





Improving Photometric Redshift Estimation for CSST Mock Catalog Using SED Templates Calibrated with Perturbation Algorithm

Yicheng Li¹, Liping Fu^{1,2}, Zhu Chen¹, Zhijian Luo¹, Wei Du¹, Yan Gong^{3,4}, Xianmin Meng⁴, Junhao Lu¹ , Zhirui Tang¹, Pengfei Chen¹, Shaohua Zhang¹ , Chenggang Shu¹, Xingchen Zhou³, and Zuhui Fan⁵

¹ Shanghai Key Lab for Astrophysics, Shanghai Normal University, Shanghai 200234, China; fuliping@shnu.edu.cn, zhuchen@shnu.edu.cn, zjluo@shnu.edu.cn

² Center for Astronomy and Space Sciences, China Three Gorges University, Yichang 443000, China

³ Key Laboratory of Space Astronomy and Technology, National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100101, China

⁴ Science Center for China Space Station Telescope, National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100101, China

⁵ South-Western Institute for Astronomy Research, Yunnan University, Kunming 650500, China

Received 2025 February 26; revised 2025 April 8; accepted 2025 April 10; published 2025 May 21

Abstract

Photometric redshifts of galaxies obtained by multi-wavelength data are widely used in photometric surveys because of their high efficiency. Although various methods have been developed, template fitting is still adopted as one of the most popular approaches. Its accuracy strongly depends on the quality of the spectral energy distribution (SED) templates, which can be calibrated using broadband photometric data from galaxies with known spectroscopic redshifts. Such calibration is expected to improve photometric redshift accuracy, as the calibrated templates will align with observed photometric data more closely. The upcoming China Space Station Telescope (CSST) is one of the Stage IV surveys, which is aiming for high precision cosmological studies. To improve the accuracy of photometric redshift estimation for CSST, we calibrated the CWW+KIN templates using a perturbation algorithm with broadband photometric data from the CSST mock catalog. This calibration used a training set consisting of approximately 4500 galaxies, which is 10% of the total galaxy sample. The outlier fraction and scatter of the photometric redshifts derived from the calibrated templates are 2.55% and 0.036, respectively. Compared to the CWW+KIN templates, these values are reduced by 34% and 23%, respectively. This demonstrates that SED templates calibrated with a small training set can effectively optimize photometric redshift accuracy for future large-scale surveys like CSST, especially with limited spectral training data.

Key words: galaxies: distances and redshifts – galaxies: photometry – methods: statistical

1. Introduction

Estimating the redshifts of astronomical objects is essential for advancing extragalactic and cosmological studies (see Salvato et al. 2019; Newman & Gruen 2022 for a review). Redshifts can be determined using two primary methods: spectroscopic and photometric observations. Spectroscopic redshifts (spec- z) are obtained by measuring wavelength shifts in spectral line features. While this method provides high accuracy, obtaining spec- z for faint, distant objects is challenging and time consuming (Newman et al. 2015). In contrast, photometric redshifts (photo- z) are estimated using multi-band photometric data, which capture broadband features such as the Lyman and Balmer breaks. Although photo- z sacrifices some accuracy compared to spectroscopic methods due to the limited resolution of photometric data, it is significantly more efficient, enabling redshift estimation from fewer exposures across all detected sources. This efficiency makes photo- z a practical choice for many observational analyses.

There are two primary methods for deriving photo- z : machine learning and the spectral energy distribution (SED)

fitting method. In principle, any regression based machine learning technique can be applied to estimate photo- z , including classical support vector machines (Wadadekar 2005; Jones & Singal 2017), decision tree based bagging methods such as random forests, or boosting algorithms (e.g., Carliles et al. 2010; Dalmaso et al. 2020; Zhou et al. 2021). Additionally, simple neural networks have been widely used (e.g., Firth et al. 2003; Collister & Lahav 2004; Cavuoti et al. 2017; Razim et al. 2021; Zhou et al. 2021).

Machine learning techniques typically use magnitudes and colors as input features, but can also incorporate more digitized features correlated with redshift, such as morphological parameters, to improve estimation accuracy (D’Isanto & Polsterer 2018; Gomes et al. 2018). Furthermore, combining digitized features with photometric images to train hybrid deep neural networks (Zhou et al. 2022) or employing advanced architectures like Long Short Term Memory (LSTM) networks (Luo et al. 2024a) has shown potential for achieving even higher precision.

However, machine learning methods have inherent limitations. Their accuracy strongly depends on the quality and size

of the training sample. Moreover, the lack of spectroscopic data for faint galaxies in the training set limits their ability to estimate photo- z accurately through extrapolation.

The basic idea of SED fitting is to find the best match between observed multi-band photometric data and redshifted galaxy SED templates. These templates can be derived from observations (e.g., Coleman et al. 1980; Kinney et al. 1996; Assef et al. 2010) or from stellar population synthesis models (e.g., Bruzual & Charlot 1993, 2003; Fioc & Rocca-Volmerange 1997; Maraston 2005). Many established SED fitting photo- z estimators (e.g., HyperZ, Bolzonella et al. 2000; *LePhare*, Arnouts et al. 1999; BPZ, Benítez 2000; EAZY, Brammer et al. 2008) can utilize either type of template set. These methods can provide redshift estimates as long as the redshifted template SEDs cover the relevant wavelength range of the observed filters.

A key limitation of SED fitting is the representativeness of the template set for a given observational data set. Stellar population synthesis models involve numerous physical assumptions, and empirical templates often derived from local, bright galaxies may not capture the full range of galaxy properties across different epochs. As a result, evolutionary effects at higher redshifts can introduce additional uncertainties, reducing the accuracy of SED fitting approaches.

The current and next generation of large scale surveys, such as the Dark Energy Survey (DES, Abbott et al. 2018), the Vera C. Rubin Observatory Legacy Survey of Space and Time (Ivezić et al. 2008; Abell et al. 2009), the Euclid Survey (Laureijs et al. 2011; Euclid Collaboration et al. 2025), and the Roman Space Telescope Survey (Green et al. 2012; Akeson et al. 2019), will provide unprecedented amounts of photometric data. For many extragalactic objects in these data sets, obtaining spec- z will be impractical. To fully utilize these extensive data sets, it is crucial to improve and thoroughly assess the performance of photo- z methods when applied to these surveys (Newman & Gruen 2022).

The same considerations apply to the China Space Station Telescope (CSST; Zhan 2011; Zhan 2018; Cao et al. 2018). The CSST will employ seven photometric bands, ranging from near-ultraviolet (NUV) to near-infrared (NIR), namely NUV, u , g , r , i , z and y , that cover wavelengths of approximately 2500–10,000 Å. The 5σ magnitude limits for these bands are 25.4, 25.4, 26.3, 26.0, 25.9, 25.2, and 24.4, respectively. During its 10 yr mission, CSST will survey a total of 17,500 deg² of the sky, performing photometric and spectroscopic observations simultaneously. One of its primary objectives is to test theoretical cosmological models through weak gravitational lensing, which heavily relies on accurate redshift estimations.

This paper is part of a series that evaluates the performance of various photo- z techniques to meet the scientific goals of the CSST using a mock CSST catalog. Lu et al. (2023) estimate

photo- z s from CSST mock fluxes using a weighted random forest method. Luo et al. (2024b) employ a deep learning method called generative adversarial imputation networks (GAIN) to impute missing photometric data, thus reducing the impact of data gaps on photo- z estimation and improving precision. Luo et al. (2024a) propose a new approach that relies solely on flux measurements from different observed filters to accurately predict CSST photo- z by constructing a deep learning model based on Recurrent Neural Networks (RNN) with LSTM units.

In this work, we focus on improving the accuracy of photo- z derived from the CSST mock catalog using SED fitting and calibrated SED templates (i.e., a template learning approach). Because the actual SEDs of galaxies may differ from the initial input templates, calibration of these templates is essential. Several studies have addressed template calibration and optimization (Budavári et al. 2000; Benítez et al. 2004; Crenshaw & Connolly 2020). Here, we employ the template perturbation algorithm introduced in Crenshaw & Connolly (2020, hereafter CC20) and demonstrate that it significantly improves photo- z accuracy for the CSST mock catalog, even with a very small training sample.

In Section 2, we briefly introduce the CSST mock catalog and the initial SED templates used for calibration. Section 3 describes the perturbation algorithm from CC20. Details of the calibration process are presented in Section 4. Section 5 compares the photo- z results obtained using the SED templates before and after calibration. Finally, our conclusions are summarized in Section 6.

2. Data

A CSST mock catalog was employed to calibrate the SED templates, evaluating the improvement in photo- z performance achieved by applying calibrated SED templates within the SED fitting algorithm. This catalog has also been used in previous CSST photo- z assessment studies (Zhou et al. 2021, 2022; Lu et al. 2023; Luo et al. 2024a, 2024b), which explored both machine learning and SED fitting methods.

In contrast to traditional approaches that rely on the mock catalogs generated from the light cone of dark matter N -body simulations, our mock catalog is derived from a CSST-like galaxy image simulation. The simulation process is summarized below, and more detailed information can be found in Zhou et al. (2022) and Lu et al. (2023). First, HST/F814W images from COSMOS were rescaled to match the CSST pixel size, and square stamped images for each galaxy were created. The size of stamps is based on the length of galaxies' semimajor axes. Second, 31 SED templates from *LePhare* (Arnouts & Ilbert 2011, Ilbert et al. 2006) were assigned to each galaxy by fitting the photometric data at the given photo- z s from the COSMOS2015 catalog (Laigle et al. 2016). The theoretical flux calculated by convolving the response function

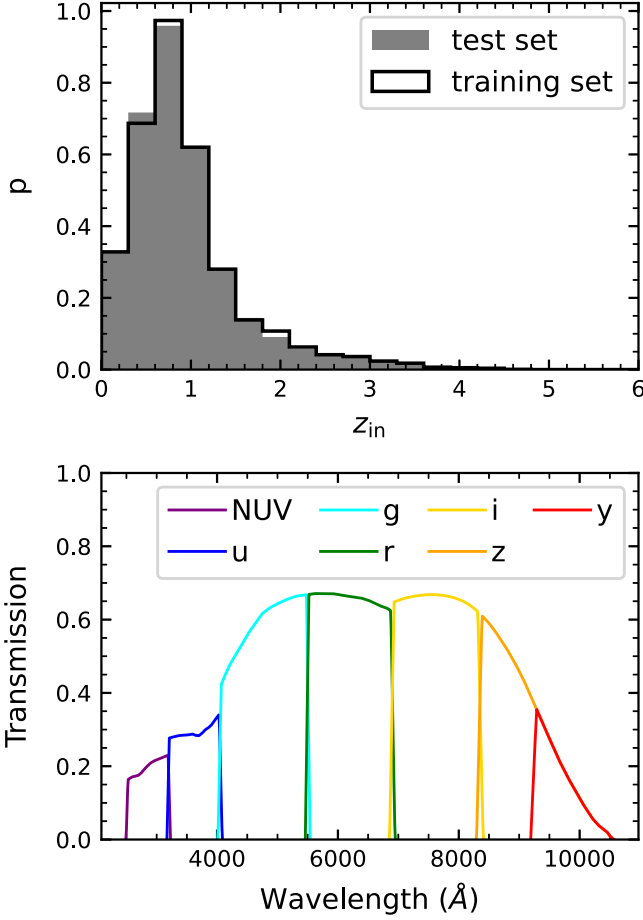


Figure 1. Top: normalized redshift distribution of training set and test set from CSST mock catalog. Bottom: response function of filters at NUV, u , g , r , i , z and y band for CSST.

of CSST filters with SED templates was taken as the simulated flux for each galaxy. Finally, simulated CSST-like images were generated in these seven bands based on the simulated flux, taking into account the exposure time, aperture size, and background noise level of CSST. Forced photometry was then applied to the simulated images to produce the mock catalog used in this work. Figure 1 displays the response function of each filter, with a wavelength coverage from ~ 2500 to $\sim 11,000$ Å.

3. The Perturbation Algorithm

In this section, we provide a brief introduction to the perturbation algorithm described in CC20 and Budavári et al. (2000). The basic idea of the perturbation algorithm is to calibrate the SED templates according to the discrepancies between the photometric data and the SED templates. When redshifts are already known, the photometric data can be assigned to the SED templates with the most similar colors.

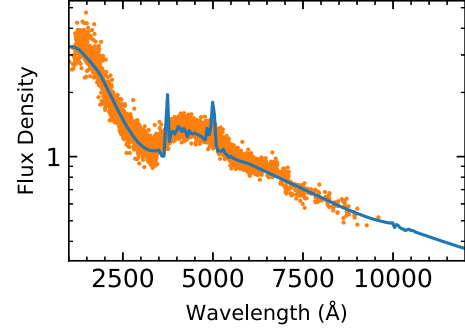


Figure 2. The photometric data of all the galaxies matched to the example SED template. The SED template has been normalized to 5500 Å, and the y-axis is expressed in logarithmic scale.

Figure 2 shows an example in which all photometric data points are overlaid with their best matching SED template in the rest frame. With a sufficiently large number of galaxies in the catalog, the overlapping region of their photometric data points can effectively reconstruct the SED's continuum. As the sample expands, it can even recover higher resolution features of the SED to a considerable degree.

However, discrepancies may exist between the distribution of photometric data and the SED templates. To address this, we calibrate the SED templates by applying perturbations, enabling better alignment between the templates and the photometric data. The template perturbation is done in the rest frame. The flux densities derived by SED templates are

$$F_n^{\text{temp}} = \sum_k s_k r'_{n,k}, \quad (1)$$

where s_k and $r'_{n,k}$ represent the flux density of the SED templates and the normalized filter response function in discrete wavelength bins in the rest frame, respectively.

All galaxies in the training set will be matched to the SED template with the most similar color before perturbation. The galaxies are separated into different groups referred to as the sub-training sets of each SED template. The cost function is in the form of

$$\chi^2 = \sum_i \sum_n \frac{(\hat{F}_{i,n}^{\text{temp}} - F_{i,n}^{\text{obs}})^2}{\sigma_{i,n}^2} + \sum_k \frac{(\hat{s}_k - s_k)^2}{\Delta_k^2}, \quad (2)$$

where $F_{i,n}^{\text{obs}}$ is the normalized observed flux density (see Section 4.1), and $\hat{F}_{i,n}^{\text{temp}}$ represents the flux density derived from the perturbed SED templates, which is denoted as \hat{s}_k . The subscripts i, n indicate band n of the i th galaxy in the sub-training set. The first term of the cost function represents how well the perturbed SED fits its sub-training set, weighted by the fractional error of observed flux densities $\sigma_{i,n}$. The second term is a penalty term, weighted by the parameter Δ_k , which constrains the extent of the shape change during template perturbation and thus helps stabilize the shape of the perturbed

SED. By minimizing this cost function, the shape change of the SED template, defined as $\xi_k = \hat{s}_k - s_k$, can be determined.

For the part corresponding to hyperparameter optimization, CC20 implemented a uniform metric Δ to replace Δ_k in various templates and introduced an innovative parameter w , defined as the ratio of the loss function to the penalty term, thus regulating the pace of training.

$$w = \frac{\sum_k \Delta^{-2} (\hat{s}_k - s_k)^2}{\sum_i \sum_n \sigma_{i,n}^{-2} (\hat{F}_n^{\text{temp}} - F_n^{\text{obs}})^2}, \quad (3)$$

by assuming

$$\frac{\sum_k (\hat{s}_k - s_k)^2}{\sum_i \sum_n (\hat{F}_n^{\text{temp}} - F_n^{\text{obs}})^2} \sim \frac{N_k}{N}, \quad (4)$$

where N is the number of photometric data points and N_k is the number of wavelength bins. Then, Δ can be approximated as

$$\Delta \approx \bar{\sigma} \sqrt{\frac{N_k}{wN}}, \quad (5)$$

where $\bar{\sigma} = \sum_i \sum_n \sigma_{i,n}$.

Since the profile of SED templates changes after every perturbation, the photometric data matching results may differ from the previous perturbation. The algorithm will be iterated multiple times to obtain the best match between the photometric data and the SED templates. Following the methodology outlined in CC20, we use the weighted mean squared error (w MSE) between the flux densities derived from the templates and the observed flux densities to quantify the alignment between the sub-training sets and the SED templates.

$$w\text{MSE} = \frac{1}{N} \sum_i \sum_n \frac{1}{\sigma_{i,n}^2} (F_{i,n}^{\text{temp}} - F_{i,n}^{\text{obs}})^2. \quad (6)$$

The termination of the perturbation process is determined by the relative change in w MSE, denoted as d MSE

$$d\text{MSE} = \left| \frac{w\text{MSE}_0 - w\text{MSE}}{w\text{MSE}_0} \right|. \quad (7)$$

The w MSE is calculated each time the SED template is perturbed or a new sub-training set is created. When d MSE falls below a specific threshold (4% in this work) or reaches the maximum number of perturbations, we consider the SED templates to be sufficiently calibrated for the current iteration. The calibration process will stop when it reaches the maximum number of iterations or no templates are perturbed during the current iteration. For more details on the perturbation algorithm, we refer the reader to the studies of CC20 and Budavári et al. (2000).

4. SED Template Calibration

In this section, we describe in detail the process of calibrating the SED templates using the perturbation algorithm, including the creation of training sets, smoothing of initial

templates to obtain more stable results, and comparisons between the calibrated and initial SED templates.

4.1. Training Set and Initial SED Templates

To improve the precision of photo- z estimation, we applied a signal-to-noise ratio (SNR) cut with a threshold of ≥ 10 in either the g or i bands for both training and test sets. This criterion is consistent with the sample selection used in previous photo- z evaluation studies for the CSST project (Zhou et al. 2021, 2022; Lu et al. 2023; Luo et al. 2024a, 2024b). The final sample consists of 44,991 galaxies.

Instead of training the SED templates from scratch, we calibrate existing SED templates, enabling the use of a relatively small training set. However, utilizing a larger training set when available generally will result in a more stable outcome. In this work, we selected 10% of the galaxies as the training set, which is ~ 4500 galaxies with a maximum redshift of about 4.83, while the remaining 90% of galaxies were the test set. The impact of the size of the training set will be discussed in Section 5. Figure 1 illustrates the redshift distribution for both training and test sets, which represent an ideal scenario where the redshift distributions of the two sets are identical.

Perturbation will be applied to the SED templates based on their sub-training sets. For each galaxy, the photometric data are then matched to the SED template with the smallest χ^2 value, representing the closest color match.

$$\chi^2 = \sum_n \frac{1}{\delta F_n^{\text{obs}}} (F_n^{\text{temp}} - F_n^{\text{obs}})^2, \quad (8)$$

where δF_n^{temp} is the uncertainty of normalized flux densities.

Due to variations in the intrinsic luminosity and distance of galaxies, the observed photometric data f_n^{obs} and their uncertainty δf_n^{obs} must be normalized before matching with the SED templates. Instead of using the median of $F^{\text{temp}}/f^{\text{obs}}$ as the normalization coefficient which is adopted in CC20, we apply the normalization method from the EAZY code, which is formulated as

$$c_{\text{norm}} = \frac{\sum_n F_n^{\text{temp}} f_n^{\text{obs}} / (\delta f_n^{\text{obs}})^2}{\sum_n (F_n^{\text{temp}})^2 / (\delta f_n^{\text{obs}})^2}, \quad (9)$$

where the uncertainties of observed flux density are considered.

Since the template perturbation process relies on sub-training sets of each SED template, different normalization methods result in varied sub-training sets, which in turn lead to variations in the training results. An accurate match between photometric data and SED templates is crucial for creating well-suited sub-training sets for each template, thereby enhancing the accuracy of photometric redshift (photo- z) estimation. The impacts of using different normalization coefficients will be discussed in Section 5.

Table 1

 Photo- z Results for Different SED Template Sets by Using EAZY Single Template Fitting Mode

SED Template	η	σ_{nMAD}
CWW+KIN	3.85%	0.047
tweak_fsps_QSF_12_v3	7.47%	0.052
COSMOS31	6.46%	0.044
EAZY_V1.2_dusty	17.25%	0.073

In principle, the initial templates can be naive log normal curves (CC20) or any set of SED templates. The initial templates should be comprehensive enough to cover a variety of galaxy types. Simultaneously, the number of templates should be minimized to ensure that each template has sufficient photometric data during the calibration process. Table 1 shows the results of photo- z derived from different SED templates,⁶ where the fraction of outliers η and σ_{nMAD} are defined in Section 5. In this work, the CWW + KIN (Coleman et al. 1980; Kinney et al. 1996; Benítez et al. 2004) template set is chosen as the initial SED templates (Figure 3), due to its good photo- z results and a relatively uniform distribution of the number of galaxies across each sub-training set.

4.2. Smoothing SED Templates

A high resolution SED template with dense sampling on the wavelength grid will significantly increase the computational time required for the perturbation algorithm. To simplify the computation, the initial SED template is sampled using a fixed interval wavelength grid. However, for those SED templates with many high resolution features, such as strong emission or absorption lines, this sampling process can greatly impact F_i^{temp} .

Figure 4 illustrates these effects using the SB2_B2004a template (hereafter SB2) as an example, showing the differences in the SED and F_i^{temp} between the original template and a smoothed version. Both templates are sampled with a fixed wavelength step of 100 Å. At 5000 Å, where the peak of the O III emission line aligns with the sampling grid, the flux density derived from the i band (F_i^{temp}) can increase by up to 30%. In addition, at 6500 Å, where the H α emission line is missed due to sampling, the F_i^{temp} is 25% lower than actual.

These discrepancies in F_i^{temp} can lead to overcorrection in the perturbed SED template, as shown in the lower left panel of Figure 4. Although the shape changes in the perturbed SED template effectively compensate for the discrepancies shown in the middle left panel, the presence of negative flux values (at 5000 Å) and an exaggerated, unrealistic absorption feature renders the SED template unusable.

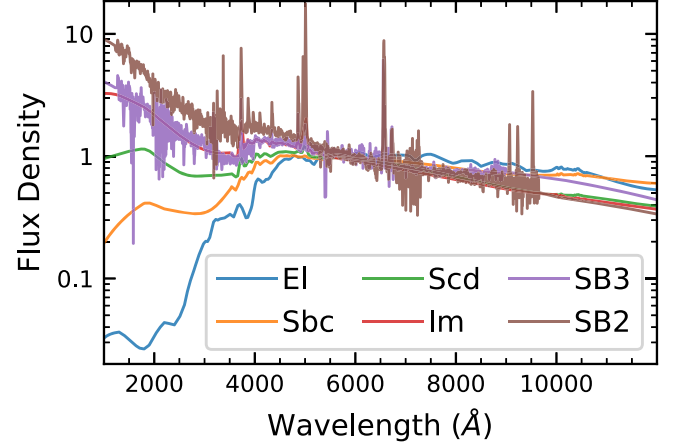


Figure 3. Initial SED templates of CWW+KIN. All templates are normalized at 5500 Å.

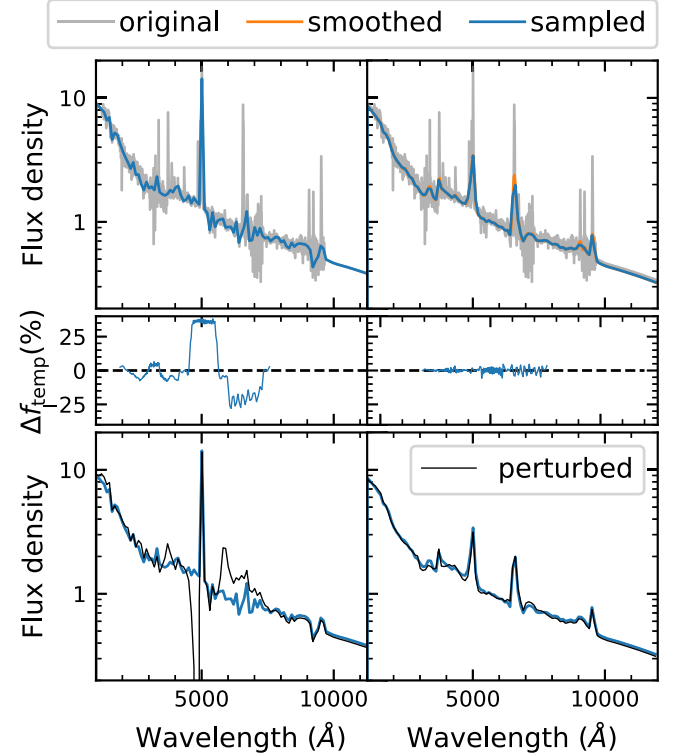


Figure 4. Impacts of sampling the SED templates. The left and right panels display the SED template before and after the smoothing process, respectively. The upper panels show the SED template before and after sampling, represented by the gray and blue curves respectively, with the smoothed template shown in orange in the upper right panel. The middle panels show the relative error of the F_i^{temp} between the initial and sampled SED template at different redshifts. Note that the wavelength here indicates the mean wavelength of i band at different redshifts. The lower panels illustrate the sampled SED template before and after perturbation, shown in blue and black curves, respectively. All panels share the same x-axis, and the horizontal panels share the same y-axis.

⁶ Before calibrating the SED templates.

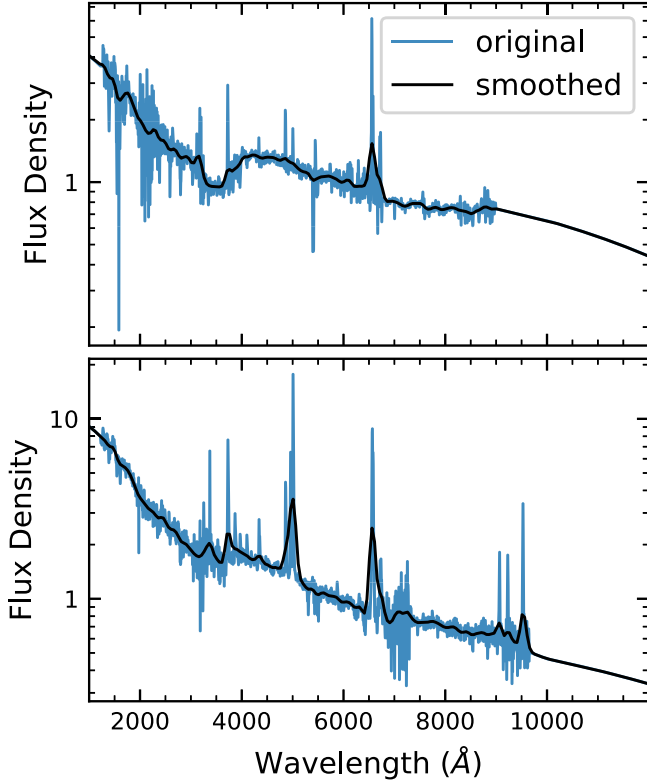


Figure 5. Two starburst templates before and after smoothing: SB3_B2004a (top) and SB2_B2004a (bottom). The blue and black curves represent the original and smoothed templates, respectively, with all templates normalized at 5500 Å.

To avoid overcorrection, one approach is to modify the cost function by introducing an additional penalty term that further constrains changes in the SED shape, as suggested in Feldmann et al. (2006, ZEBRA). Increasing the sampling grid density, smoothing the templates before calibration or smoothing the templates before perturbation can also effectively reduce the overcorrection. In this work, we choose to smooth the SB2 template using a one-dimensional Gaussian filter with a σ of 100 Å.

As shown in the right panels of Figure 4, the flux densities of the smoothed SED templates change smoothly near emission lines. The difference in F_i^{temp} between the sampled and smoothed initial templates is significantly reduced, and no overcorrection is observed in the perturbed SB2 template. Two starburst templates, SB3_B2004a and SB2_B2004a, due to their strong emission lines and other dense small features, are smoothed before calibration (Figure 5).

4.3. Calibrating SED templates

We utilized the template perturbation algorithm described in Section 3 to calibrate the CWW+KIN templates, employing photometric data sourced from the CSST mock catalog as detailed in Section 2. According to CC20, the training results are robust to w , and the CWW + KIN templates are calibrated

with $w = 1.0$, which is similar with the w used in the work of CC20. Given that the final few iterations have minimal impact on the SED templates (as shown in Figure 8), we set a maximum of three perturbations per iteration and a maximum of five iterations for the calibration process. The threshold $d\text{MSE}$ is set to 4%. In Figure 6, we present the initial SED templates together with their corresponding sub-training sets. The distribution of galaxies among the sub-training sets appears relatively uniform for each SED template, ranging from approximately 1000 galaxies to at least 500 galaxies per template. Most templates show similar initial $w\text{MSE}$, approximately 5. However, the initial $w\text{MSE}$ of the bluest template, SB2, is approximately one order of magnitude larger than the others. This difference arises from the photometric data matching process, during which all templates are normalized to a wavelength of 5500 Å. The higher $w\text{MSE}$ of SB2 is attributed to its larger template flux density at the blue end compared to the other templates.

After five iterative processes, the calibrated SED templates and the initial SED templates are shown in Figure 7. The changes in the profile of the SED templates are not obvious, and in order to make the changes in the SED profile more apparent, we use a linear scale for the y-axis below 2, while maintaining a logarithmic scale for values above 2. Each template shows a decrease in $w\text{MSE}$ to varying degrees after calibration, suggesting a more accurate alignment with the photometric data, when comparing with the original templates.

The SB3 template is illustrated as an example (Figure 8) to demonstrate the evolution of $w\text{MSE}$ and $d\text{MSE}$ throughout the calibration process. The initial stages show a rapid decline in $w\text{MSE}$, indicating that the first few perturbations are usually the most influential, which suggests it is crucial to achieve a precise alignment between the photometric data and the SED templates during this stage. After five iterations, the final $w\text{MSE}$ of the calibrated SB3 template experiences a substantial reduction of over 60%, dropping from 5.211 to 2.036. This demonstrates the efficacy of the calibration process in enhancing the alignment between the template and the observed photometric data.

In addition to the perturbation algorithm proposed in CC20, the SED fitting code EAZY (Brammer et al. 2008) can also iteratively calibrate the SED templates according to the distribution of photometric data, a process known as Absolute Calibration (Weaver et al. 2022). Additionally, the flux of each band can also be corrected with an offset. The calibration process is similar to the perturbation method: the photometric data are first matched with the SED templates, and then the median value of $F_n^{\text{obs}}/F_n^{\text{temp}}$ in each rest-frame wavelength bin is calculated, as well the offsets for each band. The SED templates are calibrated by multiplying the calculated ratio with the flux density of the SEDs (Ilbert et al. 2006), and the calibration process will be iterated until the flux offsets in each

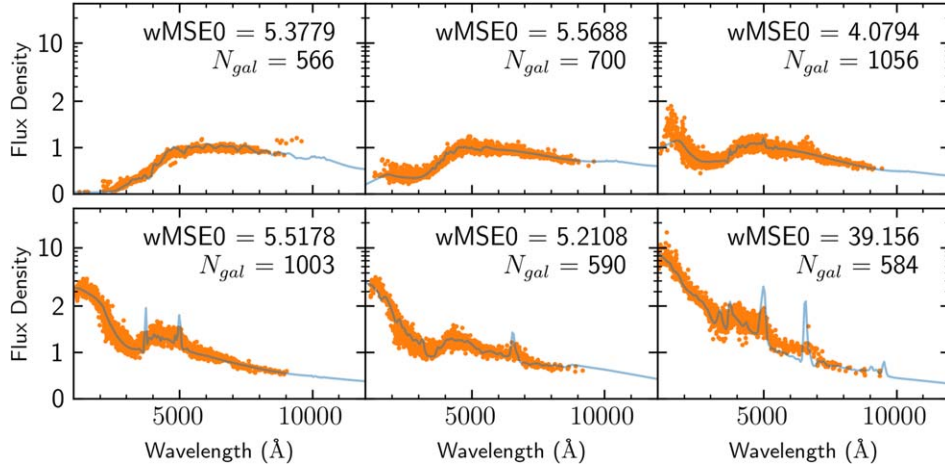


Figure 6. Initial SED templates with their sub-training sets for the first iteration. The blue curves are SED templates; orange dots are the photometric data. The flux density of all SED templates is normalized to 1 at 5500 Å. Note that we use a linear scale for the y -axis below 2, while maintaining a logarithmic scale for values above 2 in this figure.

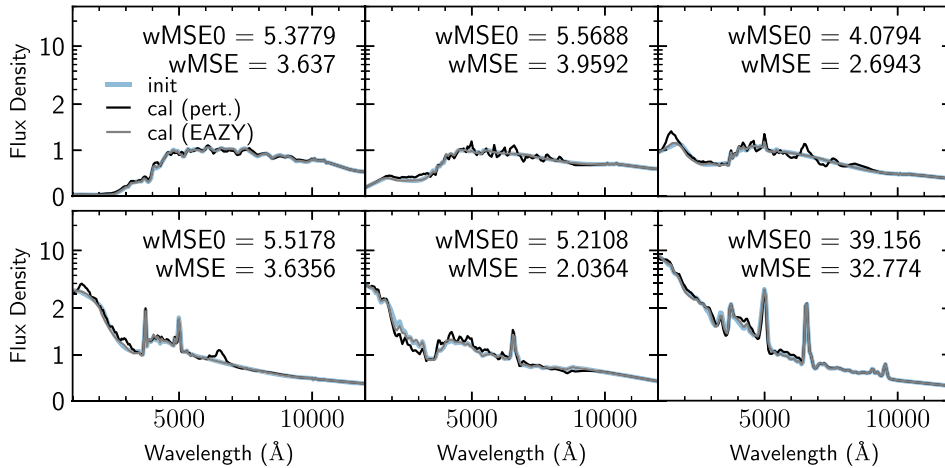


Figure 7. Calibrated templates and initial templates. The blue curves represent the initial templates, while black and gray curves are the templates calibrated with perturbation algorithm and EAZY respectively. Only initial templates are normalized at 5500 Å. To make the changes in the SED profile more apparent, we use a linear scale for the y -axis below 2, while maintaining a logarithmic scale for values above 2.

band change by less than 4%. EAZY can utilize either spec- z or photo- z for flux and SED calibration.

In this work, we used the same training set as in the perturbation algorithm, where all galaxies have available spec- z values and parameters, to calibrate the SED templates using EAZY. The process converged after four iterations, and the final flux offsets for the different bands are shown in Table 2. The calibrated SEDs are displayed in Figure 7 as gray curves.

While the calibrated SED templates from EAZY exhibit the same general trend as those calibrated with the perturbation algorithm, the amplitude of the changes in the SED profile is less apparent. This is because the calibration method in EAZY splits the wavelength into larger bins when using a small training set, resulting in smaller median values for the residuals and, consequently, smaller changes in the SED profiles. The changes

Table 2

Flux Offsets Used for Absolute Calibration given by EAZY, i Band is Taken as the Reference Band

Band	NUV	u	g	r
flux offset	0.9229	0.9730	0.9683	0.9695
band	i	z	y	
flux offset	1.0000	1.0376	1.0549	

in the SED profiles would be more apparent when using the full set of galaxies ($\sim 45,000$) as the training set, which further suggests that misalignment between the photometric data and the SED templates may occur on smaller wavelength scales. In contrast, the perturbation algorithm effectively corrects misalignment on a small wavelength scale, as it directly adjusts the SED

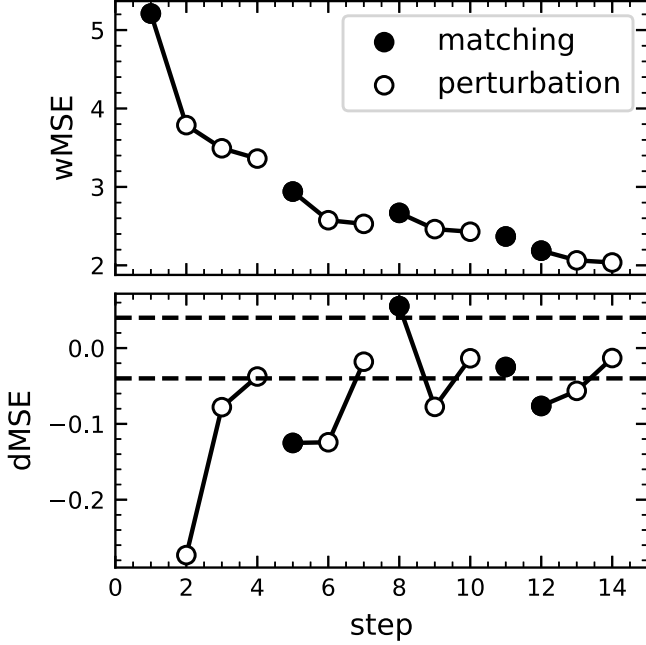


Figure 8. An example of the perturbation process for the SB3_B2004a template. The upper and lower panels show the w MSE and d MSE throughout the perturbation process, respectively. Open circles represent the w MSE and d MSE after each perturbation, and solid marks represent the w MSE and d MSE values after the creation of new sub-training sets. Changes in MSE and d MSE during the same iteration are connected by solid lines. The dashed lines in the lower panel signify the $\pm 4\%$ threshold for d MSE.

templates along the wavelength grid without binning. This precise adjustment is particularly efficient when the discrepancies between F^{temp} and F^{obs} are not large.

5. Results

We used the photo- z results of galaxies in the test set, which were derived from the calibrated templates, to quantify improvements in the SED templates after calibration. This section details the photo- z results estimated with the SED fitting code EAZY, and discusses the impact of different setups during the calibration process.

5.1. Photo- z Estimation

The SED fitting code EAZY determines redshifts by maximizing the redshift probability density function (PDF) which has the form of

$$p(z) = e^{-0.5(\chi_z^2 - \chi_{\text{min}}^2)}, \quad (10)$$

where the χ_z^2 is derived from

$$\chi_{z,i}^2 = \sum_{n=1}^{N_{\text{fit}}} \frac{(F_{z,i,n}^{\text{temp}} - F_n^{\text{obs}})^2}{(\delta F_n^{\text{obs}})^2}. \quad (11)$$

Photo- z s may become catastrophic outliers due to degeneracies in the color-redshift space, an incomplete set of

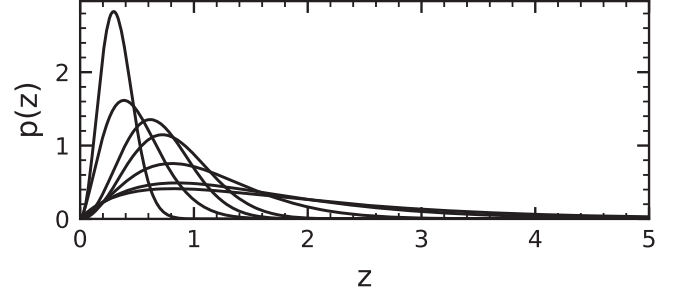


Figure 9. Prior probabilities for different i band magnitudes, ranging from 20.5 (narrowest) to 26.5 (widest).

templates, or poor photometric data (Newman & Gruen 2022). Despite a slight decrease in accuracy at high redshifts, a straightforward and effective method to overcome these degeneracies is to apply a Bayesian prior based on the redshift distribution. Benítez (2000) pioneered this approach in BPZ, and EAZY implements a similar prior to constrain the redshift PDF, as described in (Brammer et al. 2008, Equation (3))

$$p(z|m_{0,i}) \propto z^{\gamma_i} \exp[-(z/z_{0,i})^{\gamma_i}], \quad (12)$$

where γ_i and $z_{0,i}$ represent the parameters, and the PDF $p(z)$ is dependent on magnitude $m_{0,i}$ of the reference band. The PDF of galaxies can be calibrated by multiplying with prior probability density directly. Figure 9 depicts the part of the Bayesian prior used in this work, which is based on the i band photometric data and the distribution of z_{in} .

For each galaxy, we use z_{peak} as its best photo- z , which is estimated by the mean value of the largest peak in PDF, where $z_{\text{peak}} = (\int zp(z)dz)/(\int p(z)dz)$.

We employ the normalized median absolute deviation σ_{nMAD} , redshift bias b , and the fraction of outliers η to quantify the accuracy and precision of photo- z estimations, defined as follows:

$$\sigma_{\text{nMAD}} = 1.48 \times \text{median}\left(\left|\frac{\Delta z - \text{median}(\Delta z)}{1 + z_{\text{in}}}\right|\right), \quad (13)$$

$$b = \langle dz \rangle, \quad (14)$$

$$dz = \frac{z_{\text{peak}} - z_{\text{in}}}{1 + z_{\text{in}}}. \quad (15)$$

Outliers are defined as galaxies with $|dz| > 0.15$. In this work, all photo- z s are derived using single template fitting mode, with a Bayesian prior based on the i band. The amplitude of the template error function is set to 0, and the other parameters remain their default values in EAZY.

5.2. Photo- z Results

The photo- z results for the test set are displayed in Figure 10, using the template calibrated with the perturbation algorithm. The corresponding values for η and σ_{nMAD} are 2.55% and

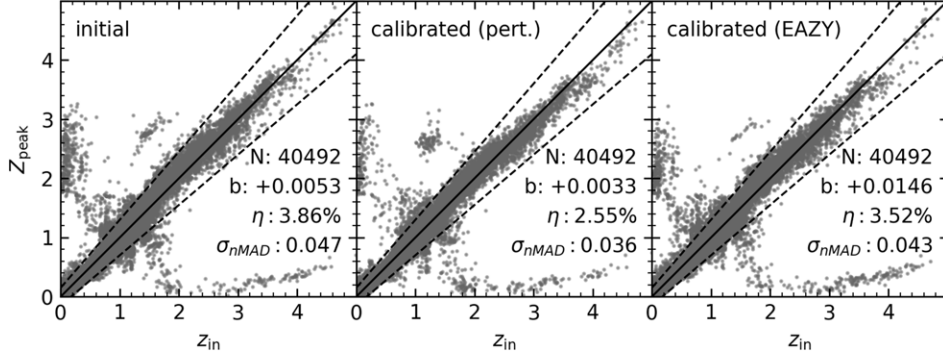


Figure 10. Photo- z results of 40492 galaxies in the test set. Each panel shows the z_{peak} vs. z_{in} from initial CWW+KIN templates (left), CWW+KIN templates calibrated with perturbation algorithm (middle) and CWW+KIN templates calibrated with EAZY (right). The dashed lines and solid line indicate $dz = 0.15$ and $z_{\text{in}} = z_{\text{peak}}$ respectively.

0.036, respectively. Compared to the results obtained using the initial CWW + KIN template (3.86% and 0.047), we observe reductions of approximately 30% in η and 20% in σ_{nMAD} . The general bias also decreases from +0.0053 to +0.0033. Templates calibrated with EAZY, which exhibit negligible changes in SED, yield photo- z results similar to the initial ones, with values η and σ_{nMAD} of 3.52% and 0.043. In order to focus on the enhancements provided by the perturbation algorithm, and to simplify subsequent figures, we exclude the results from templates calibrated with the EAZY method.

The z_{peak} estimates for $z_{\text{in}} > 3$ are slightly lower than the corresponding z_{in} values for both calibrated templates. This discrepancy is caused by the insufficient high redshift galaxies in our training set, which suggests when in a more realistic situation where the number of high redshift galaxies with spec- z is limited, the photo- z may be biased at higher redshift after template calibration.

In Figure 11, we present dz , the fraction of outliers η , and σ_{nMAD} across different z_{peak} bins. Compared to the initial CWW + KIN template, the calibrated CWW + KIN template results in reduced or comparable η and σ_{nMAD} for $z_{\text{peak}} < 3$. The most significant improvements in η and σ_{nMAD} occur in the range $1 < z_{\text{peak}} < 2.2$, where the Balmer break has shifted out of the middle part of the CSST filters, and the Lyman break has not yet reached the middle part of the CSST filters. In this region, the continuum plays a crucial role in photo- z estimation. The substantial improvements in the photo- z results for $1 < z_{\text{peak}} < 2.2$ are consistent with the results shown in Figure 7, where the perturbation algorithm primarily modifies the continuum portion of the SEDs.

The application of the prior eliminates a large number of outliers; however, there are still quite a few outliers in both photo- z s derived by calibrated and initial SED templates. Most of these outliers exhibit very low SNR in the NUV and/or u bands. The lack of constraint of photometric data with high

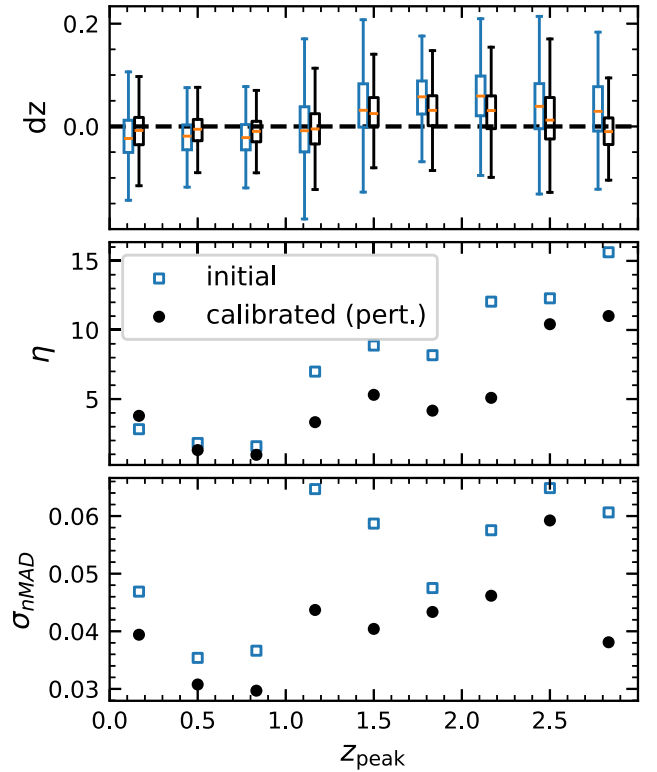


Figure 11. The accuracy parameters of different photo- z bins. From top to bottom are dz , η and σ_{nMAD} respectively. Results of calibrated and initial template are shown in solid dots and open squares respectively. dz is shown in a boxplot, where the orange bars indicate the median value, boxes indicate the quartile and error bars are derived with 1.5 times the inter-quartile range.

SNR makes these galaxies have multiple peaks in redshift PDFs, which make their photo- z s less reliable.

A redshift quality exclusion criterion, based on EAZY's quality parameter Q_z (Brammer et al. 2008, Equation (8)), is applied to remove possible outliers. Q_z describes the reliability

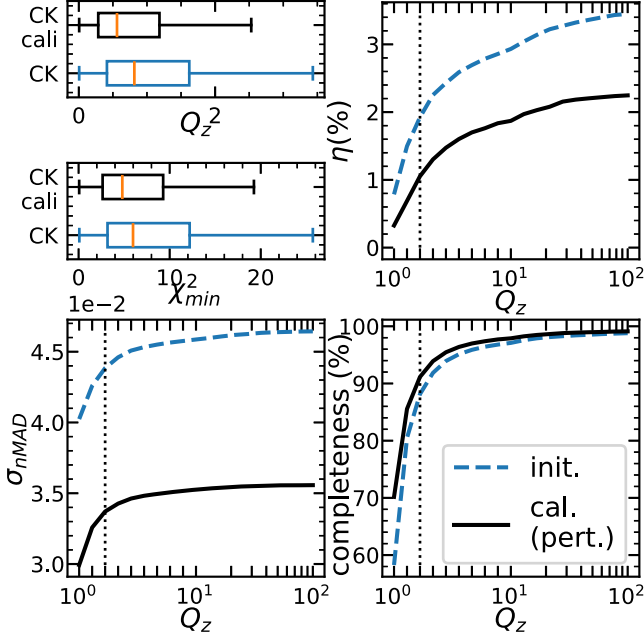


Figure 12. The distribution of Q_z and χ_{\min}^2 for two template sets and the photo- z results with different Q_z threshold. The upper left panels are the boxplots of the distribution of Q_z and χ_{\min}^2 . The photo- z results before and after calibration are shown as blue dashed and black solid curves respectively. The orange bars are the median value of the distribution. The dashed line indicates the $Q_z = 3$ cut we applied.

of photo- z estimations; it is related to redshift PDF, χ^2 and number of filters N_{filt} used in fitting, which has formula of

$$Q_z = \frac{\chi_{\min}^2}{N_{\text{filt}} - 3} \frac{z_{\text{up}}^{99} - z_{\text{lo}}^{99}}{p_{\Delta z=0.2}}, \quad (16)$$

where $z_{\text{up}}^{99} - z_{\text{lo}}^{99}$ represent the 3σ confidence intervals computed from the redshift PDF. χ_{\min}^2 is the minimum chi-square value between the observed photometric data and the SED template, and $p_{\Delta z=0.2}$ is the probability of redshift in $z_{\text{peak}} \pm 0.2$.

Though sacrificing completeness, selecting based on Q_z can further improve the photo- z accuracy. Figure 12 shows that the Q_z and χ^2 distribution of using the CWW + KIN template before calibration is much smaller than the values after calibration in the photo- z estimation. This indicates that calibrated templates achieve a better match between photometric data and SEDs. This is consistent with previous results for the smaller $w\text{MSE}$ displayed in Figure 8. A smaller Q_z distribution is crucial for applying the redshift quality cut. Figure 12 also displays the photo- z results after applying various Q_z thresholds. The calibrated templates consistently show improved performance across different Q_z thresholds. We find that $Q_z = 3$ is an effective threshold for maintaining lower η and σ_{nMAD} while retaining a sufficient number of galaxies.

However, the number of galaxies significantly decreases when $Q_z \leq 3$, as depicted in the lower right panel of Figure 12.

Figures 13 and 14 show the photo- z results after excluding galaxies with $Q_z > 3$. The fraction of galaxies excluded is reported as f_{cut} . The fraction of outliers η decreased significantly after quality cuts, and for CWW + KIN templates, the η and σ_{nMAD} are 1.94% and 0.044 respectively, with a f_{cut} of 11.8%. Compared to the initial templates, the result from calibrated CWW + KIN yields even lower η and σ_{nMAD} , which are 1.06% and 0.034 respectively. Additionally, the calibrated template set retains slightly higher completeness after the Q_z cut, with only 8.71% of galaxies excluded. When compared to Figure 11, Figure 14 shows a similar trend in the photo- z results across different z_{peak} bins.

5.3. Impact of Various Setups

To evaluate the impact of training set size on the results, we used training sets of three different sizes: 1% (~ 450 galaxies), 10% ($\sim 4 \times 10^3$ galaxies), and 50% ($\sim 2 \times 10^4$ galaxies) from the catalog. For each training set size, samples were randomly selected from the catalog, and the templates were calibrated accordingly. This process was repeated 100 times to reduce potential biases introduced by sample selection. The results shown in Figure 15 indicate that a training set with ~ 450 galaxies is sufficient to improve the photo- z accuracy, especially in terms of σ_{nMAD} . However, using such a small training set results in unstable improvements; for instance, 4 out of 100 tests exhibited a higher η compared to the initial templates. Increasing the training set size can enhance the stability of the photo- z results. For the CWW+KIN template, we find that training sets with about 4500 galaxies consistently yield improved photo- z results.

Additionally, we tested the impact of using different normalization methods during the photometric data matching process. All templates are calibrated with the same parameters. These results are also shown in Figure 15. For SED templates calibrated with small size training sets (450 galaxies), normalizing with Equation (9) yields better photo- z results from calibrated templates. The photo- z results become similar for both methods when the size of the training set increased to ~ 4500 galaxies.

Figure 16 shows an example of the $w\text{MSE}$ and $d\text{MSE}$ history during the calibration process using different normalization methods. The variations in the initial $w\text{MSE}$ primarily result from the different definitions of two normalization coefficients. When normalizing with Equation (9), $d\text{MSE}$ caused by the matching process is much smaller compared to normalizing with median values or ratios. This indicates a more accurate alignment between the SED templates and photometric data within sub-training sets, which proves particularly beneficial when the training set size is small.

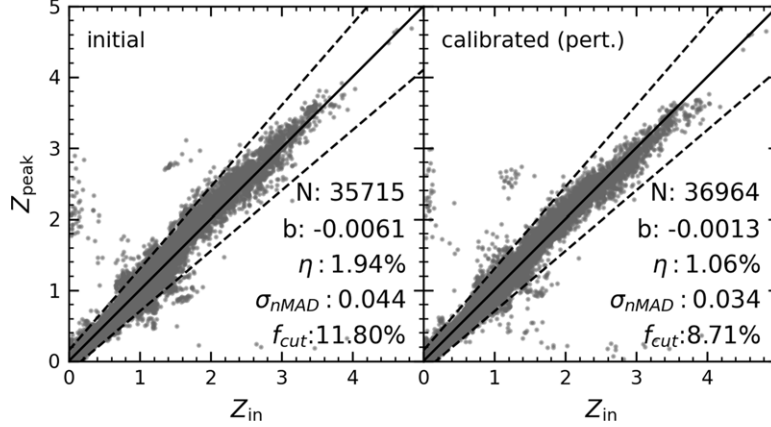


Figure 13. Results of photo- z estimations after applying Q_z cut.

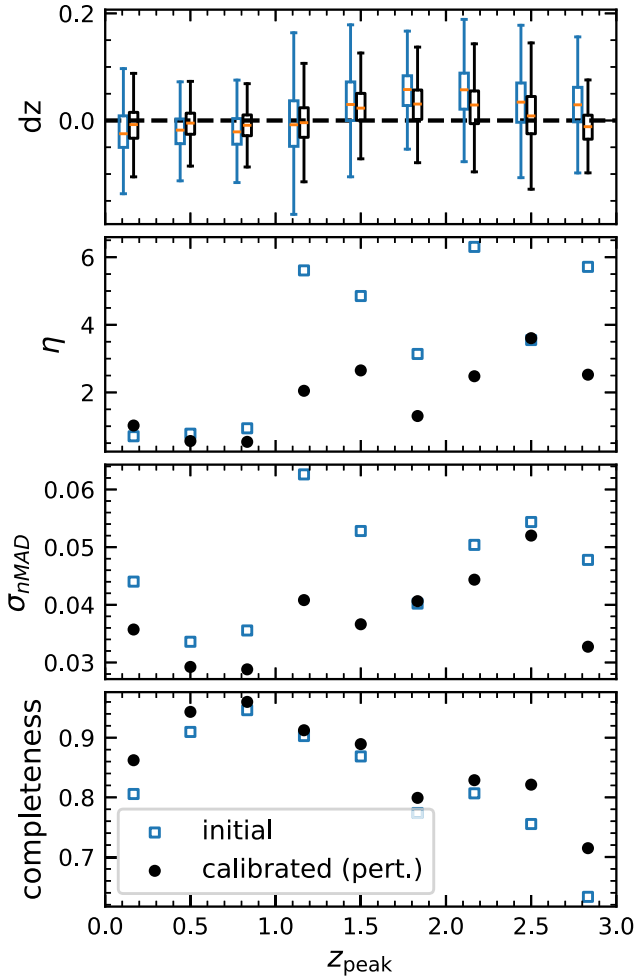


Figure 14. Photo- z results at different redshift bins after quality cuts. Panels from top to bottom are bias, fraction of outliers, σ_{nMAD} and completeness at different redshift bins, respectively.

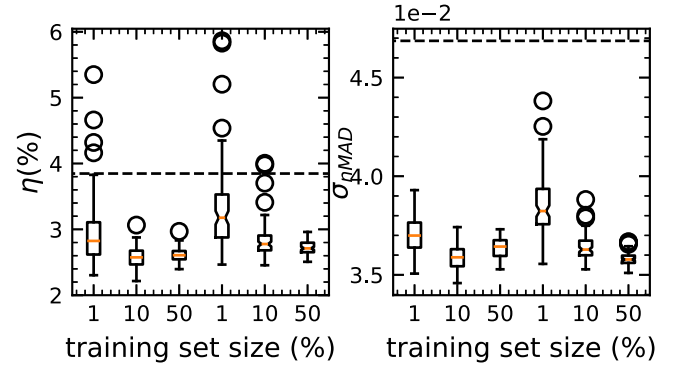


Figure 15. Boxplot represents photo- z results across varying training set dimensions and different normalization methods. Dashed lines signify the photo- z results derived from the initial CWW+KIN template. The orange lines represent the median values of the photo- z results obtained from the calibrated SED templates, while the boxes delineate the inter-quartile range. Error bars extend to 1.5 times the inter-quartile range, and circles identify outliers beyond these error bars. The results of normalizing with Equation (9) and median of ratios are marked with box and box with notch, respectively.

As shown in Figure 10, the photo- z results could be biased at high z_{in} due to the lack of high redshift samples in our training sets. To address this, we conducted an additional test by excluding galaxies with $z_{in} > 1$ from the training sets. Figure 17 shows the dz values from 100 different test sets for different training set sizes, divided into different z_{in} bins. For galaxies with $z_{in} > 4$, the photo- z results are almost unaffected, which is expected since there are very few galaxies with redshifts greater than 4 in our training sets. However, for galaxies in the range $1 < z_{in} < 4$, the templates calibrated without high redshift galaxies show slightly lower dz values, indicating that the estimated photo- z s become smaller after

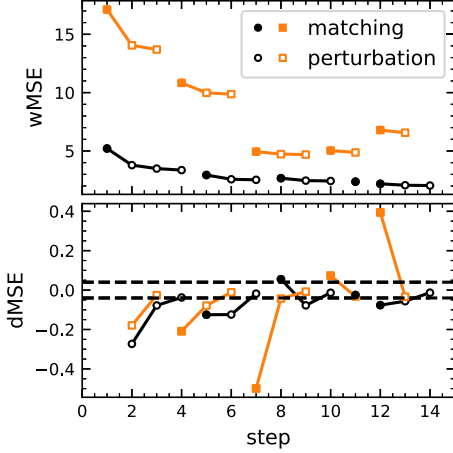


Figure 16. The variation of $wMSE$ and $dMSE$ during the perturbation process of SB3_B2004a template. The upper and lower panel show the $wMSE$ and $dMSE$ through the perturbation process, respectively. Hollow marks represent the $wMSE$ and $dMSE$ after each perturbation, solid marks represent the $wMSE$ and $dMSE$ each time new sub-training sets are created, and MSE and $dMSE$ changes in each iteration are connected with solid lines. Dashed lines in the right panel indicate the 4% threshold of $dMSE$. Normalizing cases with Equation (9) and median of ratio methods are marked with black and orange colors, respectively.

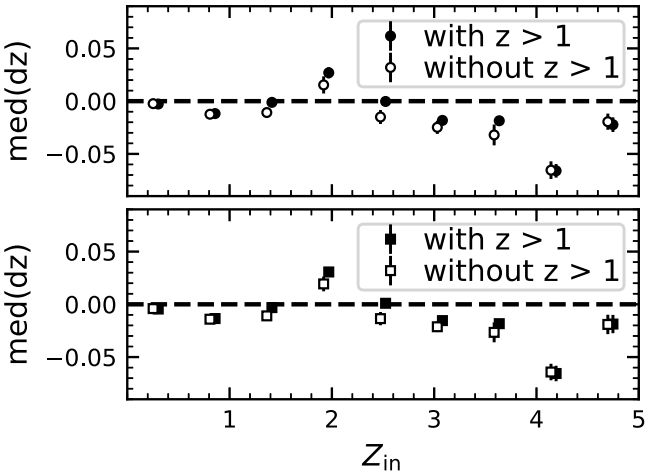


Figure 17. dz values from 100 different test sets, divided into different z_{in} bins. Solid and hollow circles (squares) represent the median values of the binned median dz values across the 100 results from SED templates calibrated with high redshift galaxies and without high redshift galaxies, respectively. Error bars signify the standard deviations of dz . The upper and lower panels show the results from training sets consisting of 10% and 50% of the SNR-cut catalog, respectively.

excluding galaxies with $z_{in} > 1$ from the training sets, potentially leading to underestimated photo- z at higher z_{in} .

6. Summary

In this work, we used the perturbation algorithm described in CC20 to calibrate CWW+KIN templates for the CSST mock

catalog. The mock catalog is derived from simulated CSST images which are based on HST-ACS F814W images. We selected galaxies with $SNR \geq 10$ in g or i band as training set and test set for SED calibration respectively.

Before the calibration, we smoothed the two SED templates of starburst galaxies with strong emission lines. The smoothed templates successfully avoided overcorrection and reduced the discrepancies in F^{temp} caused by sampling. We tested both the absolute calibration method used in EAZY and the perturbation algorithm for template calibration. Compared to absolute calibration, the perturbation algorithm can modify the SED templates directly without binning the wavelength, making it more sensitive to misalignments at small wavelength scale when the training set is small. For the CSST mock catalog, the SED templates calibrated with the perturbation algorithm yielded photo- z results with higher accuracy.

We used SED fitting code EAZY for photo- z estimation. The results derived from SED templates calibrated with the perturbation algorithm demonstrated moderate improvements: a $\sim 30\%$ reduction in the fraction of outliers (η), reduced to 2.55%, and a $\sim 20\%$ reduction in the normalized median absolute deviation (σ_{nMAD}), which decreased to 0.036. In contrast, templates calibrated using absolute calibration yielded photo- z results with an η of 3.52% and a σ_{nMAD} of 0.043, similar to the initial CWW+KIN templates, which had η of 3.86% and σ_{nMAD} of 0.047. The calibrated SED templates provided a tighter fitting during the photo- z estimation process, resulting in an improved redshift quality parameter (Q_z). Consequently, there was less loss in completeness after applying quality cuts of $Q_z \leq 3$, and the photo- z results showed a $\sim 60\%$ decrease in η and a similar σ_{nMAD} , which were 1.06% and 0.034, respectively. However, this quality cut still resulted in an exclusion of 8.71% of the galaxies.

Additionally, we tested the impact of varying training set sizes and different normalization coefficients on the photometric data matching process. We found that a training set of approximately 4500 galaxies is sufficient to consistently improve the photo- z results derived from calibrated templates, regardless of the normalization coefficient used. For smaller training sets (450 galaxies), employing the normalization coefficient defined by Equation (9) yielded slightly better performance.

The limited representation of high redshift galaxies in our training set introduces a minor systematic underestimation of their photo- z s compared to the input redshift values, even in scenarios where the redshift distributions of both the training and test sets are identical, thereby establishing an ideal condition. In realistic observations, it is likely that fewer high redshift galaxies will have spec- z available, potentially introducing additional bias into the photo- z results. The exclusion of galaxies with $z_{in} > 1$ exacerbates the photo- z underestimation for high redshift galaxies, relative to SED templates calibrated with $z_{in} > 1$ training data. Future extensions of this study will focus on incorporating hierarchical Bayesian priors in order to mitigate

high redshift biases while preserving the method's computational efficiency and small-sample adaptability.


The template calibration approach, which utilizes a perturbation algorithm, exhibits the capacity to enhance accuracy with the deployment of a comparatively small training set. Moreover, it provides a feasible route for conducting initial cosmological analyses in stage-IV surveys such as CSST, where template calibration is capable of producing immediately applicable photo- z catalogs prior to the implementation of comprehensive spectroscopic campaigns.

Acknowledgments

Z.J.L. acknowledges support from the Shanghai Science and Technology Foundation Fund under grant No. 20070502400. L.P. F. acknowledges the support from the Innovation Program of Shanghai Municipal Education Commission (grant No. 2019-01-07-00-02-E00032). Y.G. acknowledges the support from National Key R&D Program of China grant Nos. 2022YFF0503404, 2020SKA0110402, the CAS Project for Young Scientists in Basic Research (No. YSBR-092). S.H.Z. acknowledges support from the Program for Professor of Special Appointment (Eastern Scholar) at Shanghai Institutions of Higher Learning and the Shuguang Program of Shanghai Education Development Foundation and Shanghai Municipal Education Commission. This work is also supported by the National Natural Science Foundation of China (NSFC, Grant Nos. U1931210, 12141302, 12173026, and 11933002), and China Manned Space Project with grant Nos. CMS-CSST-2021-A01, CMS-CSST-2025-A02, CMS-CSST-2025-A03, CMS-CSST-2025-A05 and CMS-CSST-2025-A20.

ORCID iDs

Junhao Lu  <https://orcid.org/0009-0004-9425-5027>

Shaohua Zhang  <https://orcid.org/0000-0001-8485-2814>

References

Abell, S. K., Rogers, M. A. P., Hanuscin, D. L., Lee, M. H., & Gagnon, M. J. 2009, *JSTED*, 20, 77

- Akeson, R., Armus, L., Bachelet, E., et al. 2019, arXiv:1902.05569
- Arnouts, S., Cristiani, S., Moscardini, L., et al. 1999, *MNRAS*, 310, 540
- Arnouts, S., & Ilbert, O. 2011, LePHARE, ascl:1108.009
- Assef, R. J., Kochanek, C. S., Brodwin, M., et al. 2010, *ApJ*, 713, 970
- Benítez, N. 2000, *ApJ*, 536, 571
- Benítez, N., Ford, H., Bouwens, R., et al. 2004, *ApJS*, 150, 1
- Bolzonella, M., Miralles, J. M., & Pelló, R. 2000, *A&A*, 363, 476
- Brammer, G. B., van Dokkum, P. G., & Coppi, P. 2008, *ApJ*, 686, 1503
- Bruzual, A. G., & Charlot, S. 1993, *ApJ*, 405, 538
- Bruzual, G., & Charlot, S. 2003, *MNRAS*, 344, 1000
- Budavári, T., Szalay, A. S., Connolly, A. J., Csabai, I., & Dickinson, M. 2000, *AJ*, 120, 1588
- Cao, Y., Gong, Y., Meng, X.-M., et al. 2018, *MNRAS*, 480, 2178
- Carliles, S., Budavári, T., Heinis, S., Priebe, C., & Szalay, A. S. 2010, *ApJ*, 712, 511
- Cavuoti, S., Amaro, V., Brescia, M., et al. 2017, *MNRAS*, 465, 1959
- Coleman, G. D., Wu, C. C., & Weedman, D. W. 1980, *ApJS*, 43, 393
- Collister, A. A., & Lahav, O. 2004, *PASP*, 116, 345
- Crenshaw, J. F., & Connolly, A. J. 2020, *AJ*, 160, 191
- Dalmaso, N., Pospisil, T., Lee, A. B., et al. 2020, *A&C*, 30, 100362
- D'Isanto, A., & Polsterer, K. L. 2018, *A&A*, 609, A111
- Euclid Collaboration, Mellier, Y., Abdurro'uf, et al. 2025, *A&A*, 697, A1
- Feldmann, R., Carollo, C. M., Porciani, C., et al. 2006, *MNRAS*, 372, 565
- Fioc, M., & Rocca-Volmerange, B. 1997, *A&A*, 326, 950
- Firth, A. E., Lahav, O., & Somerville, R. S. 2003, *MNRAS*, 339, 1195
- Gomes, Z., Jarvis, M. J., Almosallam, I. A., & Roberts, S. J. 2018, *MNRAS*, 475, 331
- Green, J., Schechter, P., Baltay, C., et al. 2012, arXiv:1208.4012
- Ilbert, O., Arnouts, S., McCracken, H., et al. 2006, *A&A*, 457, 841
- Ivezić, Ž., Axelrod, T., Brandt, W. N., et al. 2008, *SerAJ*, 176, 1
- Jones, E., & Singal, J. 2017, *A&A*, 600, A113
- Kinney, A. L., Calzetti, D., Bohlin, R. C., et al. 1996, *ApJ*, 467, 38
- Laigle, C., McCracken, H. J., Ilbert, O., et al. 2016, *ApJS*, 224, 24
- Laureijs, R., Amiaux, J., Arduini, S., et al. 2011, arXiv:1110.3193
- Lu, J., Luo, Z., Chen, Z., et al. 2023, *MNRAS*, 527, 12140
- Luo, Z., Li, Y., Lu, J., et al. 2024a, *MNRAS*, 535, 1844
- Luo, Z., Tang, Z., Chen, Z., et al. 2024b, *MNRAS*, 531, 3539
- Maraston, C. 2005, *MNRAS*, 362, 799
- Newman, J. A., Abate, A., Abdalla, F. B., et al. 2015, *APh*, 63, 81
- Newman, J. A., & Gruen, D. 2022, *ARA&A*, 60, 363
- Razim, O., Cavuoti, S., Brescia, M., et al. 2021, *MNRAS*, 507, 5034
- Salvato, M., Ilbert, O., & Hoyle, B. 2019, *NatAs*, 3, 212
- Abbott, T. M. C., Abdalla, F. B., Alarcon, A., et al. 2018, *PhRvD*, 98, 043526
- Wadadekar, Y. 2005, *PASP*, 117, 79
- Weaver, J. R., Kauffmann, O. B., Ilbert, O., et al. 2022, *ApJS*, 258, 11
- Zhan, H. 2011, *SSPMA*, 41, 1441
- Zhan, H. 2018, 42nd COSPAR Scientific Assembly, 42, E1.16
- Zhou, X., Gong, Y., Meng, X.-M., et al. 2021, *ApJ*, 909, 53
- Zhou, X., Gong, Y., Meng, X.-M., et al. 2022, *RAA*, 22, 115017