Image Simulations for the *Nancy Grace Roman Space Telescope* Containing Injected Type Ia Supernovae

Kevin Wang

February 12, 2022

1 Introduction

1.1 Cosmology

Cosmology is the study of the structure and development of the universe as a whole. This is in contrast to astrophysics, which is the study of particular objects and processes that occur in space, although in the course of cosmological research we often use knowledge derived from astrophysics. The current standard cosmological model, known as the Λ-CDM model, includes two components that we do not understand fully, dark matter and dark energy. Dark matter behaves like matter gravitationally but does not interact electromagnetically, and is needed in the model because on many scales of the universe there appears to be more matter than is visible. Dark energy refers to a property of the universe that causes the universe to tend toward expansion rather than contraction like regular energy. We are interested in using type Ia supernovae (SNe Ia) to probe the nature of dark energy.

1.2 The Role of Supernovae

Fundamentally, it is difficult to determine how far objects are in space. Although objects which are further away from us are less bright, different objects have different intrinsic luminosities. Therefore, we first need objects for which we have a known intrinsic luminosity, called standard candles, to calibrate distances in space.

Type Ia supernovae arise when a white dwarf in a binary star system accretes enough matter from its companion star to reach the Chandrasekhar limit, which is the point when the electron degeneracy pressure is no longer sufficient to maintain the white dwarf against its own gravitational attraction. Since this is a specific physical process, the resulting supernova exhibits a consistent evolution and has a known luminosity.

Because SNe Ia can be as bright as an entire galaxy and thus can be seen out to very large distances, they are a key tool for probing cosmology.

1.3 Redshift and the Expansion of the Universe

By taking the universe as homogeneous and isotropic, which is an accurate description on large scales, we can obtain an exact solution to Einstein's field equations called the Friedmann-Lemaitre-Robertson-Walker (FLRW) metric. This metric allows for the universe to expand and contract uniformly, which we parametrize using a scale factor a that is set to 1 at present time. Then in 1922, Friedmann published a set of equations based on this metric describing the expansion of the universe in terms of the mass density of the universe as well as a cosmological constant that can be added in Einstein's field equations as a coefficient to the metric tensor. This cosmological constant is one possible form that dark energy could take in our universe.

Redshift describes the increase in wavelength of light as is usually denoted z and described by the equation

$$1 + z = \frac{\lambda_{obs}}{\lambda_{emit}}$$

The three sources of redshift are Doppler shift, caused by the relative motion of emitter and observer, gravitational redshift as described by general relativity, and cosmological redshift, caused by the expansion of the universe.

In 1929, Hubble published an article in which he described a proportionality between the distance of a galaxy from Earth and the redshift of that galaxy. This relationship is evidence that the universe is expanding. Figure 1 shows a Hubble diagram from **?** created using a modern sample of SNe Ia. (Note that the redshift is logarithmic, and the distance is expressed by a distance modulus which is also logarithmic in distance).

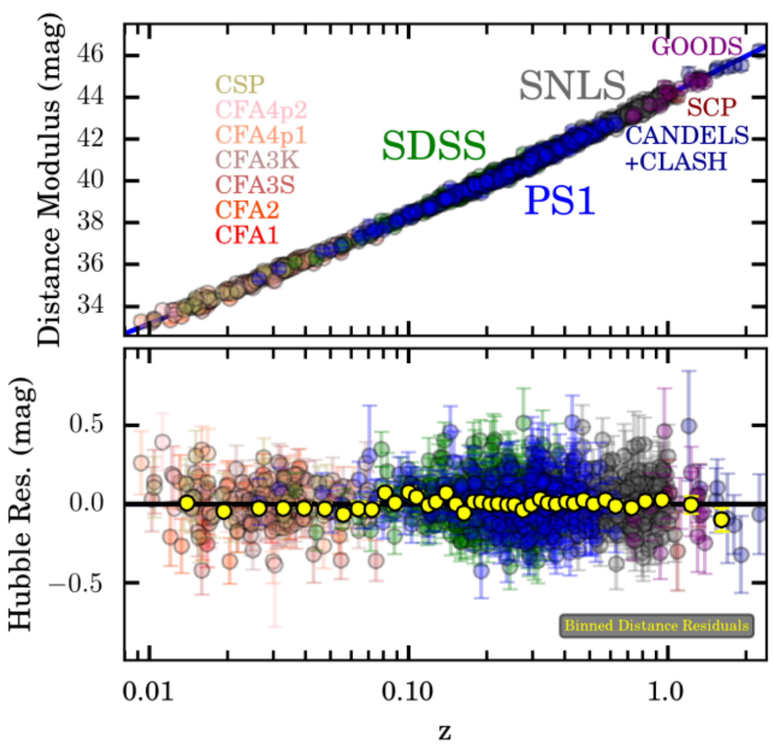

Figure 1: Hubble diagram from **?**

Then in 1998, two research teams studying SNe Ia found that distant supernovae have brightnesses that are consistent with a universe that containing a significant cosmological constant component, which also means the expansion of the universe is actually accelerating at this moment. The SNe were fainter than would be expected if the universe contained only matter (**?** and **?**).

1.4 Cosmological Models

This finding forms one basis of the current standard cosmology, which postulates a flat universe that contains about 5% baryonic matter, 25% cold dark matter, and 70% dark energy in the form of a cosmological constant. However, this is not the only possible form that dark energy could take. In a simple cosmology, we treat matter, radiation, and dark energy as perfect fluids characterized by an equation of state

$$w \equiv \frac{p}{\rho}$$

which gives the relationship between the pressure and the energy density of that fluid.

Using this equation, a cosmological constant has $w = -1$. First, this means that we can essentially treat a cosmological constant as a fluid with negative pressure. Second, this means that unlike matter and radiation, whose energy density decreases as the size of universe increases, a cosmological constant has constant energy density.

However, we are also interested in probing other models of dark energy that are dynamical rather than constant. One such parametrization takes the form of an addition to the dark energy equation of state that depends linearly on the scale factor

$$w = w_0 + w_a(1 - a)$$

1.5 The Role of ***Roman***

One limitation of SNe Ia is that they are quite rare, and therefore we do not currently have a large enough sample of SNe Ia to strongly constrain the dark energy parametrization.

Current samples of SNe Ia number around 1,000 and only a small subset have both $z > 1$ and a well-sampled light curve. SNe Ia with a high redshift are needed to see further back into the history of the universe in order to see any effects of dark energy evolution. For all SNe, we want measurements of the SN to be closely spaced in time in order to obtain the a good fit for its light curve. Currently, SNe Ia results combined with other probes like Planck's CMB data give the best measurements of dark energy as $w_0 = -1.007 \pm 0.089$ and $w_a = -0.222 \pm 0.407$, which is still not a strong constraint on the dynamic dark energy model (**?**).

NASA will launch the *Nancy Grace Roman Space Telescope* (*Roman*) in the mid-2020s. *Roman*'s cosmological observations will have two main components, an extensive wide-area survey for measuring weak lensing and clustering and a smaller-area survey covered at regular time intervals, called the time-domain survey, for detecting supernovae and other transients.

Over two years of observing, Roman is expected to observe about 10,000 SNe Ia, which will provide much more precise measurements of dark energy. *Roman* will have a dedicated time-domain survey that will sample these SNe Ia at a regular cadence of 5 days, and two subsections of the survey are currently planned to target SNe Ia at redshifts of $z = 1$ and $z = 1.7$. Referring back to Figure 1, having high quality data at these higher redshifts would allow us to fill out and extend our Hubble diagrams, which is ultimately an indication of how far back into the universe's history we can probe.

? perform simulations of a variety of *Roman* survey strategies to obtain estimates of *Roman*'s future constraints on dark energy. With optimistic estimates on systematic uncertainties, they get standard deviations for w_0 and w_a of 0.035 and 0.16 respectively for a imaging strategy like the one simulated, which is a substantial increase over current measurements.

2 Preparing for *Roman*

2.1 Overview

Well before *Roman* is launched, many preparations need to be done in order to ensure that the telescope will actually be a useful tool for cosmology. Survey strategies need to be developed that balance the requirements of diverse projects with the observing time and other resource constraints of the telescope. Simulations need to be created to investigate potential sources of systematic bias as well as the effects on the resulting cosmological measurements, allowing us to make adjustments to survey strategy.

Here we can make a distinction between two types of simulations for supernova cosmology. Catalog-level simulations involve generating and selecting a collection of SNe Ia based on a survey strategy and a theoretical model. Their main advantage is that they take fewer resources, and therefore they can be adjusted more easily and can cover a wide range of survey and generating strategies.

Simulated images are needed in order to prepare the tools that will be used to process the images, the pipeline, as well as perform analyses that cannot be done without pixel-level information. In addition, results from analyzing images can then be used to adjust assumptions used for catalog-level simulations.

The base of this project is simulation code that for weak lensing analysis from **?**. Adding on to this code, I created the first image simulations containing SNe Ia for the time-domain survey. The set of images consists of 40,000 images over 4 filters totalling 3 TB of space, covering one deg^2 in area and containing about 1,000 SNe Ia.

We plan to release these images publicly so that they can be used by others for cosmological analysis and are currently discussing with scientists from IPAC at Caltech about the logistics of transferring and hosting these images there.

Figure 2 provides an overview of the image simulation process with the products that will be provided publicly highlighted in red. Parts of this process will be discussed in more detail in the following sections.

2.2 Generating Supernovae

We generated the SNe Ia used in the images with SNANA, a program written by Rick Kessler at University of Chicago that generates realistic supernova light curves (**?**). In supernova cosmology, SNe Ia are actually often referred to as standardizable candles because their absolute luminosities are not completely consistent but are found to be associated both with the width of the light curve as well as the color of the SN. SNANA uses a model called SALT2 for these effects. Many details of our supernova generation are also based on earlier catalog-level simulations (**?**).

The primary addition in our work relates to host galaxies, since each supernova is associated with its own host by SNANA during the generation process. Since

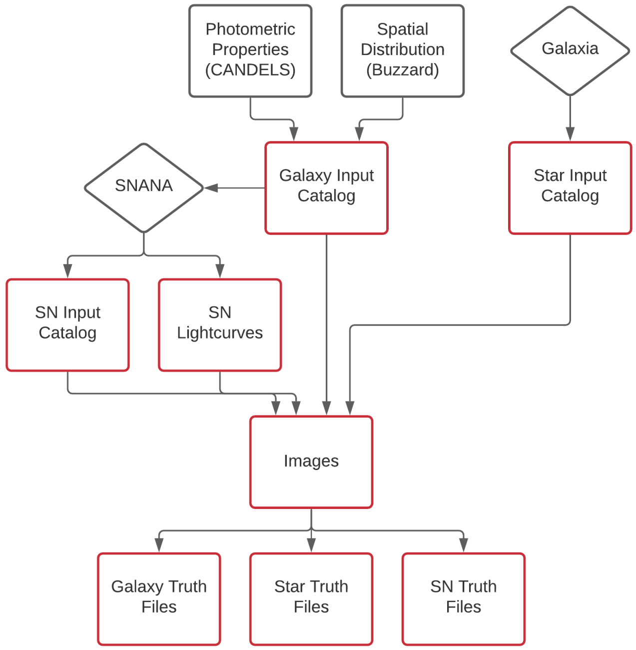

Figure 2: Flowchart depicting the process through which output products are created. Boxes labeled in red are the output files that are included in the data release.

we need to have an input galaxy catalog for drawing galaxies in order to create realistic images, it makes sense to use these galaxies as the hosts. Following **?**, galaxies are generated with positions drawn from a Buzzard simulation, which uses an algorithm to populate dark matter halos created using an N-body dark matter simulation with galaxies. Their properties are drawn from the CANDELS survey, which was taken using the Hubble Space Telescope. We use a different CANDELS catalog as **?** because they made certain selection cuts for weak lensing purposes that are undesirable for host selection purposes.

Steve Rodney, a professor who we worked with at the University of South Carolina, then helped with creating a version of the host library input for SNANA that also took into account the mass and star formation rates of the host galaxies. The goal was to make supernovae more likely to occur in more massive and more active galaxies, resulting in a more realistic distribution of host galaxies, which we succeeded at.

2.3 Generating Images

With the generated light curves in hand, I can inject the SNe onto the images as point sources. The basic procedure is as follows: a position, angle, and date on the sky is given. All objects from the input catalogs falling onto the region of the sky are selected and are drawn onto the image using GalSim (**?**), open-source software used to create image simulations [1], and truth files are created containing information about each object. GalSim has a *Roman* module containing details about the telescope that is used for the following steps.

[1] `https://github.com/GalSim-developers/GalSim`

Imaging systems are described using a point spread function (PSF), the response of that system to the light from a point source. The PSF of all telescopes are affected by diffraction limits but also by the particular aberrations of the telescope. Importantly, as a space-based telescope, *Roman*'s PSF will not be affected by atmospheric turbulence, unlike those of ground-based telescopes.

The image result from an object is the convolution of the object's light and the PSF, which is usually done in Fourier space because of the convolution theorem which says that a product in Fourier space is the convolution in image space.

After the objects have been drawn in this manner, each image is run through a series of modifications that simulate various effects that occur in the physical detector to obtain a final image. These include reciprocity failure, the breakdown of the linear relationship between flux and response at low light; dark current, Poisson noise caused by thermal electrons; and interpixel capacitance, where pixels electrically affect neighboring pixels.

Figure 3 shows a time series of a SN next to its host galaxy. The changing brightness of the SN can be seen to match the variation in brightness in the generated light curve.

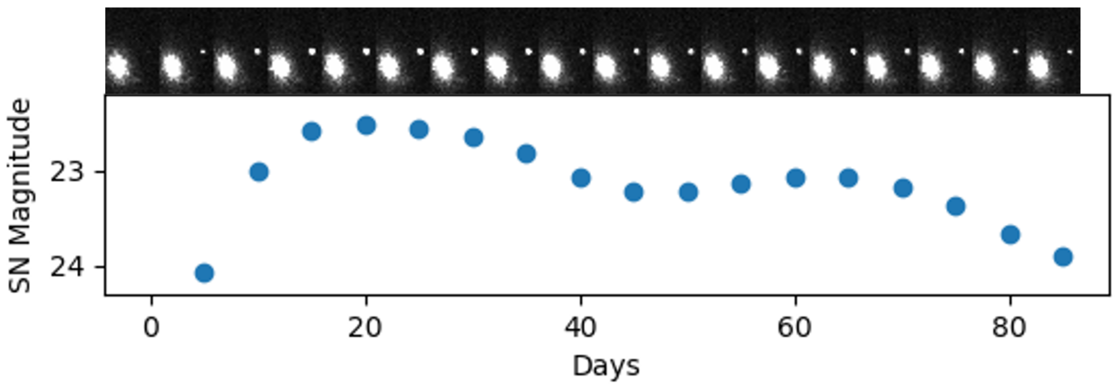

Figure 3: Simulated time series of a supernova in intervals of five days with corresponding light curve underneath in Y band. In each of the images, the supernova is on the right with its host galaxy on the left. The images and light curve are aligned so that the points on the light curve match the image at the corresponding time.

3 Using the Images

3.1 Image Processing

Several processing steps need to be done in order to reach the full potential use of the images taken by *Roman*. One step is coadding images, which means combining the information from multiple images that overlap on the sky. For various reasons, astronomical images are not usually aligned exactly, so we use a software package called AstroDrizzle which handles the details (**?**). Figure 4 shows a comparison between a single image and the corresponding area in a coadded image.

Coadded images will have a higher signal-to-noise ratio and we will be able to detect fainter objects in them compared to single images. We have set up the process so that we can run the coaddition process on an entire set of images, although there

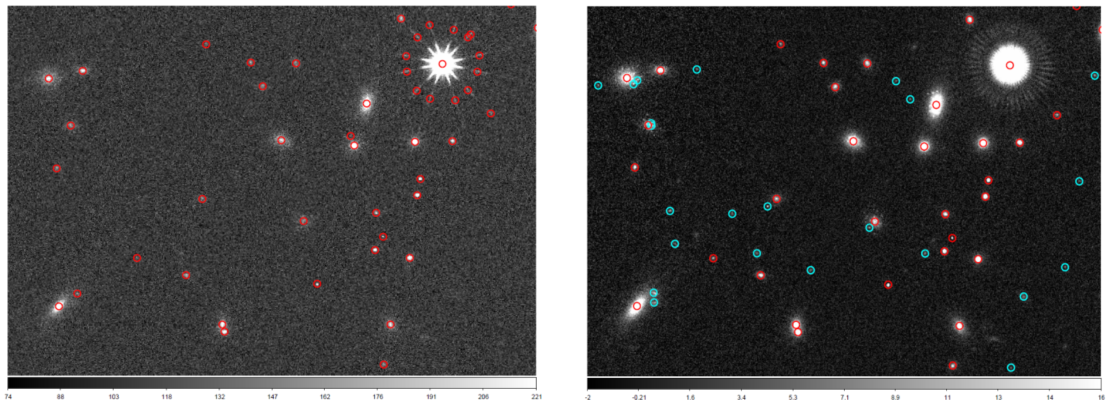

Figure 4: Left: A single image with objects detected using Source Extractor circled in red. Right: The same area in the sky in a coadded image made by combining 64 images. The actual pixel depth is between 10-15 for most areas on the coadd. Objects found only in the coadded image are circled in cyan while objects found in the coadded image that were also found in the single image are circled in red.

are still limitations right now about how many images can be put into one coadded image.

Another step that is used specifically for transients is difference imaging. Beginning with a given image, we want to know whether there are any SNe in it. We first need what is called a template or reference image, which is usually a coadded image in order to get the best possible signal-to-noise ratio. By subtracting the two images (and again there are details of alignment that need to be dealt with), any objects which have constant luminosity over time will be removed, leaving only transients. Figure 5 shows an image along with a cutout of an area containing a SN as well its difference image. Note that this difference image was not created using a coadd but with a single image from a different time since we do not have the entire process set up at this time, but were simply making an example image.

3.2 Science Analysis

In addition to the processing steps, I also performed basic science analyses in order to check for bugs in the all of previous work as well as to demonstrate potential uses of these images. Objects were detected in the simulated images using Source Extractor, a piece of software that generates a catalog of objects in an image (**?**). The basic behind SExtractor is that it first creates an estimate of the background level in an image, and then objects can be counted as detected if they fall above a threshold above the background. Afterwards, a set of algorithms is used to separate objects, classify them as galaxies or stars, and extract their properties. I manually adjusted the detection thresholds to obtain reasonable detection rates without too many false positives, but otherwise kept the default settings and did not look at other tools for detecting objects.

The most basic analysis is measuring the detection rates of galaxies of various brightnesses. We found that galaxies of up to 25 magnitude could be detected in our single images and up to 26 magnitude in our coadded images (which are 10-15 images deep). For stars, which are point sources and thus easier to detect, we can

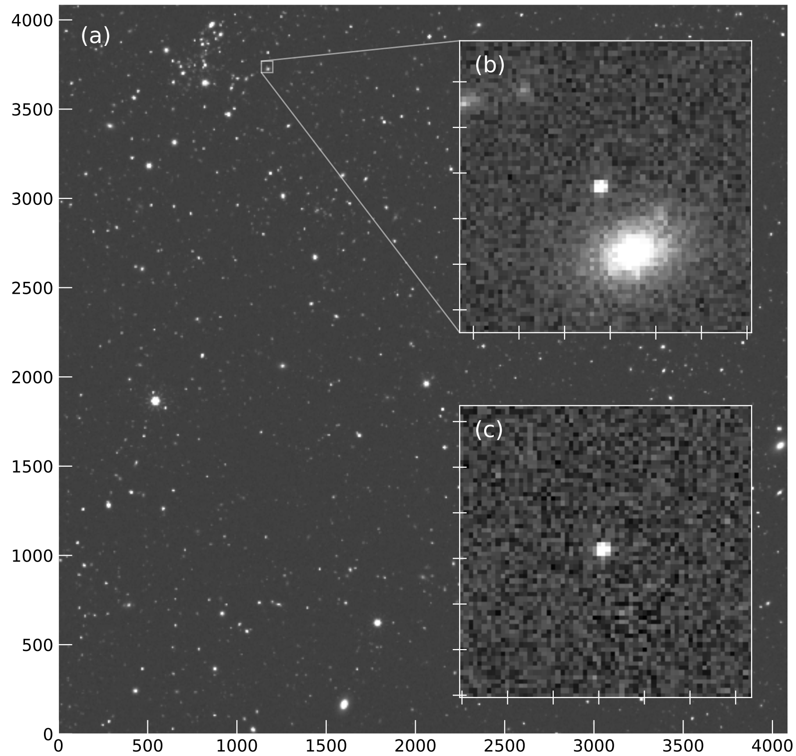

Figure 5: A simulated Y band image for one *Roman* SCA. (a) The full frame of the simulated image, comprising 4k×4k pixels (roughly 7.5′ on a side). (b) Zoom-in on a 10″ ×10″ region, centered on a bright SN with magnitude 22.59. (c) The difference image of the same 10″ ×10″ region, generated by subtracting a single-epoch template image taken after the SN has faded, with the same rotation and pointing center.

detect up to 26 magnitude in single images and up to 27 magnitude in the coadds.

We also did a very basic test of detection rates of supernovae, but this is limited by the fact that the SNe can be on bright galaxies. Since we do not have a complete difference imaging setup, we instead choose to look only at SNe far from their hosts. For these SNe, we have some detections when they are below 25 magnitude and do not have detections above that value. Although the limitations mean these numbers are not useful for cosmological purposes, they do provide an indication that the supernova generation and image creation process were successful.

Finally, we are interested in how reliably we are able to determine the true hosts of the SNe from the images. For cosmological analyses, associating the wrong galaxies as host for SNe could cause a systematic bias in our measurements. For this host association problem, I take the closest galaxy to each SN and check whether it matches the assigned host (I also tried a more sophisticated association method that includes the shape of the galaxy but found that it does not give better results). Figure 6 plots the redshift of the host galaxy against the redshift of the closest galaxy.

Blue dots on the line indicate correctly associated galaxies; green dots were incorrectly associated because the host galaxy was not detected at all; orange dots indicate host galaxies that were detected but were not the closest object to their SN. Association rates were quite high at low redshifts but fall off at higher redshifts when many host galaxies are too faint to be detected. For a cosmological analysis

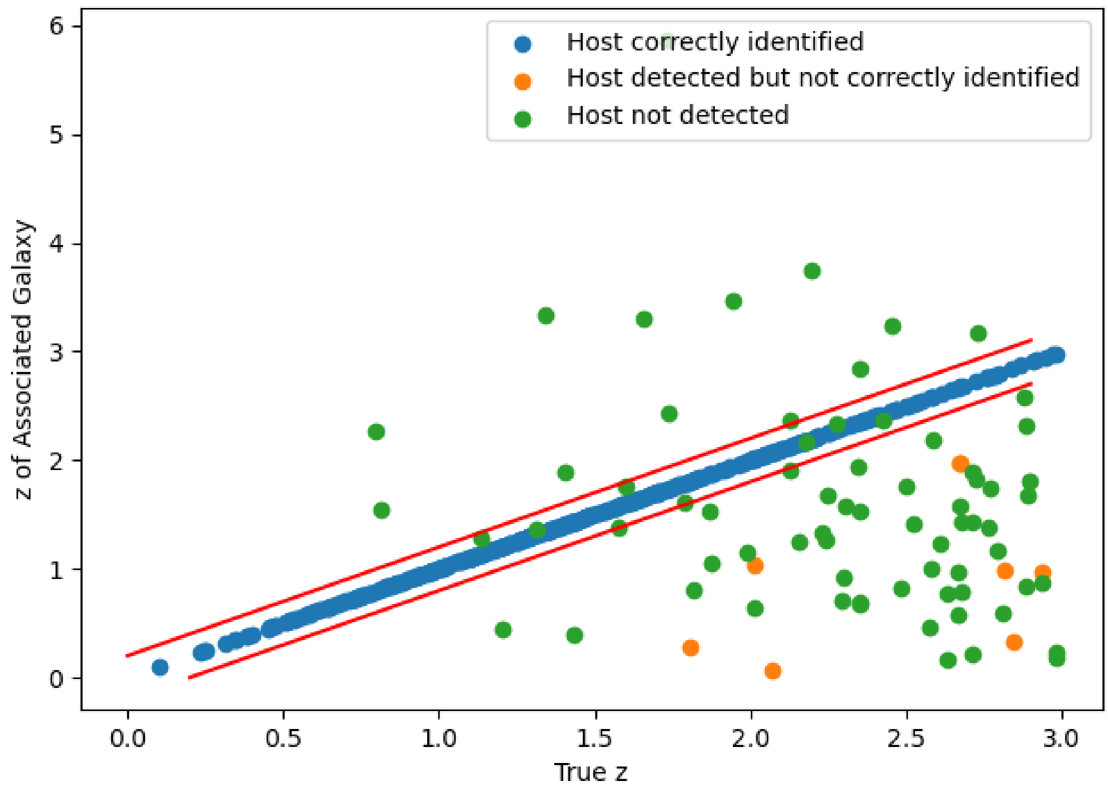

Figure 6: A scatterplot of the host galaxy redshift taken from truth files and the redshift of the closest galaxy detected from each SN. Correct assignments are shown in blue, when the host is detected but another galaxy is assigned as the host is shown in orange, and when the host is not detected and another galaxy is assigned as host is shown in green.

we would want to exclude obvious redshift mismatches between a SN and potential host galaxies. The red lines indicate where SNe and galaxy redshifts differ by more than 0.2.

4 Next Steps

Once the set of images is made publicly available, supernova cosmologists can use them for any analysis they choose. As far as the analyses presented in this report, each one has more or less work that can be done to make it more useful.

Starting with the detection efficiency for galaxies and stars, the detection limits are not unreasonable. However, we actually expect to be able to see SNe down to 26.7 magnitude whereas we only see stars to 26 magnitude (remember both SNe and stars are point sources) in our images. In this case the task is to determine whether this is an issue with the threshold I chose for detection, whether another detection method would be superior, or whether it is a real difficulty in detecting objects in images that is responsible.

The actual detection and analysis of SNe is the furthest from being suitable for cosmology, because it necessitates a complete process including both coaddition and difference imaging. Because it is so important to have these, if the images are to be widely used, a lot more needs to be done on these aspects. However, it is also the release of the images that will enable this work to be done by more than just us at Duke and our collaborators.

As far as the results from the host association go, they are encouraging for the good performance at low redshift and interesting because we found that the simpler association method based on distance was more successful. However, there is still work to be done on making the association procedure more sophisticated. One big point is that instead of cutting out redshift mismatches at the end, to incorporate it into the association process. We might find that if this is implemented that the distance-based method is no longer superior. The association process is also dependent on detection, so changes and improvements there will also affect it.

Finally, I am working on a paper about these image simulations. After the paper is complete, I will likely move on to another project. Some potential projects, like looking at the effect of lensing on SNe, are connected to the image simulations and would use them. Regardless of whether my future project is related to these simulations, I will still be available for support or clarification to scientists who are using these images. This could include up to rerunning the simulations major bugs are discovered or there are important changes or features that need to be added.