



Going Forward with the Nancy Grace Roman Space Telescope Transient Survey: Validation of Precision Forward-modeling Photometry for Undersampled Imaging

David Rubin^{1,2} , Aleksandar Cikota^{2,3} , Greg Aldering², Andy Fruchter⁴, Saul Perlmutter^{2,5}, and Masao Sako⁶

¹Department of Physics and Astronomy, University of Hawai‘i at Mānoa, Honolulu, HI 96822, USA; drubin@hawaii.edu

²E.O. Lawrence Berkeley National Laboratory, 1 Cyclotron Rd., Berkeley, CA 94720, USA

³European Southern Observatory, Alonso de Cordova 3107, Vitacura, Casilla 19001, Santiago de Chile, Chile

⁴Space Telescope Science Institute, 3700 San Martin Drive Baltimore, MD 21218, USA

⁵Department of Physics, University of California Berkeley, Berkeley, CA 94720, USA

⁶Department of Physics and Astronomy, University of Pennsylvania, Philadelphia, PA 19104, USA

Received 2021 February 10; accepted 2021 March 31; published 2021 June 28

Abstract

The Nancy Grace Roman Space Telescope (Roman) is an observatory for both wide-field observations and coronagraphy that is scheduled for launch in the mid-2020s. Part of the planned survey is a deep, cadenced field or fields that enable cosmological measurements with type Ia supernovae (SNe Ia). With a pixel scale of $0''.11$, the Wide Field Instrument will be undersampled, presenting a difficulty for precisely subtracting the galaxy light underneath the SNe. We use simulated data to validate the ability of a forward-model code (such codes are frequently also called “scene-modeling” codes) to perform precision supernova photometry for the Roman SN survey. Our simulation includes over 760,000 image cutouts around SNe Ia or host galaxies ($\sim 10\%$ of a full-scale survey). To have a realistic 2D distribution of underlying galaxy light, we use the VELA simulated high-resolution images of galaxies. We run each set of cutouts through our forward-modeling code which automatically measures time-dependent SN fluxes. Given our assumed inputs of a perfect model of the instrument point-spread functions and calibration, we find biases at the millimagnitude level from this method in four red filters (*Y106*, *J129*, *H158*, and *F184*), easily meeting the 0.5% Roman inter-filter calibration requirement for a cutting-edge measurement of cosmological parameters using SNe Ia. Simulated data in the bluer *Z087* filter shows larger $\sim 2\text{--}3$ mmg biases, also meeting this requirement, but with more room for improvement. Our forward-model code has been released on Zenodo.

Unified Astronomy Thesaurus concepts: [Surveys \(1671\)](#); [Infrared telescopes \(794\)](#); [Space telescopes \(1547\)](#); [Dark energy \(351\)](#); [Type Ia supernovae \(1728\)](#)

1. Introduction

The Nancy Grace Roman Space Telescope (Roman) is an observatory for both wide-field observations and coronagraphy that is scheduled for launch in the mid-2020s. The Wide Field Instrument (WFI) covers 0.281 square degrees and performs both imaging and low-resolution slitless spectroscopy. One of the primary science objectives of the Roman mission is to investigate the expansion history of the universe using thousands of Type Ia Supernovae (SNe Ia). Although the Roman supernova survey strategy is not yet finalized, the survey is planned to have two components: a ~ 5 days cadence multi-band imaging survey to discover transients and measure their light curves, and a spectroscopic component to classify a subset of the transients and measure redshifts. The cadenced observations would take place over a period of about two years (146 visits), with each visit rotating $\sim 5^\circ$ from the previous visit (two full rotations over two years) to keep the solar panels pointed at the Sun.

To balance field of view, read noise, and point-spread function (PSF) sampling, the pixel scale of the WFI was set at $0''.11$, leaving the imaging PSF undersampled (Table 1). This undersampling is not mitigated by the SN survey strategy, since much of the survey will likely only have one (undithered) exposure per filter per epoch in order to minimize overheads and read noise. Undersampled, undithered imaging poses a challenge for photometry methods based on image resampling (e.g., Alard & Lupton 1998). Much existing undersampled photometry is thus done with codes that use forward modeling (e.g., Suzuki et al. 2012; Hayden et al. 2021). Forward modeling (called “scene modeling” by Holtzman et al. 2008) bypasses image resampling to model each image as observed.⁷ Figure 1 shows an example and Figure 2 shows the recovered high-resolution galaxy model.

⁷ The ground-based SuperNova Legacy Survey developed a similar sort of code that used resampled and aligned images (Astier et al. 2006; Guy et al. 2010). The goal in that work was to avoid image subtraction and take the time-variable PSF into account.

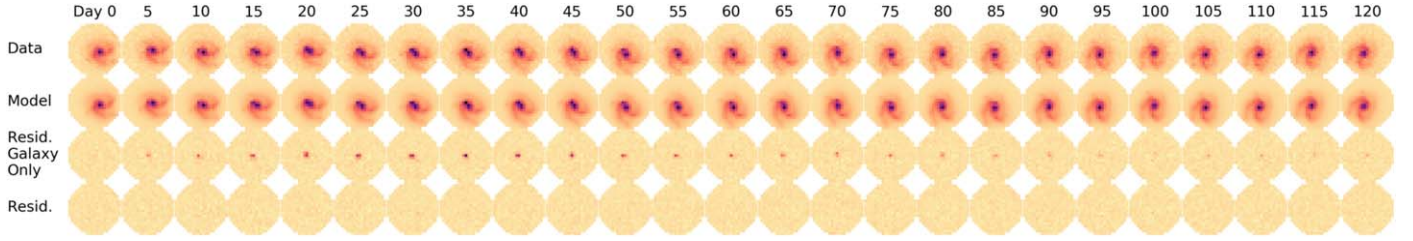


Figure 1. Forward modeling an example simulated SN in Y106. Each epoch (each column) consists of only one undithered exposure per filter; subpixel sampling is provided by a $\sim 5^\circ$ rotation between subsequent epochs. The top row of panels shows a set of cutouts around the SN; only epochs with SN light are shown. The next row of panels shows the model inferred from this data and 49 reference epochs without the SN. The galaxy model is an analytic function, parameterized with a 2D set of spline nodes. For each image, the galaxy model is sampled at high resolution ($11\times$ oversampling, across each pixel, then convolved with the PSF and the pixel and sampled at the native pixel scale. This accuracy is sufficient for $\sim 10^{-4}$ accuracy in representing the spline. The image-dependent SN light (also convolved by the PSF and the pixel and sampled at the native scale) and background are added to this galaxy model. The next row of panels show the residuals where the SN models are not subtracted. Finally, the bottom row of panels show the residuals when the modeled SN and modeled galaxy are subtracted.

Table 1
Filter-dependent Simulation Quantities

Filter	Z087	Y106	J129	H158	F184
Background ($e^- \text{ pix}^{-1} \text{ s}^{-1}$)	0.349	0.384	0.376	0.365	0.381
Exposure Time (s)	300	300	300	300	600
Read Noise (e^-)	8.3	8.3	8.3	8.3	6.9
Fitted PSF FWHM ($''$)	0 $''$.127	0 $''$.130	0 $''$.136	0 $''$.150	0 $''$.166
Gaussian Fitted PSF FWHM (Pixels)	1.16	1.18	1.24	1.36	1.51

The gold standard for validating forward-model results is to inject simulated SNe into real survey data (Holtzman et al. 2008; Suzuki et al. 2012; Astier et al. 2013; Brout et al. 2019). As we have no Roman data to validate with, we are left with two choices for test data: inject simulated SNe into images from the Hubble Space Telescope Wide Field Camera 3 IR channel (HST WFC3 IR, which has similar filters with a similar pixel scale of $0''$.128), or use fully simulated data. We chose to test entirely with simulated data for two reasons. (1) There are a limited number of WFC3 IR visits that have many rotations in several filters (and many visits do not contain a fair sampling of the universe, e.g., they contain a galaxy cluster or a globular cluster). (2) Testing with HST data will confound HST calibration uncertainties with forward-model problems. Of particular worry are thermal variations due to the HST’s low orbit (Bély et al. 1993), PSF spatial variations (Anderson 2016), and detector effects (e.g., Zhou et al. 2017).

In Section 2 we very briefly outline the assumptions and requirements for our forward modeling algorithm and how it may interface with the Roman pipeline and cosmology analysis. In Section 3 we describe the WFI mock observations which we generate and use to test the forward modeling algorithm. In Section 4 we explain the forward-modeling assumptions in detail, and in Section 5 we present and discuss the results. Section 6 summarizes our conclusions.

2. Forward-model Inputs and Outputs

Figure 3 shows a conceptual overview of the Roman SN cosmological data-processing flowchart and how this work fits in. The downlinked data will be processed into calibrated images (top of Figure 3). Mosby et al. (2020) discuss the performance of the IR detectors in detail. In short, there are eighteen Teledyne HAWAII 4RG detectors (with $10 \mu\text{m}$ $4k$ by $4k$ pixels) that non-destructively read out every 2.825 s. These multiple readouts enable lower read noise than is possible with one read, enable rejecting cosmic-ray hits during a single exposure (visible as jumps in charge versus readouts), and possibly provide better control of pointing drifts and detector effects (Rauscher et al. 2019). (To lower the required bandwidth, averaged groups or other linear combinations of readouts will be downlinked.) We assume for this work that the calibration process produces 2D images with known astrometric and photometric calibration. We acknowledge the possibility that we may need to go back earlier in the process (for example by fitting the readouts directly); this will have to be explored with better simulations of the detectors (and ultimately explored with real data).

After the images are calibrated, we assume further processing generates a PSF model. This PSF model will have to take a focal-plane position and effective wavelength into account (and possibly temporal or thermal variations as discussed above for HST). Depending on the linearity calibration (Choi & Hirata 2020), it may also have to take flux into account.

The transients in each image will have to be found and assessed. This will require a highly automated process; a 20 deg^2 survey with a five-day cadence is equivalent to searching more than 3000 WFC3 IR pointings per day. Hayden et al. (2021) demonstrated an automated transient classifier for WFC3 IR data with near-human levels of performance (as noted above, WFC3 IR has a similar pixel scale and wavelength coverage), so the search process is feasible, even for these large data sets.

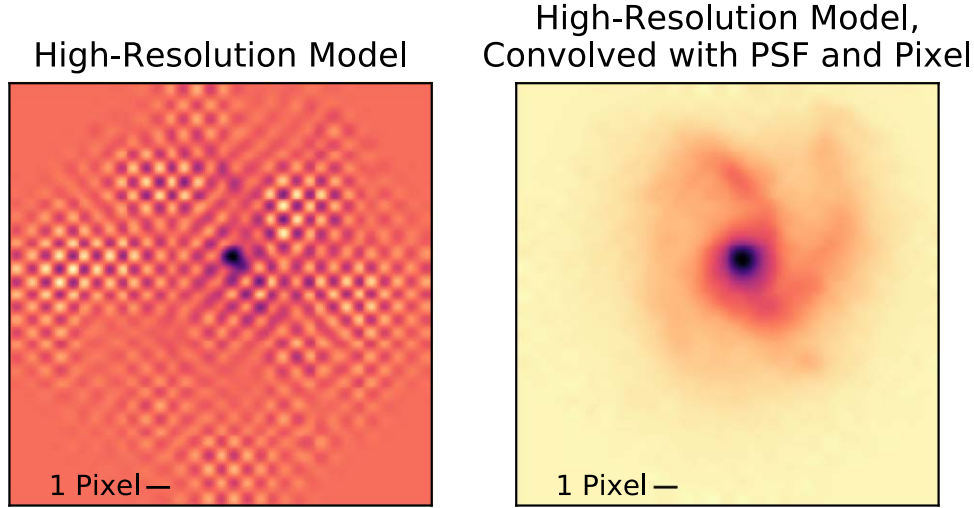


Figure 2. Recovered galaxy model from Figure 1. The left panel shows the recovered high-resolution galaxy model G (from Equation (1)), sampled at $11 \times$ the native resolution (i.e., $0''.01$). The right panel shows G convolved with the PSF and the pixel (the PSF has the same orientation), also at $11 \times$ the native resolution. Most of the unphysical high-frequency power visible in the left panel is suppressed. Note that the convolution with the pixel ensures that the total galaxy flux is not dependent on the alignment with the pixels in a given epoch.

With a series of calibrated images, a PSF model, and transient detections, the forward-modeling code can run. We describe this in more detail in Section 4. This step produces calibrated fluxes. The slitless spectroscopy will also need a separate forward-model code appropriate for 3D reconstruction (two-dimensions on the sky plus wavelength). This is a fundamentally harder problem (because of the increase in dimensionality, the increase in data volume, and the spatial/spectral degeneracy of a slitless spectrograph observing a complex scene). Ryan et al. (2018) demonstrate this concept on WFC3 IR data.

The lower half of Figure 3 outlines the steps involved in going from these calibrated imaging and spectroscopic fluxes to SN distances and cosmological results. We do not elaborate here, as many of the steps will be similar to other surveys (e.g., Scolnic et al. 2018).

3. Simulated Data Generation

To create mock observations, we use galaxies from the VELA Cosmological Simulation (Snyder 2018; Simons et al. 2019). The data set spans the cosmic time evolution of 35 galaxies over 10–50 timesteps with cosmological scale factors between 0.05 and 0.5 (redshift 1–19), each with approximately 20 viewing angles. The simulated spatially dependent galaxy spectral energy distributions (SEDs) are integrated over Roman filters (and other existing and proposed observatories), making simulated images. These simulated images are high resolution (oversampled by a factor ~ 15 compared to Roman pixels), and are much larger than a PSF (800 by 800 oversampled pixels or ~ 50 by 50 native pixels), making them perfect for precision tests of galaxy subtraction. The stellar mass distribution is also

similar to SN Ia hosts. We add the time-dependent SN Ia fluxes and reproject the high-resolution images onto the WFI detector pixels at different telescope orientations (depending on the position relative to the Sun) for 74 time epochs in a range of 1 yr (which is designed to fit well within the two-year planned Roman survey).

3.1. Supernova Light Curves

So that our light curves would be reasonably realistic, we generated a sample of SN Ia fluxes using the SALT2-Extended model in the SNCosmo Python package (Barbary 2014). SALT2 is a two-parameter (light-curve shape x_1 and light-curve color c) spectro-temporal model (Guy et al. 2007) that was extended into the UV and NIR with the Hsiao et al. (2007) template.⁸ To ensure good sampling of redshift, we assumed a random uniform redshift distribution of the SN sample between $z = 0.7$ and 2, instead of following SN rates and cosmological volume. Our assumed cadence is five days (Spergel et al. 2015; Hounsell et al. 2018). We generated a random time of peak brightness compared to the cadence (so there is not always an epoch right at maximum, nor is maximum always between two epochs). We used a random normal absolute magnitude distribution of $-19.1 - 0.14x_1 + 3.1c + \Delta m$ mag, where x_1 and c are drawn from random normal distributions centered around 0 with standard deviations of 1 and 0.1, respectively. These are similar to the distributions in, e.g., Scolnic & Kessler (2016), and are intended to span a representative range of signal-to-noise ratio (S/N) in the simulated imaging. Δm is

⁸ More than one version of SALT2-Extended has been trained. We use the SNCosmo version, not the published one (Pierel et al. 2018); the SNCosmo version seems to have more accurate rest-frame UV fluxes.

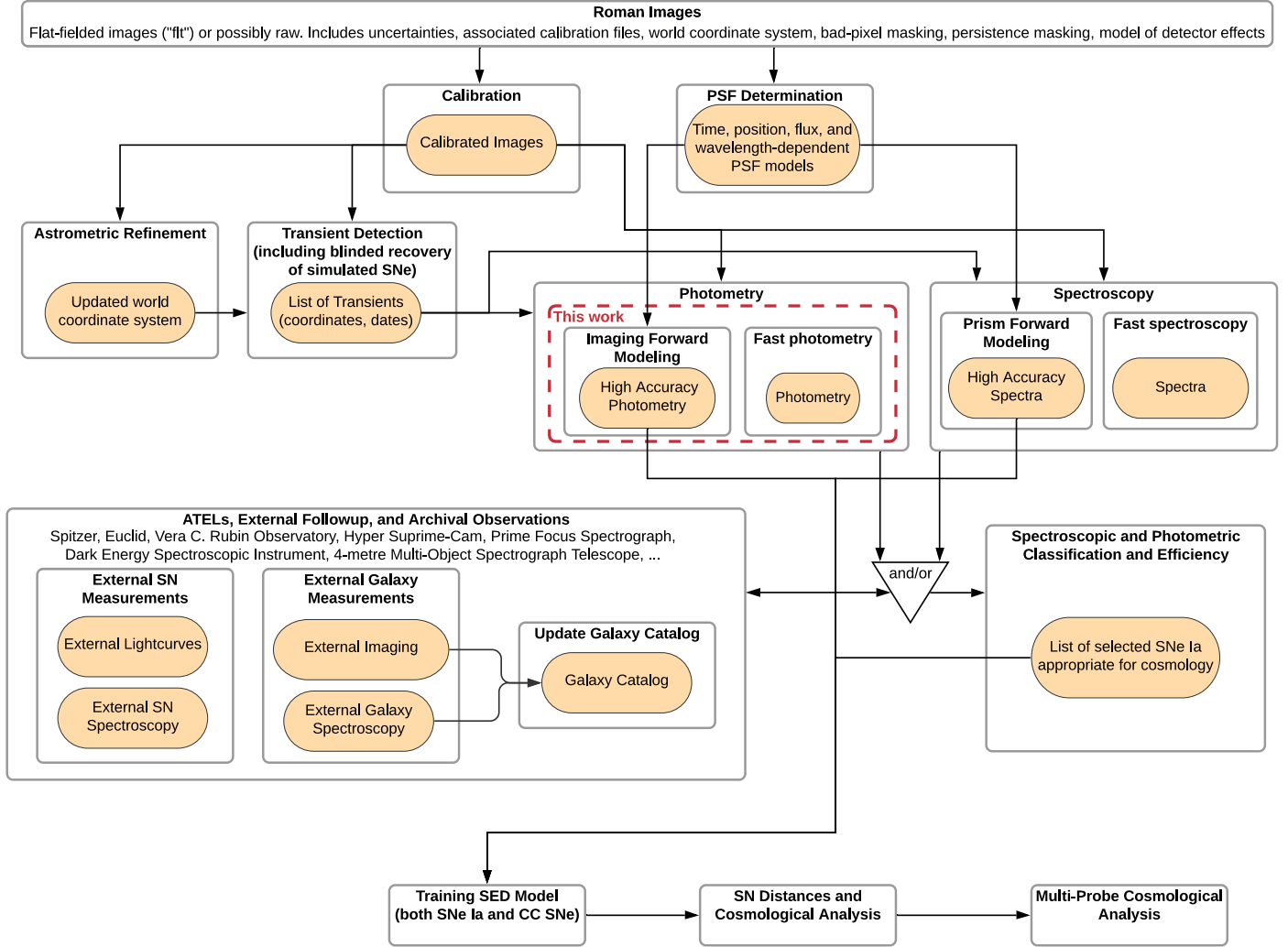


Figure 3. Conceptual Roman SN cosmological data-processing flowchart. The red dashed square denotes the photometry component, which is described and validated in this work. We show each step as a single box, but many surveys use more than one semi-independent analyses of the same data as a cross-check (e.g., Guy et al. 2010).

also assumed to be Gaussian, with a standard deviation of $\sqrt{0.1^2 + (0.055z)^2}$. The $0.055z$ is the magnitude dispersion due to the weak gravitational lensing of galaxy halos along the line of sight (Jönsson et al. 2010). Finally, we calculated the integrated flux of the SNe in the five Roman bands (Z087, Y106, J129, H158, F184) over the SALT2 phase range of (−15 to +45 rest-frame days) in 5 observer-frame day steps (an average of 31 epochs).

3.2. Supernova Positions

We assume that SNe Ia is distributed following the optical light of the galaxy (Anderson et al. 2015), and convolve the high-resolution Y106 VELA image by a Gaussian with a 1 kpc radius before choosing locations to plant SNe.

3.3. Point Spread Functions

We use PSFs generated by WebbPSF (Perrin et al. 2014). We convolve the PSFs with square $0''.11$ pixels with uniform sensitivity. As described in Section 2, we assume that detector effects such as count nonlinearity, count-rate nonlinearity, and inter-pixel capacitance have been perfectly calibrated and can be neglected. We also assume that the dependence of the PSFs on the SED of the source is negligible.⁹ All of these simplifying

⁹ In practice, one can use an iterative process of generating a PSF, measuring photometry, estimating the SED from a light-curve fit, and re-estimating the PSF (Suzuki et al. 2012). We neglect this iteration for simplicity. For sufficiently well sampled images, an even simpler approximation suffices: using a single PSF for all SEDs and modifying the filter bandpass instead of the PSF (e.g., Guy et al. 2010; Suzuki et al. 2012).

assumptions are in keeping with our philosophy for this work of focusing on the “Photometry” box in Figure 3.

3.4. Reprojection

With a full set of oversampled live-SN and reference images in hand, we use `Astropy reproject_exact` to rotate the oversampled data to match the rotation angle for each epoch and give each epoch and filter a random sub-pixel dither offset. This procedure technically convolves the images by each oversampled pixel twice: once when generating the data, and once when rotating the data. We use $30\times$ oversampled images (pixel doubling the $15\times$ oversampled VELA images), so this limits the accuracy of our simulations to $\mathcal{O}(30^{-2}) \approx 0.1\%$. We then convolve the rotated image by the PSF and the pixel, and sample it at the native $0''.11$ resolution. Figure 4 shows a randomly selected set of simulated images (near maximum light for the SNe) before the addition of noise.

WebbPSF defines the PSF as the PSF at exactly the sampled locations. The image-reprojecting code effectively defines the PSF as convolved with the subpixel (in the sense that adding together all the subpixels of a pixel of the PSF should exactly equal the PSF convolved with the pixel). We sample the PSFs at very high resolution ($\sim 100\times$) and integrate over each subpixel to ensure that the pixel convolution is done with high accuracy. Such definition considerations will need to be kept in mind as the Roman software stack is built.

3.5. Sources of Noise

It is likely that the Roman SN survey strategy will consist of tiers (Spergel et al. 2015; Hounsell et al. 2018; Rubin 2020), with each “wedding cake” layer of the survey trading area on the sky against depth. The deepest tier will likely have ~ 300 s exposure times and reach redshift ~ 2 with reasonable S/N; the wider tier(s) will have $\lesssim 100$ s exposure times and reach redshift ~ 1 or less. The VELA data are generated only as low as redshift 1, so we will have the most fidelity simulating the deepest tier. To extend the redshift span, we plant simulated SNe as low as redshift 0.7 on the $z = 1$ VELA images (the angular scale is only about 12% different between $z = 0.7$ and 1). As higher-redshift SNe have lower contrast against their host galaxy (due to the loss of physical resolution with increasing distance for $z \lesssim 1.6$), the deepest tier presents the most difficult host-galaxy-subtraction problem. Validating the forward-model code in the deep tier will thus validate it in any other tiers as well.

We assume that this deep tier consists of 300 s exposures in Z087, Y106, J129, and H158, and a 600 s exposure in F184. It is also possible that the R062 filter or the proposed K213 band could be used, but these were too recent to be in the Vela simulations, so they are not included here. We briefly note that the R062 is comparably undersampled to the Z087 after

convolution with the pixel, but will have more high-frequency power as the pixel is assumed to be a perfectly sharp square. Thus if we could test R062, we may have found worse results than for the Z087. Fortunately, R062 is currently only planned to be used in the shallower and wider $z \lesssim 1$ tier with ~ 100 s exposure times. This yields an AB magnitude depth of almost exactly 1 mag shallower than our 300 s Z087 exposure. Thus R062 can tolerate a factor of a few worse host-galaxy residuals than can the Z087. In the other extreme, K213 is even better sampled than the data simulated in this work, so it should be straightforward to model and remove the K213 galaxy light.

We generate a copy of the images with noise included, but also retain the images without noise for testing. In addition to Poisson noise from the scene (SN + host galaxy), we include zodiacal (based on the model of Aldering 2002) and thermal background (Rubin 2020) given in Table 1. For read noise, we assume 106 reads for the 300 s exposures and 212 reads for 600 s, using 20 electrons per read with a 5 electron floor, giving values also displayed in Table 1 (Garnett & Forrest 1993; Vacca et al. 2004; Rauscher et al. 2007).¹⁰

Some forward-modeling codes for ground-based data (e.g., Astier et al. 2013) use only the sky noise (not the source Poisson noise or detector noise) to eliminate biases that would otherwise be caused by using noisy observations to estimate the Poisson noise. Our photometry is in an even more complex regime: SN Poisson noise, galaxy+sky Poisson noise, and detector noise all matter. We assume that the up-the-ramp readouts have been accurately fit, yielding count-rates with known, Gaussian-distributed uncertainties. As discussed in Section 2, we may have to forward model starting with the original detector readouts (which will be read-noise-limited for these faint sources) for satisfactory performance.

If the SN survey takes place over two years, with visits to the SN field(s) every five days, then there will be 146 visits. If a SN goes off at a random epoch, and every image taken after explosion at that location is considered contaminated with SN light, then the number of references will vary up to about 140 (assuming an absolute minimum of ~ 6 epochs required for a good light curve), but be ~ 70 on average. (Including lost epochs due to chip gaps, these numbers are ~ 124 and ~ 62 .) We simulate only a year of data, with an average of 43 reference images per filter, representing a below-average reference set compared to the planned survey, but still relatively representative. As we only simulate one year out of the planned two years of the survey, we do not incorporate

¹⁰ As most of the noise in the simulated observations is Poisson noise, we take a simple quadrature sum of the read noise and Poisson noise, without the $\sqrt{6/5}$ scaling on the Poisson noise appropriate for the read-noise-dominated regime. See Fadeyev et al. (2006) or Kubik et al. (2016) for a more detailed discussion of possible up-the-ramp estimators.

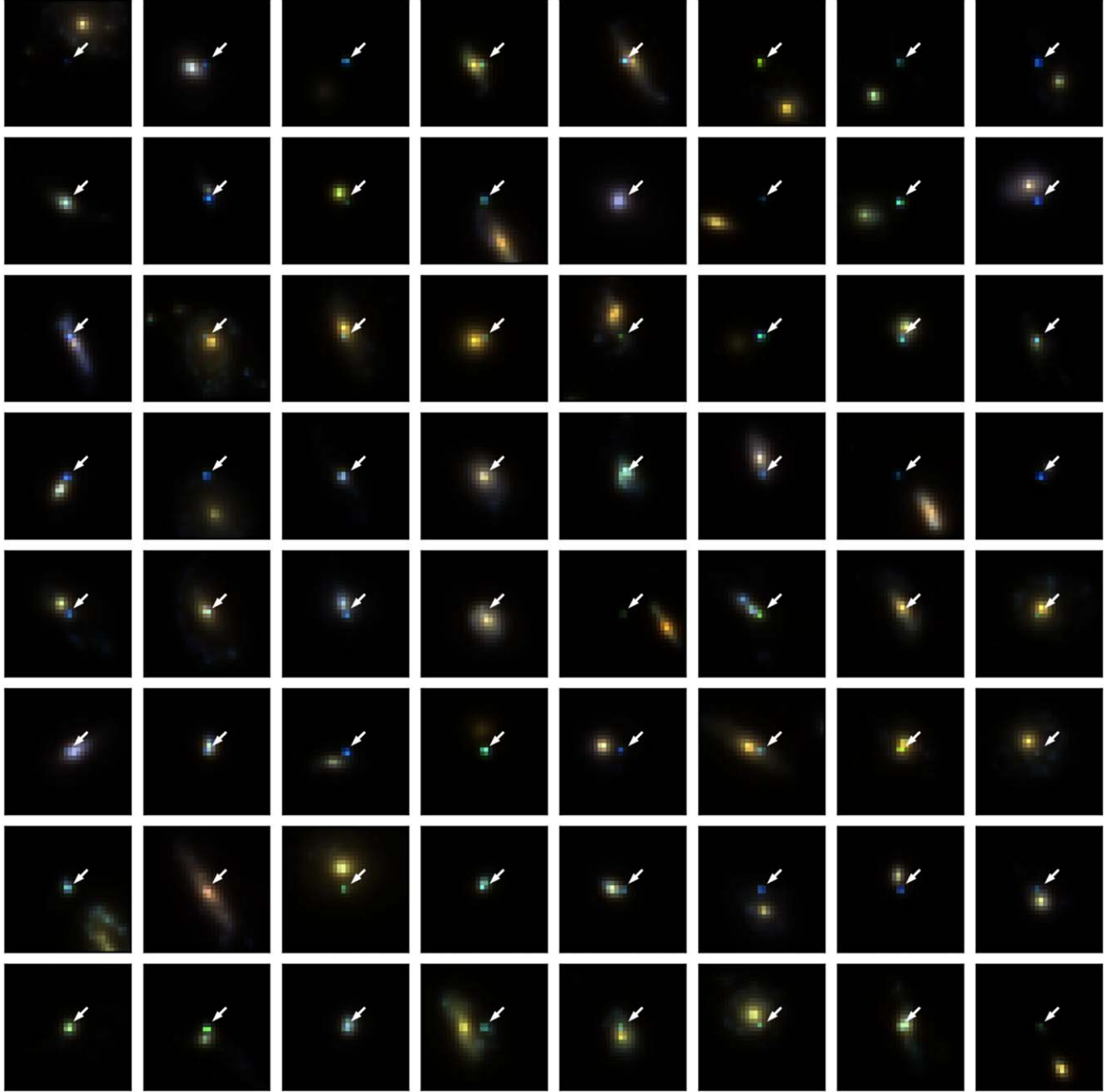


Figure 4. A randomly selected set of postage stamps from our simulations near maximum light for each supernova before the addition of noise. The arrows point out the SN locations. The color channels are Z087 (blue), J129 (green), and F184 (red). The subpixel positions are different for each filter, so we resample the Z087 and F184 images to match the J129 so that the colors align. The VELA galaxies are qualitatively similar to real SN host galaxies (e.g., Figure 2 of Riess et al. 2007). Some of the VELA viewing angles are aligned by angular momentum axis, giving a similar orientation for some galaxies in this figure.

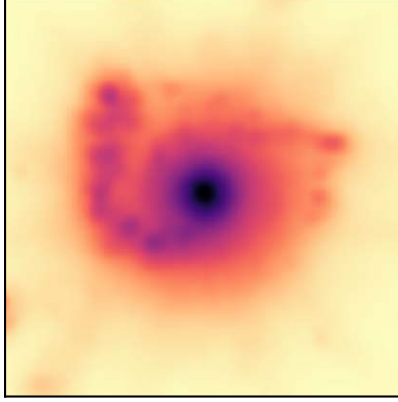
missing data, e.g., due to bad pixels or detector gaps, into the simulated images.

4. Photometry

Following Suzuki et al. (2012), we performed photometry of the SNe Ia in the WFI mock observations using a forward-model code

(the same one used by Hayden et al. 2021 and F. Ori et al. 2021, in preparation). This code fits analytic 2D-spline galaxy models (one independent model for each filter) which are convolved with PSFs (including convolution with the pixel) and resampled to match the images. As in Suzuki et al. (2012), the modeling uses $0''.01$ subpixels ($11\times$ oversampling). Our minimizer of choice is Levenberg–Marquardt (Levenberg 1944; Marquardt 1963).

Galaxy Convolved by PSF and Pixel,
Inverse Hyperbolic Sine Scale



Laplacian,
Linear Scale

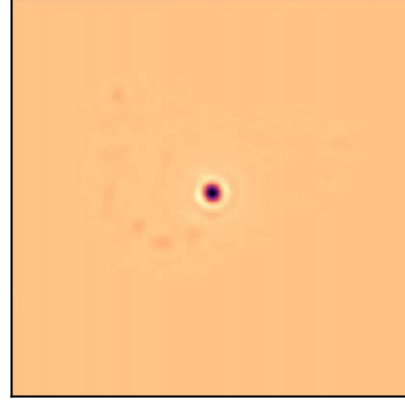


Figure 5. Visualization of the second derivative (Laplacian) for a galaxy. The left panel shows the central area around a VELA simulated galaxy, convolved with the PSF and the pixel. To better show faint features, this panel uses inverse hyperbolic sine scaling. The right panel shows the second derivative of the left panel. The largest deviations from zero are in a small region of the galaxy around the core. It is these regions that most require spatial flexibility in a galaxy model.

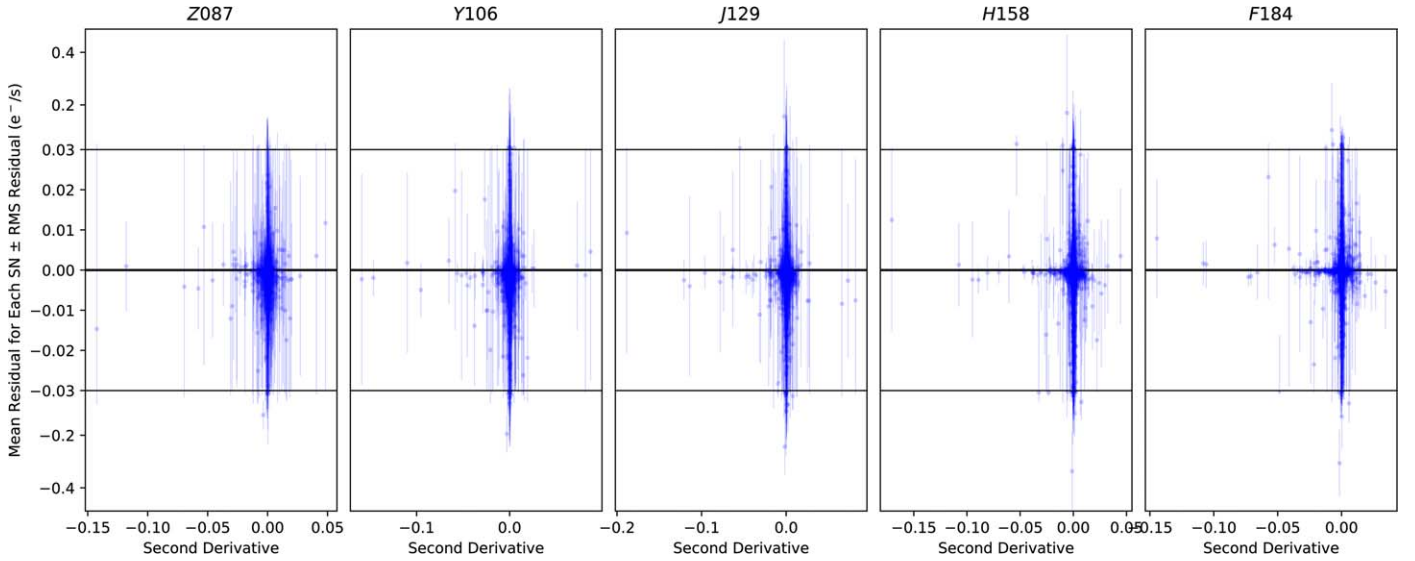


Figure 6. For each band, for each simulated SN, we show the mean flux residual (recovered flux – true flux) over the whole light curve (blue points) plotted against the second derivative of the galaxy flux at the SN location. As illustrated in Figure 5, the second derivative is close to zero over much of the galaxy after convolution with the PSF and the pixel. The error bar on each point shows the rms across epochs for that SN. To search for any possible trends with the highest sensitivity, we use the forward-model runs on the images that have no noise added. The typical flux at maximum light in this redshift range is $3 \text{ e}^- \text{ s}^{-1}$ for the four bluest bands and $1.5 \text{ e}^- \text{ s}^{-1}$ for *F184*, thus these results show very small residuals for almost all the simulated SNe. Note the piecewise-linear scale with an expanded view of -0.03 to $0.03 \text{ e}^- \text{ s}^{-1}$ (roughly $\pm 1\%$ of peak flux). Any trends with galaxy second derivative would indicate that the photometry may need a more flexible galaxy model (e.g., spacing the spline nodes closer together).

The flux of image i on a pixel x, y near the SN location is modeled as the sum of galaxy light $G(\alpha, \delta)$, background light s_i , and SN light (F_i , which is zero for epochs before explosion or ~ 1 rest-frame year after peak). The temporally unchanging galaxy model is evaluated in sky coordinates (R.A. α and decl. δ). These must be mapped to the $11 \times$ oversampled pixel coordinates with WCS functions $\mathcal{A}_i: x^{11 \times}, y^{11 \times} \rightarrow \alpha$ and

$\mathcal{D}_i: x^{11 \times}, y^{11 \times} \rightarrow \delta$. In our code, this is done using `Astropy all_pix2world`. These α and δ values are slightly adjusted with $\Delta\alpha_i$ and $\Delta\delta_i$ values for each image (it remains to be seen if the Roman WCS solution will be good enough to avoid these adjustments). The $11 \times$ oversampled galaxy model is convolved with the PSF and the pixel (also $11 \times$ oversampled), then this high-resolution model is sampled every 11 subpixels

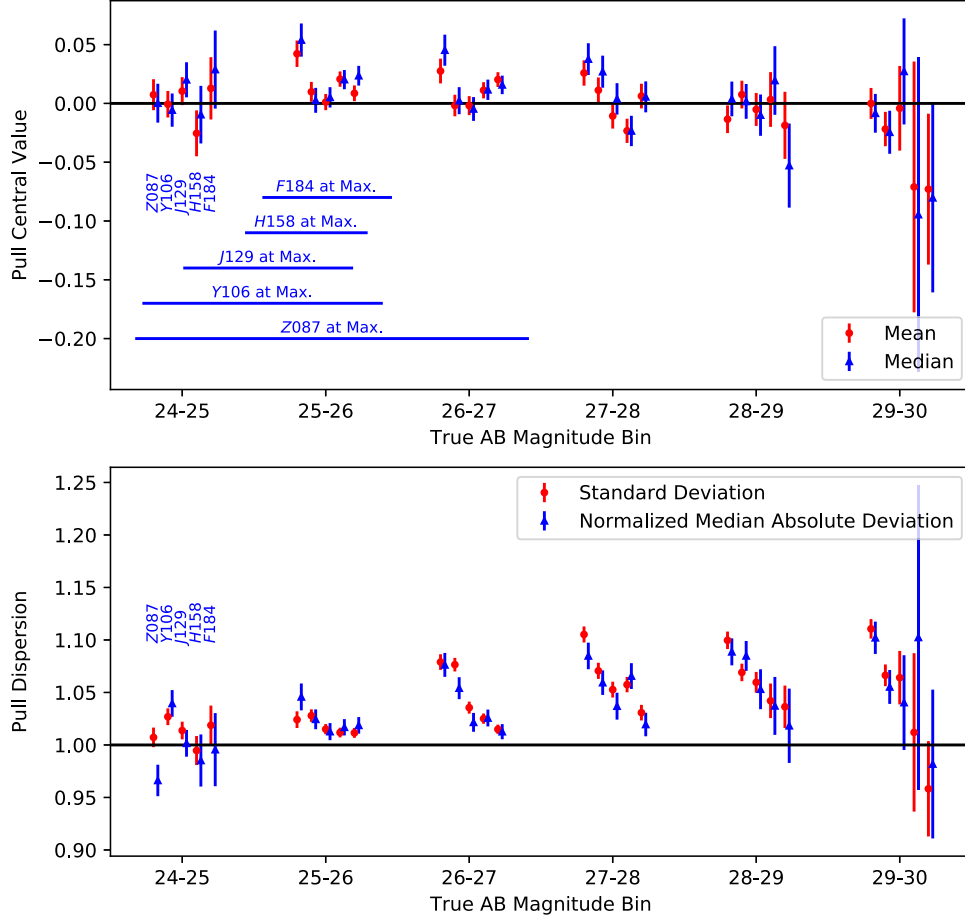


Figure 7. Summary statistics for pulls: (recovered flux – true flux)/recovered flux uncertainty. The results are binned in AB magnitude and separate results are shown for each filter (left to right is Z087 to F184). The top panel shows central values (mean with red dots, median with blue triangles). In general, there is no evidence for biases (offsets from zero) except in the Z087 filter in the middle of the magnitude range (~ 25 – 28). We also show the 16th to 84th percentile of SN fluxes at maximum for our simulated sample. The bottom panel shows dispersions (standard deviation with red dots, the normalized median absolute deviation with blue triangles). The uncertainties on the NMAD are computed with bootstrap resampling. If all uncertainties are correct and Gaussian, the dispersion values should be unity. There is evidence of mildly underestimated (by $\lesssim 10\%$) flux uncertainties in bluer filters for magnitudes fainter than ~ 26 .

(including at x, y). Thus,

$$G_i(x, y) = G(\mathcal{A}_i(x^{11\times}, y^{11\times}) + \Delta\alpha_i, \mathcal{D}_i(x^{11\times}, y^{11\times}) + \Delta\delta_i) \otimes \text{PSF}_i(x^{11\times}, y^{11\times})|_{x,y}. \quad (1)$$

The SN coordinates are also stored in sky coordinates ($\alpha^{\text{SN}}, \delta^{\text{SN}}$), but the PSF is in pixels. To convert, we need the inverse of the above WCS transform (`Astropy all_world2pix`): $\mathcal{X}_i: \alpha^{\text{SN}}, \delta^{\text{SN}} \rightarrow x_i^{\text{SN}}$ and $\mathcal{Y}_i: \alpha^{\text{SN}}, \delta^{\text{SN}} \rightarrow y_i^{\text{SN}}$. These are also adjusted with $\Delta\alpha_i$ and $\Delta\delta_i$, giving

$$x_i^{\text{SN}}, y_i^{\text{SN}} = \mathcal{X}_i(\alpha^{\text{SN}} + \Delta\alpha_i, \delta^{\text{SN}} + \Delta\delta_i), \quad (2)$$

$$\times \mathcal{Y}_i(\alpha^{\text{SN}} + \Delta\alpha_i, \delta^{\text{SN}} + \Delta\delta_i).$$

Finally, we can combine both models with the image-dependent, spatially flat sky s_i , giving the model $M_i(x, y)$:

$$M_i(x, y) = G_i(x, y) + \frac{F_i}{A(x, y)} \times \text{PSF}_i(x - x^{\text{SN}_i}, y - y^{\text{SN}_i}) + s_i. \quad (3)$$

We assume (as is true for WFC3 IR) that the flat-fielding preserves surface brightness, but point-source fluxes are scaled down proportional to the pixel area on the sky $A(x, y)$. Thus the SN model fluxes must also be scaled down to match the data. The galaxy model goes to zero at the edge of the circular fit patch (thus breaking the degeneracy between sky and galaxy light). Note that both the galaxy model and the SN require convolution with the PSF, but we do not insert the SN as a Dirac delta function into the $11\times$ oversampled galaxy model,

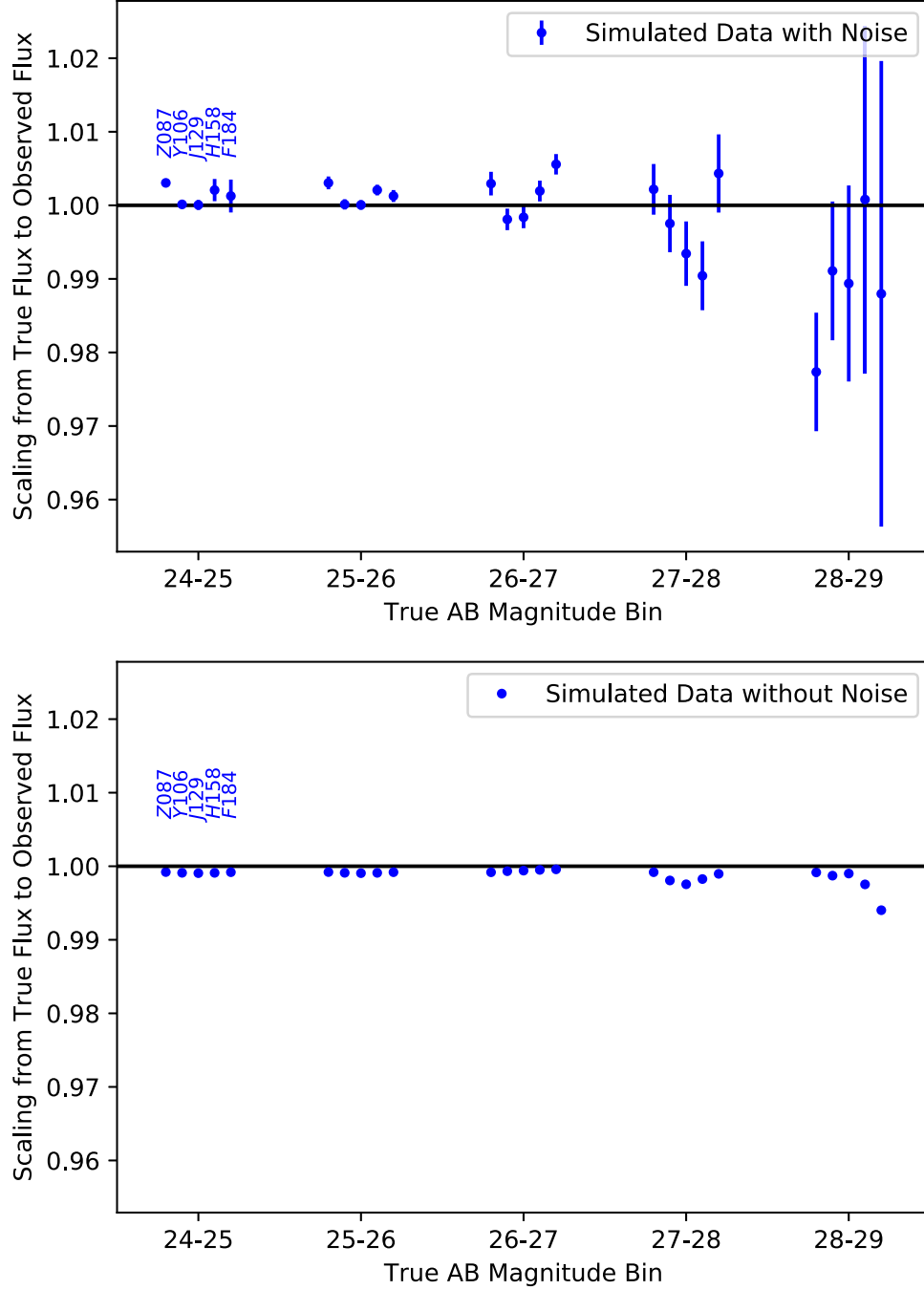


Figure 8. Average scaling on true flux to match observed flux, binned by AB magnitude and separated by filter. The top panel shows the results from the images with noise added. A small consistent bias is seen in the Z087 filter (leftmost point in each bin). The bottom panel shows the results from images without noise added. The brightest three magnitude bins show a slight bias (probably due to the way the data were generated as discussed in Section 3). The faintest magnitude bins show a larger (but still small) bias.

as the SN may land with light split between subpixels, broadening the PSF. Finally, we note that other extensions of the formalism (e.g., modifying the PSF shape as a function of image counts, Choi & Hirata 2020) are straightforward, but we do not consider these here.

There are many possible choices for the galaxy basis functions (Thevenaz et al. 2000). The general considerations for the galaxy basis functions are that they should be flexible enough to model real galaxies accurately, without being so flexible that the model is poorly constrained, amplifying noise.

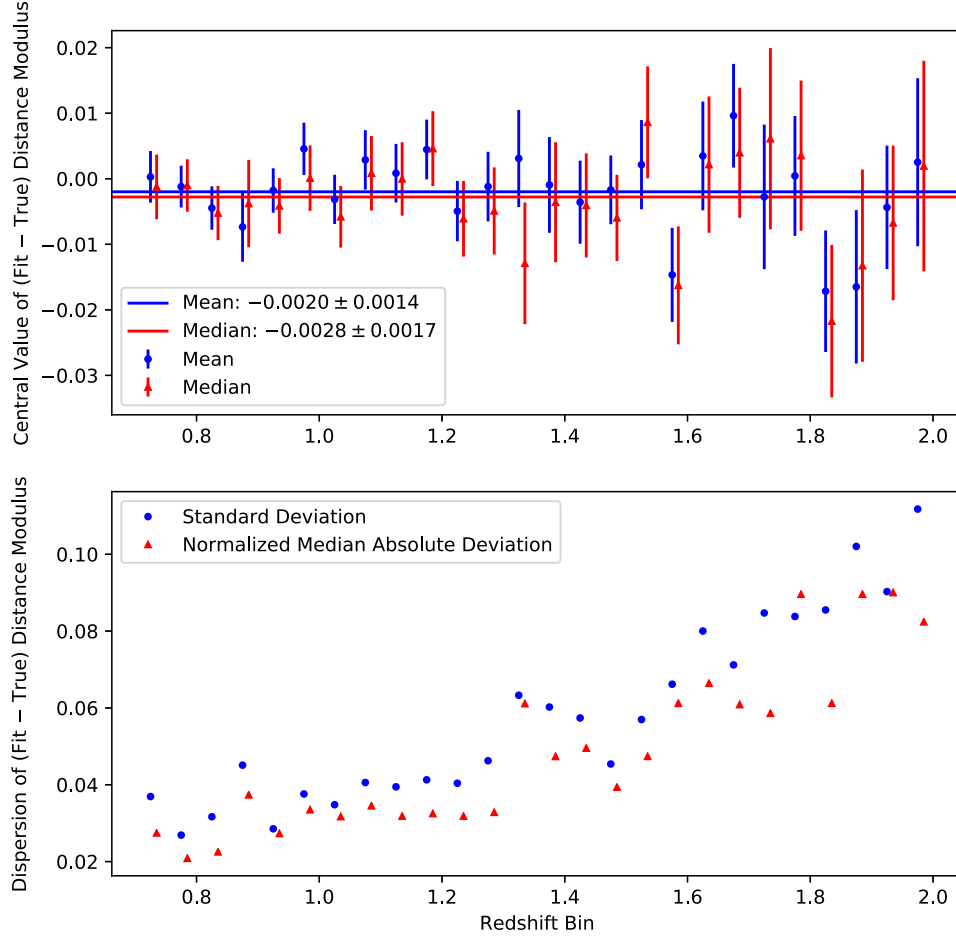


Figure 9. Summary statistics on recovered distance modulus compared to true distance modulus for each SN. The top panel shows the mean (blue circles) and median (red triangles) in bins of redshift. The blue and red lines show the result over all redshifts. No strong evidence of bias is seen. The bottom panel shows the dispersion in each bin (blue dots for the mean, and red triangles for the normalized median absolute deviation). The increase as a function of redshift is due to a combination of the lower signal to noise, and the loss of red rest-frame wavelength coverage.

Holtzman et al. (2008), Astier et al. (2013), and Brout et al. (2019) used a grid of squares of constant surface brightness. Smooth parameterizations are also used; Rodet et al. (2008) used Gaussians and Bongard et al. (2011) used sinc interpolation of a uniform grid. We want to maintain a smoother model for the undersampled data than a pixelized model or even Gaussians, and want a faster falloff than sinc interpolation (for a smaller modeled patch), so we use 2D splines. Suzuki et al. (2012) and Rubin et al. (2013) also used 2D splines, but added greater flexibility in the galaxy radial direction, giving an overall smoother model to reduce noise for SNe with limited numbers of references. Here, we have many reference epochs, so we simply use 2D splines with a uniform grid. Initially, we experimented with the spline node spacing. In the end, we settled on 2 nodes per PSF FWHM, effectively building an approximately Nyquist-sampled model. In other words, the

spline-node spacing for *J*129 was 0.62 pixels or $0''.068$ (Table 1).

5. Results and Discussion

We run several sets of analyses to investigate the results of the photometry. Our first result is that our spline-node spacing is sufficient. It is common in ground-based forward modeling to plot results against local galaxy surface brightness (e.g., Brout et al. 2019). However, the galaxy-light Laplacian (second derivative) is the better quantity for most types of modeling errors, as smooth galaxy gradients generally do not cause problems. For ground-based work, the host galaxy is frequently poorly resolved, and the local surface brightness correlates with the Laplacian. Figure 5 shows the second derivative for a typical galaxy. For each band, Figure 6 shows

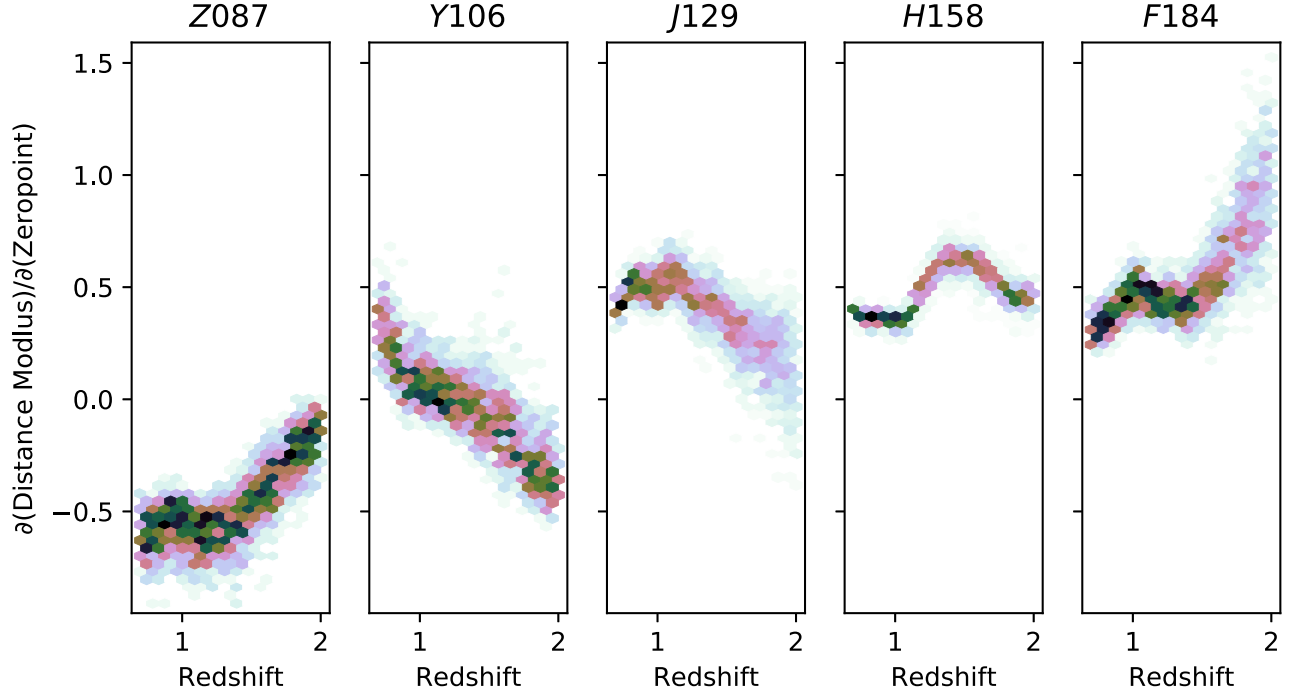


Figure 10. Sensitivity of the distance modulus of each SN to the calibration of each filter, plotted as a function of the redshift of each SN. As in Amanullah et al. (2010), this is computed by scaling the calibration for each filter in turn, refitting the SN with SALT2, and computing a new distance modulus. The distance modulus difference divided by the size of the shift in magnitudes gives the derivative. We do not propagate these calibration changes into the training of SALT2, which will change the results in detail (Guy et al. 2010). The sum of all the derivatives should be 1 (moving the calibration of each filter by 1 mag should move the distance modulus by 1 mag), and Amanullah et al. (2010) note that unstable light-curve fits frequently reveal themselves as deviations from 1. We exclude three such SNe from this plot which show up as outliers. For the five-band light-curves considered in this work, the sensitivity of the distance moduli to the calibration of any one filter is $\lesssim 1$. Limiting the wavelength range or the number of bands will significantly increase the sensitivity to calibration, resulting in much larger values than those shown here. For example, with just rest-frame B and V data (and using 3.1 for the slope of the color–magnitude relation), the distance moduli scale as $m_B - 3.1(m_B - m_V) = 3.1m_V - 2.1m_B$.

the noise-free residuals plotted against the local second derivative of the host-galaxy light; no trends are seen.¹¹

Next, we search for biases and check the uncertainties by plotting distributions of pulls: $(\text{recovered flux} - \text{true flux}) / (\text{recovered flux uncertainty})$. Figure 7 shows summary statistics, binned in true AB magnitude: $\text{zero-point} - 2.5 \log_{10}(\text{true flux})$. In general, the forward-model code has better performance in the redder filters with better sampling. The mildly underestimated ($\lesssim 10\%$) flux uncertainties in bluer filters at faint magnitudes might plausibly be due to the small galaxy-subtraction residuals shown in Figure 6.

Figure 8 shows observed flux regressed on true flux. We use both images with noise (top panel) and images without noise (bottom panel). For the images without noise only small biases

(~ 1 mmag or 0.1%) are seen until faint magnitudes. These are likely caused by the accuracy of the image reprojection (discussed in Section 3). At fainter magnitudes, the accuracy degrades, possibly due to the slight galaxy-subtraction residuals seen in Figure 6). For the results including noise, ~ 2 mmag biases are seen in Z087 but the other filters are generally consistent with unity mean scaling between true and observed fluxes.

Finally, we fit light curves using SALT2-Extended; Figure 9 shows these results. Any biases seem to be at the few mmag level or smaller. Figure 10 shows the sensitivity of our distance moduli to the calibration of each filter as a function of redshift. Our constraints on the cosmological bias are thus expected from the accuracy with which we recover the light-curve fluxes, but this test still uniquely measures any correlated effects of host-galaxy subtraction on the full light curves.

6. Summary

We validate a forward-model code for performing SN photometry in simulated undersampled images for the Roman transient survey. As there are no real images to inject simulated SNe into, we use the VELA simulated galaxy images, which

¹¹ For a simple comparison, we also perform photometry using a simple image-resampling code for the host-galaxy subtraction. For each image with SN light in it, we take each reference image and resample it to the pixels of that live-SN image. We use a square kernel with a size of 0.3 pixels (frequently called the *pixfrac*). After subtracting the references, we perform PSF photometry on the subtracted images. We only perform this test with the noise-free images to better examine the differences with forward modeling. This is an extremely simplistic resampling compared to more accurate procedures, e.g., Rowe et al. (2011) or Fruchter (2011). Unsurprisingly, it gives results that are much poorer than the forward model, with large negative slopes visible in all panels indicating over-smoothed galaxy models.


are generated in the Roman filters over a similar redshift range as the SN survey. We create 762,570 simulated postage stamps around the locations of 2061 simulated SNe ($\sim 10\%$ of the anticipated full survey). We describe the assumptions of our forward-model code and validate those assumptions first with noise-free images, and then with images that have noise added. Finally, we fit our simulated light curves and show that we can recover SN distance moduli with biases limited to less than a few mmag. The forward-model code has been released on Zenodo (Rubin 2021).

We thank Susana Deustua and the anonymous referee for their careful feedback. This work was supported by NASA through grant NNG16PJ311I (Perlmutter Roman Science Investigation Team). The technical support and advanced computing resources from the University of Hawai'i Information Technology Services Cyberinfrastructure are gratefully acknowledged. This work was also partially supported by the Office of Science, Office of High Energy Physics, of the U.S. Department of Energy, under contract No. DE-AC02-05CH11231. This research used resources of the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

Software: Astropy (Astropy Collaboration et al. 2013), Mathematica (Wolfram Research Inc 2020), Matplotlib (Hunter 2007), Numpy (Harris et al. 2020), Python (Van Rossum & Drake 2009), SciPy (Virtanen et al. 2020), SNCosmo (Barbary 2014).

ORCID iDs

David Rubin  <https://orcid.org/0000-0001-5402-4647>

Aleksandar Cikota  <https://orcid.org/0000-0001-7101-9831>

References

- Alard, C., & Lupton, R. H. 1998, *ApJ*, **503**, 325
- Aldering, G. 2002, SNAP Sky Background at the North Ecliptic Pole, Technical Report, LBNL-51157, doi:[10.2172/842543](https://doi.org/10.2172/842543)
- Amanullah, R., Lidman, C., Rubin, D., et al. 2010, *ApJ*, **716**, 712
- Anderson, J. 2016, Empirical Models for the WFC3/IR PSF, Instrument Science Report, *WFC3* 2016-12
- Anderson, J. P., James, P. A., Förster, F., et al. 2015, *MNRAS*, **448**, 732
- Astier, P., El Hage, P., Guy, J., et al. 2013, *A&A*, **557**, A55
- Astier, P., Guy, J., Regnault, N., et al. 2006, *A&A*, **447**, 31
- Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, *A&A*, **558**, A33
- Barbary, K. 2014, sncosmo/sncosmo: v1.4.0, Zenodo, doi:[10.5281/zenodo.592747](https://doi.org/10.5281/zenodo.592747)
- Bély, P., Hasan, H., Miebach, M., et al. 1993, Orbital Focus Variations in the Hubble Space Telescope, Instrument Science Report, SESD 93-16
- Bongard, S., Soulez, F., Thiebaut, É., & Pecontal, É 2011, *MNRAS*, **418**, 258
- Brout, D., Sako, M., Scolnic, D., et al. 2019, *ApJ*, **874**, 106
- Choi, A., & Hirata, C. M. 2020, *PASP*, **132**, 014502
- Fadeyev, V., Aldering, G., & Perlmutter, S. 2006, *PASP*, **118**, 907
- Fruchter, A. S. 2011, *PASP*, **123**, 497
- Garnett, J. D., & Forrest, W. J. 1993, *Proc. SPIE*, **1946**, 395
- Guy, J., Astier, P., Baumont, S., et al. 2007, *A&A*, **466**, 11
- Guy, J., Sullivan, M., Conley, A., et al. 2010, *A&A*, **523**, A7
- Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, *Natur*, **585**, 357
- Hayden, B., Rubin, D. A., Boone, K., et al. 2021, *ApJ*, **912**, 87
- Holtzman, J. A., Marriner, J., Kessler, R., et al. 2008, *AJ*, **136**, 2306
- Hounsell, R., Scolnic, D., Foley, R. J., et al. 2018, *ApJ*, **867**, 23
- Hsiao, E. Y., Conley, A., Howell, D. A., et al. 2007, *ApJ*, **663**, 1187
- Hunter, J. D. 2007, *CSE*, **9**, 90
- Jönsson, J., Sullivan, M., Hook, I., et al. 2010, *MNRAS*, **405**, 535
- Kubik, B., Barbier, R., Chabanas, E., et al. 2016, *PASP*, **128**, 104504
- Levenberg, K. 1944, *QApMa*, **2**, 164
- Marquardt, D. W. 1963, *J. Soc. Ind. Appl. Math.*, **11**, 431
- Mosby, G., Rauscher, B., Bennett, C., et al. 2020, *JATIS*, **6**, 046001
- Perrin, M. D., Sivaramakrishnan, A., Lajoie, C.-P., et al. 2014, *Proc. SPIE*, **9143**, 91433X
- Pierel, J. D. R., Rodney, S., Avelino, A., et al. 2018, *PASP*, **130**, 114504
- Rauscher, B. J., Arendt, R. G., Fixsen, D. J., et al. 2019, *JATIS*, **5**, 028001
- Rauscher, B. J., Fox, O., Ferruit, P., et al. 2007, *PASP*, **119**, 768
- Riess, A. G., Strolger, L.-G., Casertano, S., et al. 2007, *ApJ*, **659**, 98
- Rodet, T., Orieux, F., Giovannelli, J.-F., & Abergel, A. 2008, *ISTSP*, **2**, 802
- Rowe, B., Hirata, C., & Rhodes, J. 2011, *ApJ*, **741**, 46
- Rubin, D. 2020, arXiv:[2010.15112](https://arxiv.org/abs/2010.15112)
- Rubin, D. 2021, Forward-Model Photometry, Zenodo, doi:[10.5281/zenodo.4587879](https://doi.org/10.5281/zenodo.4587879)
- Rubin, D., Knop, R. A., Rykoff, E., et al. 2013, *ApJ*, **763**, 35
- Ryan, R. E. J., Casertano, S., & Pirzkal, N. 2018, *PASP*, **130**, 034501
- Scolnic, D., & Kessler, R. 2016, *ApJL*, **822**, L35
- Scolnic, D. M., Jones, D. O., Rest, A., et al. 2018, *ApJ*, **859**, 101
- Simons, R. C., Kassin, S. A., Snyder, G. F., et al. 2019, *ApJ*, **874**, 59
- Snyder, G. 2018, Vela-Sunrise Mock Observations (“Vela”), STScI/MAST, doi:[10.17909/T9-GE0B-JM58](https://doi.org/10.17909/T9-GE0B-JM58)
- Spergel, D., Gehrels, N., Baltay, C., et al. 2015, arXiv:[1503.03757](https://arxiv.org/abs/1503.03757)
- Suzuki, N., Rubin, D., Lidman, C., et al. 2012, *ApJ*, **746**, 85
- Thevenaz, P., Blu, T., & Unser, M. 2000, *IEEE Transactions on Medical Imaging*, **19**, 739
- Vacca, W. D., Cushing, M. C., & Rayner, J. T. 2004, *PASP*, **116**, 352
- Van Rossum, G., & Drake, F. L. 2009, Python 3 Reference Manual (Scotts Valley, CA: CreateSpace)
- Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, *NatMe*, **17**, 261
- Wolfram Research Inc. 2020, Mathematica, Version 12.1, <https://www.wolfram.com/mathematica>
- Zhou, Y., Apai, D., Lew, B. W. P., & Schneider, G. 2017, *AJ*, **153**, 243