Mock Observations for the CSST Mission: Main Surveys-The Mock Catalog Postprint

Authors: Cheng-Liang Wei, Yu Luo, Hao Tian, Ming Li, Yi-Sheng Qiu, Guo-Liang Li, Yue-Dong Fang, Xin Zhang, De-Zi Liu, Nan Li, Ran Li, Huan-Yuan Shan, Lin Nie, Zizhao He, Lei Wang, Xi Kang, Dongwei Fan, Yang Chen, Xiaoting Fu, Chao Liu

Date: 2026-01-28T11:17:06+00:00

Abstract

The Chinese Space Station Survey Telescope (CSST) is a flagship space mission, supported by the China Manned Space Project, designed to carry

Full Text

Mock Observations for the CSST Mission: Main Surveys-The Mock Catalog

Cheng-Liang Wei1, Yu Luo2, Hao Tian3, Ming Li3, Yi-Sheng Qiu4,5, Guo-Liang Li1, Yue-Dong Fang6, Xin Zhang3, De-Zi Liu7, Nan Li3, Ran Li3,8, Huan-Yuan Shan9, Lin Nie9, Zizhao He1, Lei Wang1, Xi Kang4, Dongwei Fan3, Yang Chen10, Xiaoting Fu1, and Chao Liu3

1 Purple Mountain Observatory, Chinese Academy of Sciences, Nanjing 210023, China
2 School of Physics and Electronics, Hunan Normal University, Changsha 410081, China; luoyupmo@gmail.com
3 Key Laboratory of Space Astronomy and Technology, National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100101, China; tianhao@nao.cas.cn
4 Zhejiang University-Purple Mountain Observatory Joint Research Center for Astronomy, Zhejiang University, Hangzhou 310027, China
5 Research Center for Astronomical Computing, Zhejiang Laboratory, Hangzhou 311121, China
6 Universitäts-Sternwarte München, Fakultät für Physik, Ludwig-Maximilians-Universität München, Scheinerstrasse 1, 81679 München, Germany
7 School of Physics and Optoelectronic Engineering, Anhui University, Hefei 230601, China

Received 2025 May 28; revised 2025 August 16; accepted 2025 September 15; published 2026 January 6

Abstract

The Chinese Space Station Survey Telescope (CSST) is a flagship space mission, supported by the China Manned Space Project, designed to carry out a large-area sky survey to explore the nature of dark matter and dark energy in the Universe. The onboard multi-band imaging and slitless spectroscopic modules will enable us to obtain photometric data for billions of galaxies and stars, as well as hundreds of millions of spectroscopic measurements, advancing various scientific analyses such as galaxy clustering and weak gravitational lensing. To support the image simulations for the main survey of the CSST mission, we present a mock catalog of stars and galaxies. For stars, the mock catalog is generated using either Galaxia or TRILEGAL, both of which provide a range of stellar properties to meet the requirements of CSST image simulations. For galaxies, we built a mock light-cone up to redshift $z \sim 3.5$ from the cosmological N-body simulation and populated the mock galaxy catalog from the dark matter haloes using a semi-analytical galaxy formation model. We then performed a full-sky ray-tracing simulation of weak gravitational lensing to obtain lensing shear at the position of each galaxy in the light-cone. To support both multi-band imaging and slitless spectroscopic simulations, we computed the spectral energy distribution (SED) for each galaxy based on its star formation history using a supervised deep learning model and determined the magnitudes in each band using the CSST throughputs. Finally, the properties of our mock galaxies include positions, redshifts, stellar masses, shapes, sizes, SEDs, lensing shears and magnifications. We have validated our mock catalog against observational data and theoretical models, with results showing good overall agreement. The catalog provides a flexible data set for the development of CSST image processing and can support a wide range of cosmological analyses within the CSST mission.

Key words: catalogs -gravitational lensing: weak -methods: numerical

1. Introduction

To accurately understand the evolution of matter in the Universe, astronomers rely on large-area galaxy surveys to study the growth of cosmic structures and uncover the nature of the accelerated expansion of the Universe [?, ?, ?, ?]. Current and upcoming galaxy surveys, such as the Vera C. Rubin Observatory [?, ?] and Euclid [?, ?], will generate vast amounts of high-quality data, enabling unprecedented precision in constraining cosmological parameters, especially the equation of state of dark energy [?, ?].

The Chinese Space Station Survey Telescope (CSST), a flagship project of the Stage-IV galaxy surveys, will carry out a 10 yr space mission to perform a large-area galaxy survey with $ 17,500deg^2$ of wide field and $ 400deg^2$ of deep field [?]. CSST will determine galaxy shapes by imaging billions of galaxies within 7-bands (NUV, u, g, r, i, z, y), reaching a magnitude limit of $m_g \sim 26.3$

in wide field and $ $1.2 mag fainter in the deep field. Simultaneously, its slitless spectroscopic module will provide low-resolution spectra for hundreds of millions of galaxies to determine their redshifts (as outlined in the CSST science book). These rich datasets will support precise analyses of galaxy clustering and weak gravitational lensing, delivering robust cosmological constraints [?, ?, ?, ?].

As statistical accuracy improves in these Stage-IV surveys, systematic errors will be critically important for maximizing the scientific returns of these current and future galaxy surveys [?, ?]. It is essential to develop an end-to-end imaging simulator to optimize the data processing and conduct preliminary scientific analyses [?, ?, ?]. For CSST, a public release of the imaging simulator[1] (version 3.1.0) has been developed by the CSST scientific data processing and analysis system to simulate multi-band imaging and slitless spectroscopic observations [?]. To produce highly realistic mock images, an optical emulator has been developed to simulate high-fidelity point-spread functions (PSFs) of CSST [?] and various sources of noise have been included, such as shot noise, sky background, and comprehensive detector effects. To achieve this, we utilize Galsim[2] [?] to generate photons from a given object, taking into account the throughputs of the CSST filter system. These throughputs encompass the mirror efficiency, filter transmission, and quantum efficiency of the detector. For more details of the CSST imaging simulator we refer the reader to [?]. With the CSST imaging simulator, we have released a set of 50 $\deg^2$ data products, which has been used to develop the pipeline of data processing and optimize the survey design of CSST.

This paper provides a detailed description of the input mock catalog used in the CSST main survey simulation. The structure is organized as follows. An overview of the input mock catalog is shown in Section 2. In Sections 3 and 4, we describe the production and primary characteristics of the star and galaxy catalogs, respectively. Section 5 presents the computation of weak lensing properties by a full-sky ray-tracing simulation. Section 6 details the validation of the properties of the mock catalog against observational constraints and theoretical models. Although tailored for the CSST simulation, the catalog can be very useful for various studies due to its comprehensive inclusion of galaxy properties and consistently calculated weak lensing observables down to sub-arcminute scales. Finally, Section 7 provides a summary of our mock catalog.

2. Overview of the Mock Catalog

The input mock catalog for the recent release comprises stars, galaxies and quasars [Figure 1: see original paper]. To support development of the imaging simulator and various scientific studies, it is of great importance that the catalog includes a range of realistic features. For stars, we generate the field star catalog on demand using either Galaxia [?] or TRILEGAL [?], two widely used stellar population synthesis models for the Milky Way. Both models can generate mock

---

[1] https://csst-tb.bao.ac.cn/code/csst-sims/csst_{{{msc}}_{{sim}}}

[2] https://github.com/GalSim-developers/GalSim

star catalogs that effectively support the general requirements of CSST image simulation.

For galaxies, several methods can be employed to generate high-precision galaxy catalogs [?, ?, ?, ?]. Hydrodynamical simulations, which model complex baryonic feedback processes within galaxies, can directly produce galaxy samples with detailed physical properties, such as morphology and star formation rates [?, ?, ?]. However, current hydrodynamical simulations are computationally expensive and rely on subgrid physics models [?, ?]. A more efficient and widely used alternative is based on high-precision N-body simulation. This approach, adopted in cosmoDC2 [?] and Flagship [?], involves constructing dark matter light-cones using large-volume and high-mass resolution cosmological simulations. Dark matter haloes are identified within the light-cone and galaxies are populated using algorithms such as semi-analytical galaxy formation models (SAM) or halo occupation distribution models (HOD).

Regarding CSST, we have conducted the "JiuTian"simulations [?], a series of cosmological N-body simulations with varying box sizes and resolutions, designed to support mission optimization and scientific analysis preparation. Given the wide-field and deep observation capabilities of CSST, the mass of dark matter particle in the simulation should be $\sim 10^8 M_\odot$ to support studies of large-scale structure and weak gravitational lensing. In particular, the JiuTian-1G simulation (hereafter JT1G), with a box size of 1 h^{-1} Gpc, is utilized to construct a full-sky light-cone covering redshifts up to $z_s \sim 3.5$ for the CSST imaging simulator. Galaxies are generated using SAM [?] and weak lensing signals are incorporated through a full-sky ray-tracing simulation [?]. Specifically, to meet the needs of spectral analysis, STARDUSTER[3] [?] has been used to generate the spectral energy distribution (SED) for each galaxy in the catalog. Below, we provide a detailed description of the input mock catalogs used in the CSST imaging simulator.

3. Star Catalog

Field stars from the Milky Way are essential for the imaging simulation, especially for a wide field survey. In the CSST imaging simulator, spectra are required for all sources. However, given the observational depth and the vast number of stars, storing pre-computed spectra for the entire field would be prohibitively demanding in terms of storage. To address this, we adopt a strategy in which only the stellar catalog is stored as input, while spectra are generated on the fly based on stellar parameters and apparent magnitudes. This requires the effective temperature $T_{\rm eff}$, surface gravity $\log g$, metallicity Z, and magnitudes to be provided. Furthermore, to ensure astrometry function, the positions and velocities of all stars must also be included. For CSST, two widely used full-sky mock stellar catalogs are currently available and have been employed at different stages of development.

The first is the Milky Way stellar mock catalog generated using the TRILEGAL

---

[3] https://github.com/yqiuu/starduster

stellar population synthesis tool [?]. This catalog includes all classical components (the bulge, the disk and the halo). The luminosity function and the mass distribution are generally consistent with observational results [?]. The TRILEGAL catalog provides detailed information of each star, including stellar parameters (effective temperature $T_{\rm eff}$, surface gravity $\log g$, metallicity Z, age τ), the position (α, δ, d), the velocity relative to the Sun (v_U, v_V, v_W) and magnitudes in CSST photometric system (NUV, u, g, r, i, z, y), SDSS photometric system $(u_s, g_s, r_s, i_s, z_s)$ and PAN-STARRS photometric system $(g_p, r_p, i_p, z_p, y_p)$. What should be noticed is that only the stars with apparent magnitude brighter than 27.5 in CSST g-band are included. The dataset is accessible to all CSST simulator users via the Virtual Observatory platform[4]. Almost all spectral types of stars are included while the interacting binary systems are not included for current simulations.

The second model is Galaxia [?], which is able to generate a synthetic survey of the Milky Way based on customizable parameters. Users can define the magnitude limit, sky coverage and output photometric system via a configuration file[5]. Similar to TRILEGAL, the output catalog includes stellar magnitudes in the selected photometric system, as well as stellar parameters and chemical information. However, Galaxia does not enforce a fixed magnitude cutoff, but instead adheres to the magnitude limits defined in the configuration file. Another notable difference is that Galaxia does not provide photometric magnitudes in the CSST filter system. Instead, we use the magnitudes available in the SDSS bands. The magnitudes in other photometric systems are recalculated from the spectra generated based on the stellar parameters.

Nonetheless, both TRILEGAL and Galaxia can support the requirements of CSST imaging simulations, as they provide the stellar parameters $(T_{\rm eff}, \log g, Z)$ to reproduce the spectrum for each star, which is necessary for the simulator. At the current stage, we adopt Galaxia, which is able to generate complete stellar samples down to the faint end for all CSST bands. As an example, the input star catalog used for the current release is generated using Galaxia. At this stage, crowded stellar systems, such as globular clusters and dwarf galaxies, are not included in the mock catalog, but these larger systems will be incorporated in future versions.

Figure 2 [Figure 2: see original paper] shows the stellar density distribution in equatorial coordinates for the catalog covering the North Ecliptic Pole (NEP). Given that the NEP lies at a relatively low Galactic latitude ($b \sim 29.81$), a distinct increase in stellar density appears in the lower-right region of the figure, corresponding to areas closer to the Galactic plane.

4.1. The Cosmological Dark Matter Simulation

The "JiuTian" simulations are a series of ΛCDM cosmological N-body simulations with different simulation box sizes and resolutions, specifically designed

---

[4] https://nadc.china-vo.org/data/data/csst-trilegal/f

[5] https://galaxia.sourceforge.net/Galaxia3pub.html

for CSST [?]. The JT1G is one of the "JiuTian" simulations with both large volume and high resolution. It features a 1 h^{-1} Gpc box size and 6144^3 particles, which is 8 times larger volume and almost 27 times more particles than the Millennium simulation [?]. The corresponding dark matter particle mass is $3.72 \times 10^8 h^{-1} M_\odot$. For the purpose of weak lensing studies, JT1G stores 128 snapshots from redshift 127 to 0, with an average time interval of approximately 100 Myr. The adopted cosmological parameters are based on Planck2020 results [?]: $\sigma_8 = 0.8102$, $H_0 = 67.66$ km s^{-1} Mpc^{-1}, $\Omega_\Lambda = 0.6889$, $\Omega_m = 0.3111$, $\Omega_b = 0.0490$ ($f_b = 0.1575$).

4.2. Halo Catalog

The JT1G was executed using a lean version of the traditional Tree-PM code Gadget-3 [?]. A significant portion of the typical post-processing work was performed on the fly, including group finding via the Friends-of-Friends (FoF) algorithm [?] and the extraction of bound substructures within each FoF group using the SUBFIND algorithm [?].

The construction of merger trees follows a two-step approach. First, descendant links are established by tracking groups across adjacent snapshots. Once all unique descendant subhaloes are identified, we link them across the full snapshot sequence to construct complete merger trees. This is done by taking a subhalo at $z = 0$ and linking all subhaloes with descendant pointers to this halo, then repeating with all of those subhaloes, and so on, until no more subhaloes can be joined. Galaxies are then generated within these dark matter haloes using a galaxy formation model. Figure 3 [Figure 3: see original paper] provides an illustrative example of our mock galaxy light-cone, where a pixel sky-field of HEALPix[6] with $N_{\rm side} = 64$ is displayed, and one can identify the large-scale structure of the Universe at a given redshift snapshot.

4.3. Semi-analytical Modeling

Galaxies are generated from halo merger trees using a semi-analytical galaxy formation model. In SAMs, galaxy populations are assigned to dark matter haloes based on simplified prescriptions for a wide range of physical processes, including reionization, hot gas cooling and cold gas infall, star formation and metal production, SN feedback, hot gas stripping and tidal disruption in satellites, galaxy mergers, bulge formation, black hole growth, and AGN feedback. A brief overview of these processes is provided below.

In most SAMs, each dark matter halo is assigned a cosmic abundance of baryons. As the halo grows, a significant fraction of baryons is accreted as primordial diffuse gas. This gas is shock-heated and then either cools rapidly onto the central galaxy's disk or is added to a quasi-static hot atmosphere, which accretes at a slower rate via cooling flows. The cold gas disk fuels star formation, and when some stars die, they release energy, mass, and heavy elements into the surrounding medium. This energy reheats the cold disk gas, injecting it into

---

[6] https://healpix.sourceforge.io/

the hot atmosphere, while the hot atmosphere itself may be ejected into an external reservoir, only to be reincorporated much later.

Once a satellite crosses its host’ s virial radius, several environmental processes come into play. Tidal forces can strip both hot and cold gas and stars, while ram pressure stripping can remove gas. These processes gradually suppress star formation, especially in satellites orbiting more massive systems. As dark matter subhaloes merge, their associated galaxies merge as well, albeit with some delay. Once a subhalo is completely disrupted, its galaxy spirals into the central galaxy and merges after a dynamical friction time, forming a bulge and triggering a burst of star formation. Bulges can also form through a long-term process whenever the disk becomes dynamically unstable.

Black holes are believed to grow primarily through the accretion of cold gas during mergers, but also through static accretion of the hot atmosphere, releasing energy that can counteract cooling flows. This form of feedback eventually suppresses star formation in the most massive haloes. Finally, the light emitted by stellar populations of different ages is calculated using a population synthesis model and corrected for dust extinction.

In this project, we adopt the semi-analytical model developed by [?], which is based on the [?] and [?] model, a version of the Munich Semi-analytical model called L-Galaxies. [?] improved the prescription for low-mass galaxies, especially satellite galaxies, by including additional physics of cold gas stripping and an analytical modeling of orphan galaxies. For more details, we refer readers to [?, ?, ?, ?, ?].

We introduce two major modifications for generating mock galaxy images. The first involves the galaxy size modeling. Similar to most SAMs, the [?] model assumes that each galaxy consists of two stellar components, an exponential disk and a $r^{1/4}$-law bulge. The disk size is proportional to the ratio of specific angular momentum to maximum circular velocity. The growth of the bulge is mainly driven by mergers, with the bulge size being determined through energy conservation and the virial theorem for both minor and major mergers [?]. These generated galaxy sizes are in rough agreement with the SDSS mass-size relation, but show larger scatters for low mass galaxies. Accurately modeling the redshift evolution of the mass-size relation is also challenging.

Therefore, we use the mass-size relation fitting formula for SDSS galaxies from [?] and consider a certain redshift evolution to describe the size of a given galaxy:

$$\log_{10} R_{50} = \alpha \log_{10} M_* + \beta \log_{10}\left(1 + \frac{M_*}{M_0}\right) + \gamma + \epsilon \log_{10}(1+z),$$

where R_{50} is the half mass-radius of the galaxy, M_* is stellar mass and z is redshift of the galaxy. For spiral galaxies, $\alpha = 0.18$, $\beta = 0.78$, $\gamma = 6.34$, $M_0 = 1.57 \times 10^{11} h^{-2} M_\odot$, $\epsilon = -0.6$; for elliptical galaxies, $\alpha = 0.14$, $\beta = 0.71$, $\gamma = 1.53$, $M_0 = 1.72 \times 10^{11} h^{-2} M_\odot$, $\epsilon = -0.6$. Each galaxy is generated with a

Gaussian dispersion with $\sigma = 0.3$. In Figure 4 [Figure 4: see original paper], we show the mass-size relation of spiral galaxies and elliptical galaxies from $z = 2$ to $z = 0$ using Equation (1).

Another improvement is the approach used to generate the SED for each galaxy. Typically, in SAMs, the SED of a galaxy is generated by applying a stellar population synthesis (SPS) procedure, such as [?], with star formation history and geometry parameters of galaxy. To optimize calculation speed and storage, we employ the STARDUSTER [?] model to generate the high-resolution SEDs, which are then compressed using principal component analysis (PCA). STARDUSTER is a supervised deep learning model trained using SKIRT radiative transfer simulation [?] which can represent features of dust attenuation and emission features in galaxies. We input the star formation history, metallicity, size, and inclination angle of each galaxy from our mock galaxy catalog into STARDUSTER. We employ interpolation to resample the galaxy' s rest-frame SED at uniform intervals of 0.29 nm across the wavelength range 60–1100 nm to meet the requirements for both PCA and slitless spectroscopy simulations. With a small sample of SEDs, we generate 20 principal components (PCs) and decompose each galaxy' s SED using these 20 PCs. Therefore, the SED of each galaxy is represented by 20 coefficients (w_i), from which the full SED can be reconstructed using

$$F(\lambda) = \sum_i w_i PC_i(\lambda).$$

To assess the accuracy of the SED reconstruction, we randomly select about 10,000 galaxies at redshifts 0, 1, 2. As shown in Figure 5 [Figure 5: see original paper], the comparison between the reconstructed and original SEDs shows that the relative deviation is below 2%, indicating a high level of fidelity in the PCA-based compression.

The quasar catalog is also generated based on the galaxy catalog. We use SIMQSO[7] [?] to generate quasar SEDs. SIMQSO is a toolset that generates mock quasar spectra using a broken power-law continuum model along with Gaussian emission line templates. We calculate the total number of quasars in a target sky area using quasar luminosity functions. Specifically, we integrate the quasar luminosity functions from [?] over redshift bins of $\Delta z = 0.1$ ranging from $z = 0.1$ to $z = 3.5$ to determine the total number of quasars within the target sky area. We then generate the SEDs of individual quasars based on their luminosities. Finally, we assign these quasars to galaxies in the same redshift bins and match their luminosities to the AGN accretion rates of the corresponding galaxies from the galaxy catalog. In Figure 6 [Figure 6: see original paper], we present examples of quasar SEDs generated using SIMQSO.

5. Weak Lensing Maps

---

[7] https://simqso.readthedocs.io/en/latest

To produce weak lensing maps from a cosmological simulation, we employ a multi-plane algorithm on spherical geometry. Below, we briefly summarize the key procedures of our weak lensing simulation and more details can be found in [?].

We first cut the simulation box of JT1G into a collection of small cubic boxes with 100 h^{-1} Mpc on each side and assemble the cell boxes to cover a past light-cone from $z = 0$ to $z_{\rm max} \sim 3.5$. Next, the light-cone is decomposed into a series of spherical shells of a given width, 50 h^{-1} Mpc, around the observer, and the dark matter particles are projected onto their corresponding shells, pixelized by the HEALPix tessellation [?]. In this work, the HEALPix resolution is set to $N_{\rm side} = 8192$, which corresponds to an angular resolution of 0.43 arcmin. The projected surface mass densities are computed with an Epanechnikov kernel for each shell and the smoothing length for each particle is set to a few N-body softening lengths. Finally we obtain a surface matter overdensity $\Sigma^{(n)}$ on the nth shell by

$$\Sigma^{(n)}(\theta^{(n)}) = \frac{1}{d\Omega} \sum_i m_i K \left(\frac{|\theta^{(n)} - \theta_i^{(n)}|}{h_i} \right),$$

where K is the Epanechnikov kernel, m_i is the particle mass, and h_i is the smoothing length. The convergence field is $\kappa^{(n)}(\theta^{(n)}) = W^{(n)}\Sigma^{(n)}(\theta^{(n)})$, where the lensing kernel function $W^{(n)}$ is defined as

$$W^{(n)} = \frac{3}{2}\Omega_m \frac{H_0^2}{c^2} \frac{D_A^{(n)} D_A^{(n,m)}}{D_A^{(m)}} (1 + z^{(n)}).$$

Using the Poisson equation, the lensing potential $\phi^{(n)}$ and deflection field $\alpha^{(n)} = \nabla\phi^{(n)}$ can be derived from the mass density of a shell in spherical harmonic space

$$\phi_{\ell m}^{(n)} = \frac{2}{\ell(\ell+1)} \kappa_{\ell m}^{(n)},$$

where $\kappa_{\ell m}^{(n)}$ are the spherical harmonic coefficients of the convergence $\kappa^{(n)}$.

After obtaining the deflection field on each lens plane, we can perform ray-tracing simulation between different lens planes. Light rays are initialized at the center of each HEALPix cell in the first shell and are then propagated to the next lens plane according to

$$\hat{n}^{(n+1)} = \hat{n}^{(n)} + \alpha^{(n)},$$

where $\hat{n}^{(n)}$ denotes the radial direction of light rays on the nth shell. The lensing distortion matrix can be calculated at the position of light rays by

$$\mathcal{A}_{ij}^{(n+1)} = \mathcal{A}_{ij}^{(n)} + D^{(n)}\mathcal{U}_{ij}^{(n)},$$

where $D^{(n)}$ is the comoving angular diameter distance, and $\mathcal{U}_{ij}^{(n)}$ denotes the deformation matrix in the nth lens plane, which can be calculated as

$$\mathcal{U}_{ij}^{(n)} = \frac{\partial^2 \phi^{(n)}}{\partial\theta_i \partial\theta_j}.$$

Figure 7 [Figure 7: see original paper] displays a patch of 10×10 deg^2 from the all-sky map of the convergence field for sources at $z_s = 1.0$. In the figure, the short sticks are overlapped to indicate the lensing shear in the field. As expected, galaxies are stretched tangentially around convergence peaks by the gravitational lensing. As a basic validation of the lensing maps, Figure 8 [Figure 8: see original paper] shows the measurement of the convergence angular power spectrum in the simulation (solid) compared with theory predictions from Born approximation (gray dashed) and Halofit model (black dashed, [?]), for sources at $z_s = 1$. The power spectra of the shear E-mode (blue) and B-mode (magenta) are also included for comparison. The lower panels show the relative errors between the simulation and theoretical predictions.

6.1. Mock Star Properties

During the selection of the star catalog, we apply a magnitude cut to ensure the sample’ s completeness. Since Galaxia does not provide data in the CSST photometric system, we adopt the SDSS system and keep all stars with an apparent magnitude of $g_{\rm SDSS} < 30$. This threshold ensures completeness down to the limiting magnitudes across all seven CSST bands, as the color indices of different stellar types vary significantly. Figure 10 [Figure 10: see original paper] shows the luminosity distributions in the u, g, i, z bands of all the stars and the stars from the thin disk, the thin+thick disk and the thin+thick+halo are represented by different colors in the panels from top left to the bottom right. The vertical dashed lines indicate the CSST limiting magnitudes in the $u, g, i,$ and z bands, which are 25.4, 26.3, 25.9, and 25.2, respectively. The distributions are not always increasing exponentially versus the magnitude, because the contribution from the thin and thick disks are limited within a few kiloparsec. In contrast, the halo, being the most diffuse component, extends from a few parsecs to beyond 100 kpc. Consequently, its luminosity function closely follows an exponential distribution with magnitude.

Regarding the stellar populations, the mock Milky Way catalog includes different populations corresponding to the components, i.e., the thin disk, the thick disk and the halo from young to old. Figure 11 [Figure 11: see original paper] shows the distributions of the stars in the Kiel diagram, with the left and right panels color-coded by metallicity and age, respectively. This catalog covers a wide range of stellar populations, from metal-poor stars with $Z \sim 10^{-6}$ to metal-rich stars with $Z \sim 0.1$, from hot stars with $T_{\rm eff} \sim 14,000$ K to cool stars with

$T_{\text{eff}} \sim 2000$ K. What should be noticed is that certain special stellar types, such as chemically enhanced stars, are not included in this dataset.

6.2. Mock Galaxy Properties

In general, several model parameters in SAMs can be adjusted to better match observations. In this work, we constrain our model using stellar and cold gas mass functions at $z = 0$. Figure 12 [Figure 12: see original paper] presents a comparison of our galaxy catalog' s stellar mass function (left panel), HI mass function (middle panel) and H_2 mass function (right panel) with observational data (yellow symbols) from [**?**, **?**, **?**, **?**, **?**] at $z = 0$.

Our mock galaxy catalog includes key parameters such as redshift, position, velocity, halo mass, stellar mass (separated into disk and bulge mass), galaxy type (central or satellite), size, and SED represented by 20 coefficients. In the Appendix, we present statistical analyses, including the conditional stellar mass function, the evolution of the stellar mass function, and the stellar mass-halo mass relation from $z = 2$ to 0, comparing them with the results derived from observational data. Our mock catalog can reproduce the fundamental stellar mass and halo mass relation of various galaxy populations.

The galaxy size and morphology are two important properties to generate galaxy mock images. Normally we use B/T which is the ratio of bulge mass and total stellar mass to define the galaxy morphology [**?**]. In Figure 13 [Figure 13: see original paper] we show three galaxy population distributions with different morphology at $z = 0$. Here, we define elliptical galaxies with $B/T > 0.5$, spiral galaxies with $0.1 < B/T < 0.5$ and pure disk galaxies with $B/T < 0.1$. Our results reproduce observations roughly [**?**, **?**]. We then assign an intrinsic shape to each galaxy using the mass distribution of its dark matter halo. For an early-type central galaxy, its major axis is assumed to align with that of the host dark matter halo. For a late-type central galaxy, its spin follows that of the halo, with the major axis determined by projecting the circular disk onto the sky. For satellite galaxies, we assign their shapes using a random orientation. A detailed description of this model can be found in [**?**].

In Figure 14 [Figure 14: see original paper], we show the mass-size relation of our galaxy catalog at $z = 0$. We simply divide a small sample of our galaxy catalog into two populations, spiral galaxies with $B/T < 0.5$ (blue dots), elliptical galaxies with $B/T > 0.5$ (red dots). Solid lines represent the median R_{50}, while the corresponding dashed lines indicate the 1σ error. The open dots are from [**?**]. Since we use the fitting formula from [**?**], it can reproduce the relation of the observational data.

The magnitude of each galaxy is computed by convolving its SED with a given filter. In Figure 15 [Figure 15: see original paper], we present the luminosity functions of our mock galaxy catalog in u, g, r, i, z bands of SDSS at $z = 0$, which closely match SDSS observational data (black symbols; [**?**]). The only exception is the u-band, where our catalog contains a higher number of bright galaxies, likely due to the stellar population synthesis (SPS) model used.

6.3. Weak Lensing 2-point Statistics

Weak lensing can result in small correlated distortions of galaxy shapes, which can be analyzed using various statistical methods [?, ?, ?, ?]. A standard statistic used in weak lensing analysis is the shear 2-point correlation function [?]. The shear 2-point correlation between galaxies at separation ϑ is usually defined as

$$\xi_{tt}^{(ij)}(\vartheta) = \frac{\sum_{ab} w_i w_j \epsilon_t^i \epsilon_t^j}{\sum_{ab} w_i w_j},$$

where ϵ_t and $\epsilon_\times$ are the tangential and cross-component of weak lensing shear, w_i is a weight of the i-th galaxy. A more commonly used form is the linear combination of ξ_{tt} and $\xi_{\times\times}$,

$$\xi_{\pm}^{(ij)}(\vartheta) = \xi_{tt}^{(ij)}(\vartheta) \pm \xi_{\times\times}^{(ij)}(\vartheta),$$

which can be related to the power spectrum of convergence field by

$$\xi_{\pm}(\vartheta) = \int \frac{d\ell}{2\pi} P_\kappa(\ell) J_{0,4}(\ell\vartheta),$$

where $J_{0,4}$ denotes the zeroth and fourth Bessel function for ξ_+ and ξ_-, respectively.

In this section, we validate the weak lensing 2-point statistics by generating a mock catalog modeled on the photometric redshifts of DES Year-3 measurements (DES-Y3, [?]), which provides the largest sky coverage among the currently available dataset of Stage-3 weak lensing surveys. We adopt the same number density as DES-Y3 to populate the mock galaxies and Figure 16 [Figure 16: see original paper] shows the redshift distributions of the mock catalog. The test catalog covers an area of \$ $4200deg^{2,withanumberdensityof5.9galaxiesperarcmin}2$\$. Here we do not aim to exactly reproduce the DES-Y3 catalog. Instead, we construct a mock catalog with comparable galaxy number density and redshift binning, and directly compare the resulting galaxy shape correlation functions with those from DES-Y3.

Figure 17 [Figure 17: see original paper] shows the tomographic shear correlations $\xi_\pm$ from the mock catalog. We calculate the correlation functions using the public code TreeCorr[8] [?], where the superscript (ij) denotes the different redshift bins used for the calculation of the correlation function. First, we compare the shear-only correlation (GG, red dots) with the theoretical predictions (solid lines) derived from the halofit model using Nicaea[9] [?], using the redshift

---

[8] https://github.com/rmjarvis/TreeCorr

[9] https://github.com/CosmoStat/nicaea

distribution of galaxies in the light-cone. It can be seen that the correlations of lensing shear from the simulation are consistent with the predictions. Due to the presence of intrinsic ellipticity of galaxies in our catalog, we then measure the ellipticity correlation (GG+II+GI, black circles), which can be directly compared with the observational results from DES-Y3 (blue circles). Overall, the tomographic correlations in the mock catalog are slightly higher than those in DES-Y3. To quantify the difference between the model and the data, we calculate the reduced χ^2 following the definition in [?] and find the reduced $\chi^2 \sim 1.63$ between our mock data and DES-Y3 measurements [?]. This difference is consistent with the fact that the underlying cosmology in the JT1G simulation [?] is based on Planck results, with $S_8 = 0.827 \pm 0.019$, which is about 1.5σ higher than the DES-Y3 constraint $S_8 = 0.776 \pm 0.017$. In this comparison, we do not account for the effects of redshift uncertainties. These systematics, including shape noise and photometric redshift uncertainties, are incorporated in our CSST data challenge project[10], where more than five types of systematics have been modeled. A more detailed analysis that fully accounts for these effects will be presented in future work.

7. Summary

As one of the Stage-IV galaxy surveys, CSST will use galaxy clustering and weak gravitational lensing to constrain cosmological parameters and unravel the cause of the accelerated expansion of the Universe. In the current definition stage, it is crucial to construct a mock catalog, which includes a wide range of realistic features, for optimizing the data processing pipeline and assessing scientific performance. In this paper a mock catalog has been presented, consisting of stars, galaxies and quasars, for the CSST main survey simulator and could be used to support the analyses of CSST galaxy clustering and weak gravitational lensing.

We produce the star catalog using either Galaxia or TRILEGAL. The catalog provides various properties of each star, including the position, velocity, stellar age, metallicity, and chemical information. When using TRILEGAL, the catalog additionally includes magnitudes in the CSST photometric system. It is shown that the luminosity function and the mass distribution are generally consistent with observations.

The mock galaxy catalog is generated from the halo catalog using a galaxy formation model of SAM. We utilize JT1G, a cosmological N-body simulation, to construct the merger tree of dark matter haloes, assuming galaxies form at the centers of the dark matter haloes according to analytical prescriptions of relevant physical processes, such as gas cooling, star formation, supernova, and black hole feedback. To characterize galaxy morphology, we use B/T to classify different types of galaxies and assume that the shape of an elliptical galaxy roughly follows that of its host dark matter halo, while the rotation axis of a spiral galaxy is aligned with the spin of its halo. For each galaxy in the mock

---

[10] https://nadc.china-vo.org/events/CSSTdatachallenge

catalog, its full SED is generated using the STARDUSTER deep learning model, trained on SKIRT radiative transfer simulations. The magnitudes in each band are then determined with the throughputs of CSST. Additionally, quasars are assigned to galaxies by matching their luminosities with the AGN accretion rates in the galaxy catalog. The quasar catalog is constructed using the luminosity functions of [**?**] and SEDs for quasars are generated with SIMQSO.

With the mock galaxy catalog, we construct a light-cone up to redshift $z \sim 3.5$ over the full sky and perform the weak lensing ray-tracing simulation on the curved sky with a HEALPix resolution of 0.43 arcmin. This enables us to derive the convergence, shear and magnification at each galaxy’ s position within the light-cone. We validate the mock galaxy catalog by comparing it with observational data. Our results generally reproduce the stellar mass-halo mass relation at low redshifts but show lower stellar masses for larger halo masses at $z = 2$, consistent with the stellar mass function at that redshift. The size distributions of the mock galaxies show reasonable agreement with observations and our mock catalog includes more bright galaxies than SDSS data, likely due to the stellar population synthesis used in our data production pipeline. Additionally, we check the weak lensing properties of the mock catalog. The measured power spectra of convergence and shear E-/B-mode align well with expectations from the Halofit model. As an example of the data production pipeline, we generate a DES-Y3-like mock catalog and compare the shear correlation with observational data and halofit predictions, finding good agreement with the data.

Although our mock catalog includes comprehensive properties, some improvements can be further explored. In the catalog, galaxies are modeled as a two-component system, bulge and disk. While this is sufficient for some analyses, incorporating more complex morphologies would enhance the catalog’ s applicability, especially for validating data processing pipelines with imaging data. Another potential refinement is the calculation of galaxy SEDs. Currently, SEDs are computed with random incidence angles, which lacks accuracy for individual galaxies. A more precise method would compute SEDs based on the actual position and orientation of each galaxy within the light-cone. Although this is computationally expensive, it could significantly improve the precision of SEDs.

Acknowledgments

This work is supported by the project of the CSST scientific data processing and analysis system of the China Manned Space Project. We thank the working group of CSST- “JiuTian” to make the cosmological simulations available.

This work is supported by the National Natural Science Foundation of China (NSFC, No. 11903082) and the China Manned Space Program with grant Nos. CMS-CSST-2025-A20, CMS-CSST-2025-A05 and CMS-CSST-2021-A03. Y.C. acknowledges the support by the Natural Science Research Project of Anhui Educational Committee No. 2024AH050049, NSFC No. 12003001 and the Anhui Project (Z010118169).

Appendix

Statistics of Stellar Mass and Halo Mass

In SAM models, stellar mass is a crucial parameter that can be directly compared with observational data. We present statistical results of the conditional stellar mass function, the evolution of the stellar mass function, and the stellar mass-halo mass relation, comparing them with observational data. Our mock catalog can reproduce the fundamental relation between stellar mass and halo mass.

Figure A1 (upper panels) shows the evolution of stellar mass function of our galaxy catalog from $z = 2$ to 0 (black solid lines). The blue symbols are the combined observational data from [**?**]. The stellar mass of our galaxy catalog at $z = 2$ is slightly lower at the high mass end, while higher at the low mass end.

In Figure A1 (lower panels), we show the evolution of stellar mass-halo mass relations from $z = 2$-0 and compare to the analytical result from [**?**]. The stellar mass-halo mass relation is well reproduced by our results at low redshifts. However, at $z = 2$, the model predicts lower stellar masses in massive haloes, in agreement with the corresponding stellar mass function.

The conditional stellar mass function indicates the mass distribution of central galaxies and satellite galaxies in certain mass haloes. In Figure A2 we show the conditional stellar mass functions of our galaxy catalog at $z = 0$ which reproduce the observation data [**?**] especially in massive halo mass bins.

References

[Abbott2022] Abbott, T. M. C., Aguena, M., Alarcon, A., et al. 2022, PhRvD, 105, 023520 [Angulo2012] Angulo, R. E., Springel, V., White, S. D. M., et al. 2012, MNRAS, 426, 2046

[Ban2026] Ban, Z., Li, X., Yang, X., et al. 2026, RAA, 26, 024002 [Barrera2023] Barrera, M., Springel, V., White, S. D. M., et al. 2023, MNRAS, 525, 6312

[Chen2023] Chen, Y., Fu, X., Liu, C., et al. 2023, SCPMA, 66, 119511 [Conselice2006] Conselice, C. J. 2006, MNRAS, 373, 1389 [Crain2015] Crain, R. A., Schaye, J., Bower, R. G., et al. 2015, MNRAS, 450, 1937 [Dalal2023] Dalal, R., Li, X., Nicola, A., et al. 2023, PhRvD, 108, 123519

[Doux2022] Doux, C., Jain, B., Zeurcher, D., et al. 2022, MNRAS, 515, 1942 [Driver2012] Driver, S. P., Robotham, A. S. G., Kelvin, L., et al. 2012, MNRAS, 427, 3244

[Fu2012] Fu, J., Kauffmann, G., Li, C., & Guo, Q. 2012, MNRAS, 424, 2701 [Fu2013] Fu, J., Kauffmann, G., Huang, M.-l., et al. 2013, MNRAS, 434, 1531 [Fu2014] Fu, L., Kilbinger, M., Erben, T., et al. 2014, MNRAS, 441, 2725

[Gorski2005] Górski, K. M., Hivon, E., Banday, A. J., et al. 2005, ApJ, 622, 759 [Guo2011] Guo, Q., White, S., Boylan-Kolchin, M., et al. 2011, MNRAS, 413, 101 [Guo2013] Guo, Q., White, S., Angulo, R. E., et al. 2013, MNRAS, 428, 1351 [Habouzit2021] Habouzit, M., Li, Y., Somerville, R. S., et al. 2021, MNRAS, 503, 1940 [Han2025] Han, J., Li, M., Jiang, W., et al. 2025, arXiv:2503.21368 [Henriques2015] Henriques, B. M. B., White, S. D. M., Thomas, P. A., et al. 2015, MNRAS, 451, 2663 [Ivezic2019] Ivezić, Ž., Kahn, S. M., Tyson, J. A., et al. 2019, ApJ, 873, 111 [Jarvis2015] Jarvis, M., 2015 TreeCorr: Two-point Correlation Functions, Astrophysics Source Code Library, ascl:1508.007 [Kaviraj2017] Kaviraj, S., Laigle, C., Kimm, T., et al. 2017, MNRAS, 467, 4739

[Korytov2019] Korytov, D., Hearin, A., Kovacs, E., et al. 2019, ApJS, 245, 26 [Laureijs2011] Laureijs, R., Amiaux, J., Arduini, S., et al. 2011, arXiv:1110.3193 [Li2009] Li, C., & White, S. D. M. 2009, MNRAS, 398, 2177

[LSST2009] LSST Science Collaboration, Abell, P. A., Allison, J., et al. 2009, arXiv:0912.0201

[Martin2010] Martin, A. M., Papastergis, E., Giovanelli, R., et al. 2010, ApJ, 723, 1359

[Merson2013] Merson, A. I., Baugh, C. M., Helly, J. C., et al. 2013, MNRAS, 429, 556 [Miao2023] Miao, H., Gong, Y., Chen, X., et al. 2023, MNRAS, 519, 1132

[Plazas2020] Plazas Malagón, A. A. 2020, Symm, 12, 494

[Ross2013] Ross, N. P., McGreer, I. D., White, M., et al. 2013, ApJ, 773, 14

[Sanchez2020] Sánchez, J., Walter, C. W., Awan, H., et al. 2020, MNRAS, 497, 210 [Schaye2023] Schaye, J., Kugel, R., Schaller, M., et al. 2023, MNRAS, 526, 4978 [Shan2018] Shan, H., Liu, X., Hildebrandt, H., et al. 2018, MNRAS, 474, 1116

[Song2024] Song, Y., Xiong, Q., Gong, Y., et al. 2024, ApJ, 976, 244

[Springel2005] Springel, V., White, S. D. M., Jenkins, A., et al. 2005, Natur, 435, 629

[Torrey2015] Torrey, P., Snyder, G. F., Vogelsberger, M., et al. 2015, MNRAS, 447, 2753 [Vogelsberger2014] Vogelsberger, M., Genel, S., Springel, V., et al. 2014, MNRAS, 444, 1518 [Wei2018] Wei, C., Li, G., Kang, X., et al. 2018, ApJ, 853, 25 [Wei2026] Wei, C., Li, G., Fang, Y., et al. 2026, RAA, 26, 024001 [Xiong2024] Xiong, Q., Gong, Y., Zhou, X., et al. 2025, The Astrophysical Journal, 985, 131

[Yao2024] Yao, J., Shan, H., Li, R., et al. 2024, MNRAS, 527, 5206 [Zhan2021] Zhan, H. 2021, ChSBu, 66, 1290

[Zhang2022] Zhang, Z. J., Chang, C., Larsen, P., et al. 2022, MNRAS, 514, 2181

Note: Figure translations are in progress. See original paper for figures.