations have been created for each field by sampling different halos from hydrodynamic simulations, which propagates physically accurate cosmic variance to the simulated fields. A total of 48 lightcones presented in this work provides a total of $\simeq 12.2 \deg^2$ of (non-contiguous) simulated area.

The construction of these past lightcones leverages both the accurate representation of structure formation in cosmological N-body simulations and the efficiency of semi-analytic models. Dark matter halos sourced from state-of-the-art cosmological simulations are projected onto past lightcones that are traced by the lines of sight within the area of a specified field. The merger histories of individual halos are constructed on-the-fly using the EPS formalism, within which the formation and evolution of galaxies and a wide range of associated physical processes are predicted by the well-established Santa Cruz SAM. In previous papers in the *Semi-analytic Forecasts* series, the SAM has been shown to reproduce a variety of observational constraints at high redshift, including UV luminosity functions, stellar mass functions, star formation rate functions, and stellar-to-halo mass ratios. We also present predictions that are specifically tailored to *JWST*, including galaxies in the luminosity (or mass) range that have not been detected before, as well as NIRCam broad- and medium-band photometry for all galaxies.

In this final paper of the *Forecasts* series, we take full advantage of the mock lightcones and present new predictions that take advantage of the lightcone format, including continuous redshift evolution of galaxies and their spatial distributions. We examine the clustering of galaxies in these mock lightcones by computing projected two-point correlation functions and compare them to existing measurements from past galaxy surveys. Our results are in excellent agreement with past observations across a wide range of redshifts. Utilizing the two sets of simulated lightcones covering areas and depths that are similar to future *JWST* wide and deep surveys, we provide *forecasts* for two-point correlation functions that may be obtained from future *JWST* surveys.

We summarize our main conclusions below.

(i) We assembled a modelling pipeline that sources halos from N-body simulations along a past lightcone and populates them with galaxies with a wide range of predicted properties. The set of mock wide-field and ultra-deep lightcones delivers a wide variety of predictions for upcoming *JWST* surveys, including one-point distribution functions of physical and observable prosperities, as well as time-evolution of galaxy populations across different epochs and clustering statistics with two-point auto-correlation functions.

(ii) The wide-field lightcones cover ~ 1000 arcmin^2 each, containing galaxies with $-16 > M_{\rm UV} > -22$ between $0 \lesssim z \lesssim 10$. The ultra-deep lightcones cover 132 arcmin^2 each, containing galaxies $-12 > M_{\rm UV} > -21$ between $0 \lesssim z \lesssim 12$.

(iii) The predicted clustering of galaxies in our simulated

lightcones is in excellent agreement with observed clustering over a wide redshift range $0.2 \lesssim z \lesssim 7.5$.

(iv) We show that *JWST* wide-field surveys of ~ 1000 arcmin^2 with depth reaching $m_{\rm F200W} \sim 29$ will be capable of constraining galaxy clustering up to an angular separation of ~ 10 arcsecond.

(v) We show that the expected field-to-field variance in the angular correlation function at separation scale ~ 100 arcseconds can be up to a factor of 2 for galaxies with $m_{\rm F200W} < 29$ in fields with area ~ 1000 arcmin^2.

(vi) We quantify the expected field-to-field variance in the cosmic star formation rate density (SFRD) for galaxies with $M_{\rm UV} < -17$ in fields with area ~ 1000 arcmin^2 to be $\sim 10\%$ at $z \sim 4$ and up to $\sim 35\%$ at $z \sim 10$.

(vii) Based on the full set of 40 realizations of wide-field lightcones, we show that the expected fractional root variance evolves from $\sigma_V \sim 0.04$ to ~ 0.27 from $z \sim 4$ to 10 for galaxies with $m_{\rm F200W} < 29$ in fields with area ~ 1000 arcmin^2.

The results presented in this paper are intended as just a few examples of the predictions that can be extracted from these lightcones. All of the lightcones are made publicly available so that the community can exploit them for numerous additional applications.

DATA AVAILABILITY

The data underlying this paper are available in the Data Product Portal hosted by the Flatiron Institute at https://www.simonsfoundation.org/semi-analytic-forecasts-for-jwst/ and the Flatiron Institute Data Exploration and Comparison Hub (Flathub, http://flathub.flatironinstitute.org/group/sam-forecasts).

REFERENCES

Atek H., Richard J., Kneib J.-P., Schaerer D., 2018, MNRAS, 479, 5184
Avila-Reese V., Colin P., Gottlober S., Firmani C., Maulbetsch C., 2005, ApJ, 634, 51
Barone-Nugent R. L., et al., 2014, ApJ, 793, 17
Behroozi P. S., Silk J., 2015, ApJ, 799, 32
Behroozi P. S., Conroy C., Wechsler R. H., 2010, ApJ, 717, 379
Behroozi P. S., Wechsler R. H., Wu H.-Y., 2013a, ApJ, 762, 109
Behroozi P. S., Wechsler R. H., Wu H.-Y., Busha M. T., Klypin A. A., Primack J. R., 2013b, ApJ, 763, 18
Behroozi P. S., Wechsler R. H., Conroy C., 2013c, ApJ, 770, 57
Behroozi P., Wechsler R. H., Hearin A. P., Conroy C., 2019, MNRAS, 488, 3143
Behroozi P., et al., 2020, MNRAS, 499, 5702
Berlind A. A., Weinberg D. H., 2002, ApJ, 575, 587
Bhowmick A. K., Somerville R. S., Di Matteo T., Wilkins S., Feng Y., Tenneti A., 2020, MNRAS, 496, 754
Bigiel F., Leroy A., Walter F., Brinks E., de Blok W. J. G., Madore B., Thornley M. D., 2008, AJ, 136, 2846
Bose S., Deason A. J., Belokurov V., Frenk C. S., 2020, MNRAS, 495, 743
Bose S., Eisenstein D. J., Hadzhiyska B., Garrison L. H., Yuan S., 2022, MNRAS, 512, 837
Bouwens R. J., et al., 2015, ApJ, 803, 34
Bouwens R. J., Oesch P. A., Illingworth G. D., Ellis R. S., Stefanon M., 2017, ApJ, 843, 129
Bouwens R. J., Stefanon M., Oesch P. A., Illingworth G. D., Nanayakkara T., Roberts-Borsani G., Labbé I., Smit R., 2019, ApJ, 880, 25
Bouwens R. J., Illingworth G., Ellis R. S., Oesch P., Paulino-Afonso A., Ribeiro B., Stefanon M., 2022, arXiv:2203.14978
Bowler R. A. A., Jarvis M. J., Dunlop J. S., McLure R. J., McLeod D. J., Adams N. J., Milvang-Jensen B., McCracken H. J., 2020, MNRAS, 493, 2059
Bruzual G., Charlot S., 2003, MNRAS, 344, 1000
Bryan G. L., Norman M. L., 1998, ApJ, 495, 80
Calzetti D., Armus L., Bohlin R. C., Kinney A. L., Koornneef J., Storchi-Bergmann T., 2000, ApJ, 533, 682
Chary R., Elbaz D., 2001, ApJ, 556, 562
Cirasuolo M., et al., 2007, MNRAS, 380, 585
Coil A. L., et al., 2011, ApJ, 741, 8
Conroy C., Wechsler R. H., 2009, ApJ, 696, 620
Cool R. J., et al., 2013, ApJ, 767, 118
Davis M., et al., 2007, ApJ, 660, L1
Dooley G. A., Peter A. H. G., Vogelsberger M., Zavala J., Frebel A., 2016, MNRAS, 461, 710
Drakos N. E., et al., 2022, ApJ, 926, 194
Dunlop J. S., et al., 2021, JWST Propos. ID 1837. Cycle 1 GO/Treasury
Ellis R. S., et al., 2013, ApJ, 763, L7
Endsley R., Behroozi P., Stark D. P., Williams C. C., Robertson B. E., Rieke M., Gottlöber S., Yepes G., 2020, MNRAS, 493, 1178
Fakhouri O., Ma C.-P., 2009, MNRAS, 394, 1825
Fakhouri O., Ma C.-P., 2010, MNRAS, 401, 2245
Finkelstein S. L., et al., 2015, ApJ, 810, 71
Finkelstein S. L., et al., 2017, JWST Proposal ID 1345. Cycle 0 Early Release Science
Finkelstein S. L., et al., 2021, JWST Proposal ID 2079. Cycle 1 GO/Treasury
Finkelstein S. L., et al., 2022, ApJ, 928, 52
Gabrielpillai A., Somerville R. S., Genel S., Rodriguez-Gomez V., Pandya V., Yung L. Y. A., Hernquist L., 2021, arXiv:2111.03077
Gardner J. P., et al., 2006, Space Science Reviews, 123, 485
Giavalisco M., et al., 2004, ApJ, 600, L93
Gnedin N. Y., Kravtsov A. V., 2011, ApJ, 728, 88
Grogin N. A., et al., 2011, ApJS, 197, 35
Han S., Smith R., Choi H., Cortese L., Catinella B., Contini E., Yi S. K., 2018, ApJ, 866, 78
Harikane Y., et al., 2016, ApJ, 821, 123
Hirschmann M., Charlot S., Feltre A., Naab T., Choi E., Ostriker J. P., Somerville R. S., 2017, MNRAS, 472, 2468
Hirschmann M., Charlot S., Feltre A., Naab T., Somerville R. S., Choi E., 2019, MNRAS, 487, 333
Illingworth G. D., et al., 2013, ApJS, 209, 6
Kakos J., Primack J. R., Rodríguez-Puebla A., Tejos N., Yung L. Y. A., Somerville R. S., 2022, MNRAS, 514, 1857
Kartaltepe J., et al., 2021, JWST Propos. ID 1727. Cycle 1 GO/Treasury
Kimm T., et al., 2009, MNRAS, 394, 1131
Klypin A., Yepes G., Gottlöber S., Prada F., Heß S., 2016, MNRAS, 457, 4340
Koekemoer A. M., et al., 2011, ApJS, 197, 36
Koekemoer A. M., et al., 2013, ApJS, 209, 3
Landy S. D., Szalay A. S., 1993, ApJ, 412, 64
Livermore R. C., Finkelstein S. L., Lotz J. M., 2017, ApJ, 835, 113
Lovell C. C., Vijayan A. P., Thomas P. A., Wilkins S. M., Barnes D. J., Irodotou D., Roper W., 2020, MNRAS, 500, 2127
Madau P., Ferguson H. C., Dickinson M. E., Giavalisco M., Steidel C. C., Fruchter A., 1996, MNRAS, 283, 1388
Marshall M. A., et al., 2022, arXiv:2206.08941
Maulbetsch C., Avila-Reese V., Colin P., Gottlober S., Khalatyan A., Steinmetz M., 2007, ApJ, 654, 53
McDonough B., Brainerd T., 2022, arXiv:2206.00045
Moster B. P., Somerville R. S., Newman J. A., Rix H.-W., 2011, ApJ, 731, 113
Moster B. P., Naab T., White S. D. M., 2013, MNRAS, 428, 3121
Moster B. P., Naab T., White S. D. M., 2018, MNRAS, 477, 1822
Navarro J. F., Frenk C. S., White S. D. M., 1997, ApJ, 490, 493
Nelson D., et al., 2019, Comput. Astrophys. Cosmol., 6, 2
Newman J. A., et al., 2013, ApJS, 208
Oesch P. A., Bouwens R. J., Illingworth G. D., Labbé I., Stefanon M., 2018, ApJ, 855, 105
Oke J. B., Gunn J. E., 1983, ApJ, 266, 713
Pacifici C., Charlot S., Blaizot J., Brinchmann J., 2012, MNRAS, 421, 2002
Pacifici C., et al., 2016, ApJ, 832, 79
Pandya V., et al., 2020, ApJ, 905, 4
Peebles P. J. E., 1980, in , Large-Scale Struct. Universe. Princeton University Press, Princeton, https://ui.adsabs.harvard.edu/abs/1980lssu.book.....P
Planck Collaboration 2016, A&A, 594, A13
Popping G., Somerville R. S., Trager S. C., 2014, MNRAS, 442, 2398
Price-Whelan A. M., et al., 2018, AJ, 156, 123
Qiu Y., et al., 2018, MNRAS, 481, 4885
Reback J., et al., 2022, pandas-dev/pandas: Pandas, doi:10.5281/zenodo.6408044, https://doi.org/10.5281/zenodo.6408044
Robertson B. E., 2022, ARA&A, 60, 121
Robitaille T. P., et al., 2013, A&A, 558, A33
Rodriguez-Gomez V., et al., 2019, MNRAS, 483, 4140

Rodríguez-Puebla A., Behroozi P., Primack J., Klypin A., Lee C., Hellinger D., 2016, MNRAS, 462, 893

Rodríguez-Puebla A., Primack J. R., Avila-Reese V., Faber S. M., 2017, MNRAS, 470, 651

Scoville N., et al., 2007, ApJS, 172, 1

Sinha M., Garrison L. H., 2020, MNRAS, 491, 3022

Skibba R. A., et al., 2015, ApJ, 807, 152

Smith R., Choi H., Lee J., Rhee J., Sanchez-Janssen R., Yi S. K., 2016, ApJ, 833, 109

Snyder G. F., Lotz J. M., Rodriguez-Gomez V., da Silva Guimarães R., Torrey P., Hernquist L., 2017, MNRAS, 468, 207

Snyder G. F., Rodriguez-Gomez V., Lotz J. M., Torrey P., Quirk A. C., Hernquist L., Vogelsberger M., Freeman P. E., 2019, MNRAS, 486, 3702

Somerville R. S., Davé R., 2015, ARA&A, 53, 31

Somerville R. S., Kolatt T. S., 1999, MNRAS, 305, 1

Somerville R. S., Primack J. R., 1999, MNRAS, 310, 1087

Somerville R. S., Lemson G., Kolatt T. S., Dekel A., 2000, MNRAS, 316, 479

Somerville R. S., Lee K., Ferguson H. C., Gardner J. P., Moustakas L. A., Giavalisco M., 2004, ApJ, 600, L171

Somerville R. S., Hopkins P. F., Cox T. J., Robertson B. E., Hernquist L., 2008, MNRAS, 391, 481

Somerville R. S., Gilmore R. C., Primack J. R., Domínguez A., 2012, MNRAS, 423, 1992

Somerville R. S., Popping G., Trager S. C., 2015, MNRAS, 453, 4338

Somerville R. S., et al., 2021, MNRAS, 502, 4858

Stefanon M., et al., 2019, ApJ, 883, 99

Tacchella S., et al., 2022, ApJ, 927, 170

Torrey P., et al., 2015, MNRAS, 447, 2753

Trenti M., Stiavelli M., 2008, ApJ, 676, 767

Virtanen P., et al., 2020, Nat. Methods, 17, 261

Vogelsberger M., et al., 2020, MNRAS, 492, 5167

Wang H. Y., Mo H. J., Jing Y. P., 2007, MNRAS, 375, 633

Wechsler R. H., Tinker J. L., 2018, ARA&A, 56, 435

Wechsler R. H., DeRose J., Busha M. T., Becker M. R., Rykoff E., Evrard A., 2022, ApJ, 931, 145

Williams C. C., et al., 2018, ApJS, 236, 33

Yung L. Y. A., Somerville R. S., Finkelstein S. L., Popping G., Davé R., 2019a, MNRAS, 483, 2983

Yung L. Y. A., Somerville R. S., Popping G., Finkelstein S. L., Ferguson H. C., Davé R., 2019b, MNRAS, 490, 2855

Yung L. Y. A., Somerville R. S., Popping G., Finkelstein S. L., 2020a, MNRAS, 494, 1002

Yung L. Y. A., Somerville R. S., Finkelstein S. L., Popping G., Davé R., Venkatesan A., Behroozi P., Ferguson H. C., 2020b, MNRAS, 496, 4574

Yung L. Y. A., Somerville R. S., Finkelstein S. L., Hirschmann M., Davé R., Popping G., Gardner J. P., Venkatesan A., 2021, MNRAS, 508, 2706

Yung L. Y. A., et al., 2022, in preperation

Zehavi I., et al., 2004, ApJ, 608, 16

Zehavi I., et al., 2011, ApJ, 736

Zheng Z., Coil A. L., Zehavi I., 2007, ApJ, 667, 760

van der Walt S., Colbert S. C., Varoquaux G., 2011, Comput. Sci. Eng., 13, 22

APPENDIX A: TABULATED VALUES FOR SELECTED TWO-POINT CORRELATION FUNCTIONS

We provide tabulated data for selected angular two-point correlation functions with limiting magnitude $m_{\rm F200W,lim} \sim$ 29 and ~ 31 in Tables A1 and A2, respectively.

APPENDIX B: TABULATED VALUES FOR COSMIC SFR

We provide tabulated data for the predicted cosmic SFR presented in Fig. 13. We provide values from the median (red solid line) and the 84th and 16th percentile (red dashed lines).

This paper has been typeset from a TEX/LATEX file prepared by the author.

Table A1. This tables provide the tabulated angular correlation function computed from the wide-field lightcones for galaxies $m_{\rm F200W,lim} \sim$ 29. We provide the median value and the 16th and 84th percentile across the 40 wide-field lightcones.

			$\log w(\theta)$			
$\log(\theta/\mathrm{arcsec})$	$z = 3.50 - 4.50$	$z = 4.50 - 5.50$	$z = 5.50 - 6.50$	$z = 6.50 - 7.50$	$z = 8.00 - 9.00$	$z = 9.00 - 10.00$
0.47	$-0.472^{+0.036}_{-0.028}$	$-0.243^{+0.036}_{-0.039}$	$0.014^{+0.042}_{-0.054}$	$0.307^{+0.052}_{-0.056}$	$0.814^{+0.152}_{-0.201}$	
0.67	$-0.569^{+0.028}_{-0.036}$	$-0.371^{+0.035}_{-0.025}$	$-0.151^{+0.045}_{-0.048}$	$0.09^{+0.034}_{-0.03}$	$0.562^{+0.113}_{-0.075}$	
0.87	$-0.739^{+0.027}_{-0.036}$	$-0.559^{+0.048}_{-0.037}$	$-0.344^{+0.045}_{-0.039}$	$-0.089^{+0.045}_{-0.058}$	$0.327^{+0.132}_{-0.105}$	$0.684^{+0.219}_{-0.109}$
1.07	$-0.918^{+0.036}_{-0.036}$	$-0.716^{+0.047}_{-0.049}$	$-0.507^{+0.045}_{-0.059}$	$-0.291^{+0.068}_{-0.054}$	$0.107^{+0.162}_{-0.113}$	$0.452^{+0.206}_{-0.089}$
1.27	$-1.089^{+0.071}_{-0.023}$	$-0.858^{+0.057}_{-0.056}$	$-0.652^{+0.036}_{-0.064}$	$-0.467^{+0.062}_{-0.042}$	$-0.072^{+0.096}_{-0.121}$	$0.201^{+0.151}_{-0.11}$
1.47	$-1.23^{+0.096}_{-0.035}$	$-0.993^{+0.066}_{-0.086}$	$-0.815^{+0.054}_{-0.053}$	$-0.617^{+0.08}_{-0.099}$	$-0.31^{+0.207}_{-0.102}$	$-0.008^{+0.115}_{-0.162}$
1.67	$-1.346^{+0.107}_{-0.061}$	$-1.121^{+0.099}_{-0.123}$	$-0.955^{+0.076}_{-0.069}$	$-0.789^{+0.111}_{-0.105}$	$-0.488^{+0.186}_{-0.145}$	$-0.277^{+0.153}_{-0.271}$
1.87	$-1.482^{+0.149}_{-0.109}$	$-1.263^{+0.121}_{-0.131}$	$-1.1^{+0.06}_{-0.107}$	$-0.972^{+0.149}_{-0.143}$	$-0.71^{+0.229}_{-0.31}$	$-0.557^{+0.23}_{-0.404}$
2.07	$-1.643^{+0.19}_{-0.123}$	$-1.402^{+0.143}_{-0.179}$	$-1.293^{+0.09}_{-0.109}$	$-1.187^{+0.205}_{-0.176}$	$-0.876^{+0.231}_{-0.388}$	$-0.756^{+0.173}_{-0.371}$
2.27	$-1.825^{+0.179}_{-0.142}$	$-1.552^{+0.227}_{-0.306}$	$-1.527^{+0.14}_{-0.173}$	$-1.452^{+0.297}_{-0.271}$	$-1.152^{+0.431}_{-0.691}$	$-0.969^{+0.206}_{-1.526}$
2.47	$-2.043^{+0.272}_{-0.504}$	$-1.733^{+0.251}_{-0.493}$	$-1.804^{+0.219}_{-0.328}$	$-1.77^{+0.459}_{-0.669}$	$-1.453^{+0.457}_{-0.799}$	

Table A2. This tables provide the tabulated angular correlation function computed from the ultra-deep lightcones for galaxies $m_{\rm F200W,lim} \sim$ 31. We provide the median value and the 16th and 84th percentile across the set of eight lightcone realizations.

			$\log w(\theta)$			
$\log(\theta/\mathrm{arcsec})$	$z = 3.50 - 4.50$	$z = 4.50 - 5.50$	$z = 5.50 - 6.50$	$z = 6.50 - 7.50$	$z = 8.00 - 9.00$	$z = 9.00 - 10.00$
0.27	$-0.326^{+0.043}_{-0.021}$	$-0.377^{+0.012}_{-0.036}$	$-0.468^{+0.117}_{-0.112}$	$-0.324^{+0.041}_{-0.04}$	$0.104^{+0.232}_{-0.072}$	
0.47	$-0.716^{+0.016}_{-0.017}$	$-0.728^{+0.06}_{-0.054}$	$-0.59^{+0.111}_{-0.147}$	$-0.304^{+0.08}_{-0.034}$	$0.097^{+0.042}_{-0.063}$	$0.438^{+0.146}_{-0.292}$
0.67	$-0.986^{+0.026}_{-0.083}$	$-0.849^{+0.047}_{-0.064}$	$-0.569^{+0.046}_{-0.091}$	$-0.375^{+0.032}_{-0.095}$	$-0.009^{+0.106}_{-0.095}$	$0.279^{+0.154}_{-0.162}$
0.87	$-1.156^{+0.044}_{-0.041}$	$-0.915^{+0.074}_{-0.07}$	$-0.723^{+0.109}_{-0.128}$	$-0.551^{+0.092}_{-0.036}$	$-0.209^{+0.062}_{-0.107}$	$-0.057^{+0.176}_{-0.036}$
1.07	$-1.254^{+0.031}_{-0.059}$	$-1.051^{+0.105}_{-0.123}$	$-0.904^{+0.13}_{-0.035}$	$-0.67^{+0.019}_{-0.156}$	$-0.383^{+0.08}_{-0.124}$	$-0.212^{+0.123}_{-0.183}$
1.27	$-1.413^{+0.075}_{-0.039}$	$-1.182^{+0.117}_{-0.182}$	$-1.041^{+0.123}_{-0.081}$	$-0.832^{+0.042}_{-0.102}$	$-0.545^{+0.095}_{-0.164}$	$-0.427^{+0.096}_{-0.096}$
1.47	$-1.527^{+0.087}_{-0.037}$	$-1.309^{+0.077}_{-0.195}$	$-1.168^{+0.097}_{-0.119}$	$-0.968^{+0.046}_{-0.243}$	$-0.752^{+0.101}_{-0.106}$	$-0.67^{+0.106}_{-0.08}$
1.67	$-1.654^{+0.059}_{-0.077}$	$-1.489^{+0.118}_{-0.197}$	$-1.357^{+0.19}_{-0.021}$	$-1.138^{+0.108}_{-0.268}$	$-0.973^{+0.163}_{-0.122}$	$-0.884^{+0.097}_{-0.142}$
1.87	$-1.847^{+0.155}_{-0.085}$	$-1.782^{+0.321}_{-0.245}$	$-1.491^{+0.174}_{-0.345}$	$-1.271^{+0.044}_{-0.627}$	$-1.35^{+0.252}_{-0.216}$	$-1.45^{+0.4}_{-0.28}$
2.07	$-2.148^{+0.354}_{-0.173}$	$-1.93^{+0.31}_{-1.384}$	$-1.763^{+0.374}_{-0.381}$	$-1.582^{+0.082}_{-0.379}$		

Table B1. This tables provide the tabulated cosmic SFR predicted by the Santa Cruz SAM. We provide the median value and 16th and 84th percentile across the 40 wide-field lightcones (see Fig. 13) between $z = 2$ to 10 and the same quantity calculated from a grid run between $z = 11$ to 15 from Paper IV. These predictions are made with the same fiducial model.

z	This work $\log(\text{CSFR}/\,\text{M}_\odot\,\text{yr}^{-1}\,\text{Mpc}^{-3})$	z	This work $\log(\text{CSFR}/\,\text{M}_\odot\,\text{yr}^{-1}\,\text{Mpc}^{-3})$	z	Paper IV $\log(\text{CSFR}/\,\text{M}_\odot\,\text{yr}^{-1}\,\text{Mpc}^{-3})$
$2.0 - 2.5$	$-1.01^{+0.05}_{-0.05}$	$6.0 - 6.5$	$-1.69^{+0.06}_{-0.11}$	11	-3.94
$2.5 - 3.0$	$-0.98^{+0.03}_{-0.04}$	$6.5 - 7.0$	$-1.86^{+0.06}_{-0.02}$	12	-4.63
$3.0 - 3.5$	$-0.99^{+0.07}_{-0.05}$	$7.0 - 7.5$	$-2.09^{+0.05}_{-0.06}$	13	-5.31
$3.5 - 4.0$	$-0.99^{+0.05}_{-0.04}$	$7.5 - 8.0$	$-2.26^{+0.07}_{-0.15}$	14	-6.49
$4.0 - 4.5$	$-1.09^{+0.04}_{-0.05}$	$8.0 - 8.5$	$-2.51^{+0.05}_{-0.05}$	15	-7.73
$4.5 - 5.0$	$-1.25^{+0.06}_{-0.05}$	$8.5 - 9.0$	$-2.74^{+0.08}_{-0.10}$		
$5.0 - 5.5$	$-1.36^{+0.05}_{-0.04}$	$9.0 - 9.5$	$-2.98^{+0.12}_{-0.05}$		
$5.5 - 6.0$	$-1.54^{+0.07}_{-0.09}$	$9.5 - 10.0$	$-3.26^{+0.11}_{-0.19}$		