

No Evidence for Type Ia Supernova Luminosity Evolution: Evidence for Dark Energy is Robust

B. M. ROSE,¹ D. RUBIN,^{2,3} A. CIKOTA,³ S. DEUSTUA,¹ S. DIXON,^{4,5} A. FRUCHTER,¹ D. O. JONES,⁶ A. G. RIESS,^{1,7} AND
D. M. SCOLNIC⁸

¹*Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218, USA*

²*Physics and Astronomy Department, University of Hawaii, Honolulu, HI 96822, USA*

³*Physics Division, E.O. Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA, 94720 USA*

⁴*E.O. Lawrence Berkeley National Laboratory, 1 Cyclotron Rd., Berkeley, CA, 94720*

⁵*Department of Physics, University of California Berkeley, 366 LeConte Hall MC 7300, Berkeley, CA, 94720-7300*

⁶*Department of Astronomy and Astrophysics, University of California, Santa Cruz, CA 92064, USA*

⁷*Department of Physics and Astronomy, Johns Hopkins University, Baltimore, MD 21218, USA*

⁸*Department of Physics, Duke University, 120 Science Drive, Durham, NC, 27708, USA*

(Received February 26, 2020)

Submitted to The Astrophysical Journal Letters

ABSTRACT

Type Ia Supernovae (SNe Ia) are powerful standardized candles for constraining the cosmological model and provided the first evidence of accelerated expansion. Their precision derives from empirical correlations now measured from > 1000 SNe Ia between their luminosities, light curve shapes, colors and most recently a modest relationship with the mass of their host galaxy. As mass correlates with other host properties, these have been investigated to improve SN Ia standardization though none have been shown to significantly alter the determination of cosmological parameters. We re-examine a recent claim, based on 34 SN Ia in nearby passive host galaxies, of a 0.05 mag/Gyr dependence of standardized SN Ia luminosity on host age which if extrapolate to higher redshifts, might accrue to 0.25 mag challenging the inference of dark energy. We reanalyze this sample of hosts using both the original method and a Bayesian Hierarchical Model and find after a fuller accounting of the errors the significance for a dependence on age to be $\leq 2\sigma$ and $\sim 1\sigma$ after removal of a single poorly-measured SN. To test the claim that a trend seen in old stellar populations can be applied to younger ages, we extend our analysis to a larger sample which includes young hosts. We find the residual dependence of host age (after all standardization typically employed for cosmological measurements) to be $0.0011 \pm 0.0018 \text{ mag/Gyr}$ (0.6σ) for 254 SNe Ia from the Pantheon sample, consistent with no trend and strongly ruling out the large but low significance trend claimed from the passive hosts.

Keywords: dark energy, distance scale, supernovae: general, supernovae: individual (SN2003ic)

1. INTRODUCTION

Type Ia supernovae (SNe Ia), the thermonuclear explosion of carbon-oxygen white dwarfs, have been used as cosmic distance indicators for over 30 years. Their observed variability can be empirically corrected (Pskovskii 1977; Phillips 1993; Hamuy et al. 1996; Riess et al. 2016; Perlmutter et al. 1997), allowing for precise luminosity distances resulting in the measurements

of the energy density of matter and dark energy, driving the expansion history of the Universe (Garnavich et al. 1998a; Riess et al. 1998; Garnavich et al. 1998b; Perlmutter et al. 1999). Since then, there have been improvements in standardization methods (Jha et al. 2007; Guy et al. 2010; Burns et al. 2011; Mosher et al. 2014) and resulting cosmological measurements (Suzuki et al. 2012; Betoule et al. 2014; Riess et al. 2018; Scolnic et al. 2018; DES Collaboration et al. 2019; Jones et al. 2019; Freedman et al. 2019).

As the number of SNe Ia at cosmological distances now exceed 1000, the selection criteria have become more

stringent. Of all the SNe Ia that are observed, roughly 75% (Scolnic et al. 2018) are used for cosmology. Cosmologically useful SNe Ia are required to have sufficient data: adequate sampling around the peak of the light curve and of the decline rate. They must also pass “quality cuts” i.e. their parameters should be nearest the centers of the population distributions and thus can be precisely standardized through the empirical correlations. Even after standardization, outliers exist, resulting in outlier rejection tools (Kunz et al. 2007; Rubin et al. 2015) and even the classification of a new class of transients (Foley et al. 2013).

The accuracy of SN Ia cosmological measurements requires the absence of a redshift dependence of the standardized luminosity, which we refer to as luminosity evolution. This is not the same as a change in the mean of the properties used to standardize SNe Ia which may result from sample selection and is commonly referred to as a change in demographics. Luminosity evolution may be constrained locally ($\lesssim 400$ Mpc) by measuring differences in standardized SN Ia luminosity between galaxy types that predominately form their progenitor stars recently versus in the distant past as a proxy for a change in redshift or cosmic time. Over the last decade large samples with strict quality control have revealed correlations between host galaxy properties and standardized peak luminosity of modest significance and level (e.g. Gallagher et al. 2008; Kelly et al. 2010; Sullivan et al. 2010; Lampeitl et al. 2010; Gupta et al. 2011; Rigault et al. 2013; Jones et al. 2015; Moreno-Raya et al. 2016; Uddin et al. 2017; Kim et al. 2018; Rigault et al. 2018; Jones et al. 2018; Rose et al. 2019). Each of these measurements agree in direction of the host galaxy effect, and it is clear that these are not all by chance. Since the average galaxy changes with redshift and sample selection, it has become necessary to include such correlations in the standardization process to limit biases to the $< 1\%$ level in distance. The first recognized and most commonly used host property for such standardization has been host mass (Kelly et al. 2010; Sullivan et al. 2010; Lampeitl et al. 2010), where there has been seen a ~ 0.06 mag change in Hubble-Lemaître residual across a division in host stellar mass. Hubble-Lemaître residuals are the difference between the measured luminosity distance and the expected distance from the best-fit cosmology.

There are additional tests for luminosity evolution, beyond correlations with galaxy properties. For instance, one can determine if general light curve correlations and SN Ia spectra behave similar locally and at a redshift of one (Scolnic & Kessler 2016; Foley et al. 2008, respectively). Like all evidence for luminosity evolution, these

changes would be beyond the expected changes in population demographics.

Kang et al. (2020) (hereafter K20) claim to find a correlation between the ages of 34 early-type host galaxies — derived from spectral features — and their SN Ia distances. If extrapolated to younger ages and higher redshifts convolving look-back time and SN Ia progenitor models, could cause a significant redshift dependent luminosity evolution, $\Delta\text{mag}/\Delta z > 0.2$ mag. The original discovery of dark energy from SNe Ia (Riess et al. 1998; Perlmutter et al. 1999) ruled out such a large evolution in standardized luminosity by demonstrating consistency between SN Ia in early-type hosts and those in young, star-forming hosts. K20 claim a much larger trend between galaxy age and SNe Ia standardized luminosity, that along with other assumptions, this would bring into question the evidence for dark energy. However, we have serious concerns about the accuracy of this analysis, and here we show when these are addressed the strong evidence from SNe Ia for dark energy is unchanged.

This does not mean that we disagree with the motivation of the K20 work — that correlations between SN Ia properties and their hosts exist, and that these will need to be better characterized to significantly improve upon present cosmological measurements. However, we show here that these correlations are not significantly limiting our current ability to use SNe Ia to measure the cosmological parameters of our universe.

2. RE-EXAMINING THE K20 ANALYSIS

Between the two paper series Kang et al. (2016) and K20 observed 51 early-type low redshift SN Ia host galaxy, obtaining extremely high signal-to-noise galaxy spectra ($S/N \sim 175$, Kang et al. 2016). The SN Ia are archival, spanning from 1990 to 2010, and aggregated into a uniform re-analysis in the YONSEI SN catalog (Kim et al. 2019). It is the extremely high quality spectra that makes this data set unique, most SN Ia host galaxy research uses photometry (e.g. Gupta et al. 2011; Jones et al. 2018; Rose et al. 2019) but some use lower signal-to-noise integral field unit spectra (Rigault et al. 2013, 2018). The methodology used in K20 are not new nor controversial, for they built on previous research, such as Gallagher et al. (2008).

However, our re-examination of the claims by K20 can be grouped into three major areas: first, the quality of the SN Ia used, second, the robustness and overall significance of the trend in Hubble-Lemaître residual with stellar population age, and third, the extrapolation of

this trend to SNe Ia from younger stellar environments, and hence, redshifts.

2.1. SN Ia Data Quality

The SN Ia used in [K20](#) are archival, and as such are of varying quality. However, calculating Hubble-Lemaître residuals is directly effected by the quality of the SN Ia data. First we want to bring out that there are several SN Ia with extremely poorly sampled light-curves and as such would have very uncertain Hubble-Lemaître residuals. SN2007ap and SN2008af have no data prior to +5 days post-maximum. SN2003ch, SN2003ic, SN2003iv, and SN2007cp have less than seven nights observed, with half of these having less than four nights. Finally SN1993ac and SN2001ie have both no data prior to +5 days post-maximum and less than seven nights observed. Of these eight SN Ia with questionable to bad light curves, only SN2003ch and SN2007cp were removed from the final data set used in [K20](#). Therefore, the reliability of the Hubble-Lemaître residual measurements of 6 out of 34 SN Ia can be called into question.

To test assumptions used in cosmological analyses, it is critical that the subset of SN Ia are reasonable. They do not necessarily need to be a fully representative sample, but they should at least all pass the typical quality cuts. Using the Joint Light-Curve Analysis as an example ([JLA](#), [Betoule et al. 2014](#), Table 7), we see that 4 of the final 34 SN Ia do not pass quality cuts. Several others of the initial [K20](#) SN Ia sample also fail. SN2002do and SN2007au have $x_1 < -3$. These fast decliners are outside the valid range of the SALT2 model requiring alternative methods ([Garnavich et al. 2004](#)). In addition, SN2002do, SN2004gc, SN2006kf, and SN2008ia are highly extinguished by the Milky Way, having $E(B - V)_{MW} \geq 0.15$. The more dimming and reddening from Milky Way dust, the less accurate the SN Ia peak luminosity can be. For this reason, cosmological analyses typically use SNe Ia that are out of the plane of the Milky Way. The Pantheon analysis ([Scolnic et al. 2018](#)) performs very similar quality cuts. Other analyses include additional cuts on the phase coverage of the light curves. For example, ([Rest et al. 2014](#)) requires at least one observation between -10 and +5 rest frame days after maximum brightness, at least one observation between +5 and +20 days after maximum, and at least 5 total observations between -10 and +35 days after maximum. There are 4 SNe Ia (SN1993ac, SN2001ie, SN2007ap, SN2008af) in the final sample of [K20](#) that fail the first cut, and another SN Ia (SN2003ic) fails the second.

A summary of which SN Ia fails what cut can be seen in Table 1. Using several definitions of SN Ia quality, we

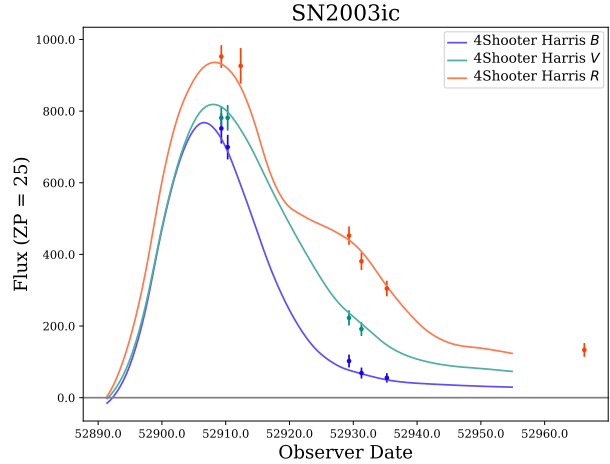


Figure 1. The light curve of SN2003ic. With no pre-maximum data points, it would traditionally be removed from cosmological samples due to difficulties constraining the peak luminosity. However this SN Ia alone takes the correlation seen in [K20](#) from a 1.8σ to a 2.3σ significance. Photometry from [Hicken et al. \(2009\)](#).

find 10 SNe Ia that fail at least one quality cut. Using just the JLA cosmology cuts, nearly 12% (4 out of 34 objects) of the final sample is not of cosmological quality, resulting in unreliable Hubble-Lemaître residuals. Our first conclusion is that [K20](#) properly cleans the data for defects in the host galaxies, but does not perform the same treatment to their SNe Ia.

2.2. Correlation Between Hubble-Lemaître Residual and Age

The reduced χ^2 statistic (χ^2_ν) of the fiducial [K20](#) correlation is ~ 2 indicating significant unaccounted for uncertainty. For SN Ia at low redshift (the [K20](#) sample is at $z < 0.04$) this is typically from local peculiar motion with an uncertainty σ_v , and the unexplained scatter seen in SN Ia post standardization ($\sigma_{\text{unexplained}}^{\text{mag}}$).¹ Further, if one accounts for expected flows using maps of large-scale structure on a SN Ia-by-SN Ia basis, as undertaken by [K20](#), a peculiar velocity uncertainty floor remains due to the unpredictable motions local to each SN Ia. Pantheon ([Scolnic et al. 2018](#)), calculate this to be $\sigma_v = 250 \text{ km s}^{-1}$ after estimating velocity corrections. An certainty floor like this significantly increases distance uncertainties at these low redshifts and reduces the significance of correlations. The total distance uncertainty of a SN Ia is comprised of many individual uncertainties. A relevant example, based on the Pantheon

¹ [K20](#) uses a common alternative name, intrinsic dispersion (σ_{int}).

Table 1. K20 SN Ia that do not pass typical “quality cuts”

	LC Quality		JLA Cuts		Rest et al. (2014)	
	≥ 1 obs.	Total num obs.	$ x_1 < -3$	$E(B - V)_{MW} \geq 0.15$	≥ 1 obs.	≥ 1 obs.
	$t < +5$ days	> 7			$-10 < t < +5$ days	$+5 < t < +25$ days
SN1993ac ^{ac}	X	X			X	
SN2001ie ^{ac}	X	X			X	
SN2002do ^b			X	X		
SN2003ic ^{ac}		X				X
SN2003iv ^a		X				
SN2006kf ^b				X		
SN2007ap ^{ac}	X				X	
SN2007au ^b			X			
SN2008af ^{ac}	X				X	
SN2008ia ^b				X		

^aHas a poorly sampled light curve.

^bFails JLA quality cuts, defined in Betoule et al. (2014).

^cFails phase coverage cuts, defined in Rest et al. (2014).

NOTE—From the final 34 SN Ia in the K20 sample, 6 SN Ia have poorly sampled light curves, 4 would not pass the JLA cuts, and 5 would not pass the Rest et al. (2014) phase coverage cuts.

analyse of Scolnic et al. (2018), is

$$\sigma_{\text{total}}^2 = \sigma_N^2 + \sigma_{\mu-z}^2 + \sigma_v^2 + \sigma_{\text{unexplained}}^{\text{mag}^2} \quad (1)$$

where σ_N^2 is the photometric error of the SN Ia distance and $\sigma_{\mu-z}^2$ is the uncertainty from the redshift. The last two terms (in bold) were ignored in K20. These ignored uncertainties increase the total distance uncertainty by $\sim \sqrt{2}$, would increase the statistical significance of any measured correlation.

If a new correlation were to explain the unexplained scatter, obviating the need to include it in the uncertainty, the result should be a value of $\chi_\nu^2 \sim 1$ which is not the case for the K20 analysis. K20 did not include any unexplained scatter in SNe Ia luminosities post-standardization or a peculiar motion error floor. Their inclusion is important because correctly estimating the significance of any trend is tied directly to accurate accounting of the uncertainty in the data and a realistic error in a trend is only determined if $\chi_\nu^2 \sim 1$. Including both $\sigma_v = 250 \text{ km s}^{-1}$ velocity uncertainty and an unexplained scatter of $\sim 0.10 \text{ mag}$ as used in most cosmological analyses (Suzuki et al. 2012; Betoule et al. 2014; Rubin et al. 2015; Scolnic et al. 2018; DES Collaboration et al. 2019) brings $\chi_\nu^2 \sim 1$ for the K20 sample and the significance of the age trend becomes 1.9σ .

In addition, the measured Hubble-Lemaître residual-age trend (K20, Figure 13) visually appears to be strongly dependent on SN2003ic, the SN Ia with the oldest host. As seen in Figure 1 and addressed in Section 2.1, the light curve of SN2003ic is very poorly sampled, including no pre-maximum measurements and only two epochs closely spaced in time to sample the the first 15 days of decline, the most valuable span of time for calibrating the light curve decline rate. If SN2003ic was removed, the trend shifts from $-0.051 \pm 0.022 \text{ mag/Gyr}$ (2.3σ) to a less significant $-0.045 \pm 0.024 \text{ mag/Gyr}$ (1.8σ) using the original K20 data and 1.5σ when attempting to get the $\chi_\nu^2 \sim 1$, as discussed previously. Removing other poorly sampled SN Ia does not effect the trend as much as SN2003ic. A summary of each Hubble-Lemaître residual stellar age correlation discussed in this paper, can be found in Table 2.

As both are $< 2\sigma$ our second conclusion is that with realistic errors or the exclusion of a single poorly observed SN Ia (i.e., as is seen in a basic jack knife test), there is no statistically significant trend with age in the K20 data.

However, looking for a trend between Hubble-Lemaître residuals and any host galaxy property can easily ignore cross correlations with the SN Ia standardization terms (Hamuy et al. 1995, 2000; Smith et al. 2020). Therefore, to further test the observed trend in

Table 2. Summary of the correlations between Hubble-Lemaître residual and stellar age

Method	Correlation [mag/Gyr]	Significance	χ^2	Num. SN Ia
K20	-0.051 ± 0.022	2.3σ	64.4	34
K20 w/o SN2003ic	-0.045 ± 0.024	1.8σ	64.1	33
K20 plus 250 km/sec velocity uncertainty	-0.047 ± 0.022	2.1σ	37.8	34
above plus 0.10 mag intrinsic error	-0.046 ± 0.024	1.9σ	27.0	34
above w/o SN2003ic	-0.037 ± 0.025	1.5σ	26.4	33
UNITY	-0.035 ± 0.023	1.5σ	...	34
UNITY w/o SN2003ic	-0.013 ± 0.022	0.6σ	...	33
UNITY with $\sigma_{\text{unexplained}}^{\text{age}}$	-0.12 ± 0.30	0.5σ	...	34
Spearman correlation coefficient	...	2.0σ	...	34
Pantheon Hubble-Lemaître residuals	-0.016 ± 0.031	0.5σ	26.3	27
Pantheon w/o SN2003ic	$+0.008 \pm 0.030$	0.3σ	20.4	26
Pantheon Sample				
standard analysis	-0.0011 ± 0.0018	0.6σ	257	254
without mass step	-0.0026 ± 0.0018	1.2σ	275	254
$\alpha=0.15$, $\beta=3.69$ from K20	-0.0021 ± 0.0018	0.6σ	280	254
late-type hosts only	$+0.0008 \pm 0.0023$	0.3σ	131	134
early-type hosts only	-0.0064 ± 0.0054	1.2σ	39.9	36

[K20](#), we sampled a simple standardization equation in the Bayesian hierarchical model UNITY² ([Rubin et al. 2015](#); [Rose et al. 2020](#)). We used a typical Tripp-like linear standardization ([Tripp 1998](#)):

$$\mu = m_B - (M_B + \alpha x_1 + \beta c + \gamma a) \quad (2)$$

where μ , m_B , M_B are the distance modulus, apparent and absolute magnitude respectively. The α , β , and γ parameters are the linear standardization coefficients corresponding to the SALT2 ([Guy et al. 2010](#)) light-curve shape (x_1) and intrinsic color (c), as well as the host galaxy age in gigayears (a). The parameters m_B , x_1 , c , and a are unique for each SN Ia, whereas M_B , α , β , and γ are data set variables that are fit simultaneously along with the cosmological parameters of interest. UNITY also fits for the remaining unexplained scatter ($\sigma_{\text{unexplained}}^{\text{mag}}$) allowing for the additional term, γ , to ex-

plain more of the observed SN Ia variability without over estimating the significance of any parameter.

The resulting UNITY parameter estimation, using the fiducial sample from [K20](#), is shown in Figure 2. Using UNITY, the significance of the [K20](#) trend with age (γ) is reduced to 1.5σ . These results from UNITY suggest that the significance of any Hubble-Lemaître residual-host galaxy correlations are typically over estimated. As is necessary for an accurate error estimation, we included the non-diagonal covariance terms from the light-curve fitting; [K20](#) only report diagonal covariance terms. Due to this missing data and some inconsistencies between the values reported in [K20](#) and the original YON-SEI SN catalog ([Kim et al. 2019](#)) (e.g. the x_1 value of SN2002G), we used the results from our own light-curve fits. Without SN2003ic, a non-zero γ drops to only a 0.6σ significance. Our third conclusion, like the second, is that the simultaneous consideration of the standardization parameters, appropriate when considering a new parameter, shows that standardizing with host galaxy

² https://github.com/rubind/host_unity

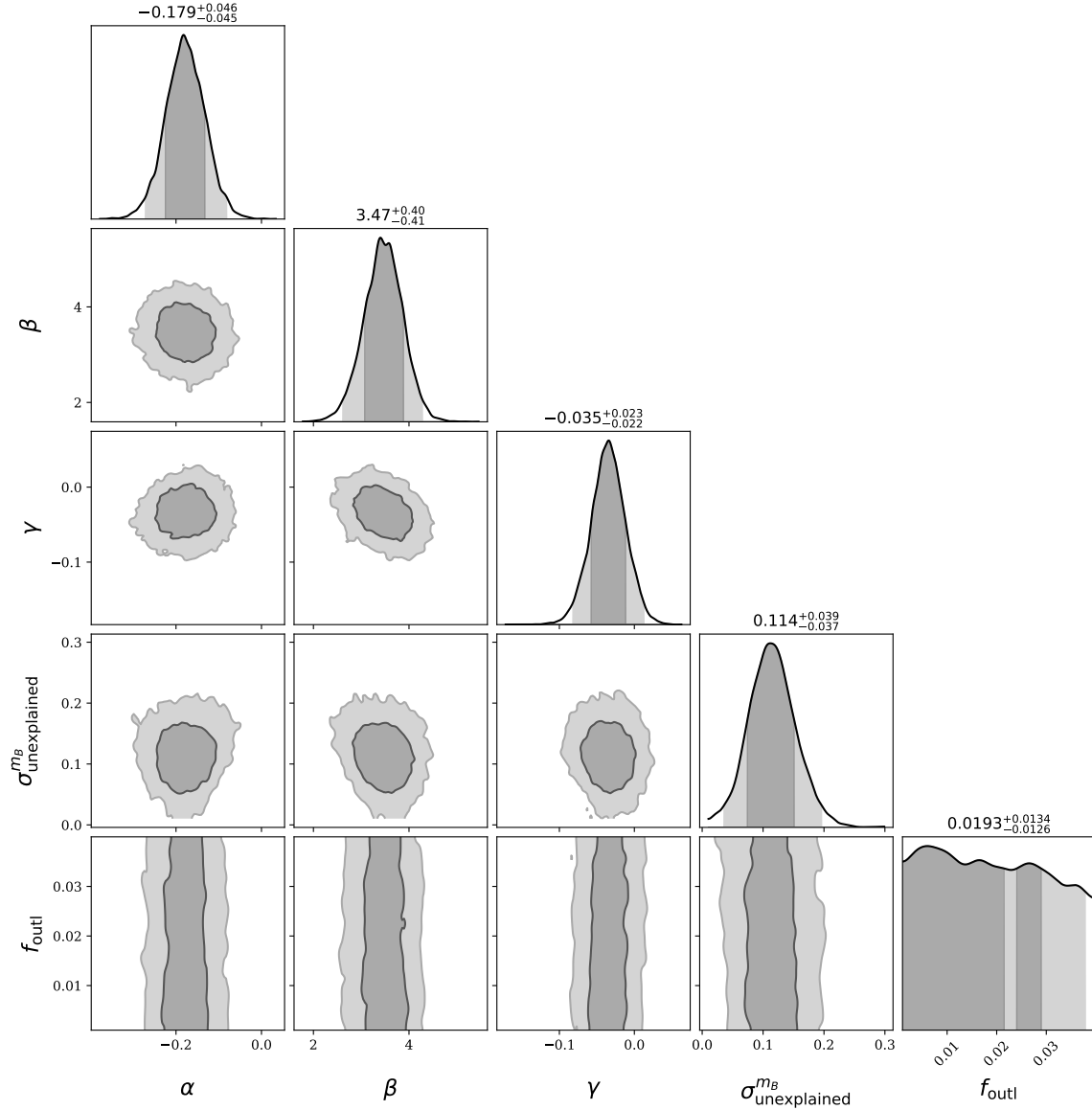


Figure 2. A UNITY re-analysis of the SN Ia standardization parameters including a term correlating peak magnitude with YEPS ages (γ). Dark and light shaded areas show 1σ and 2σ credible regions, respectively. The typical SN Ia standardization parameters (α and β) are within expected ranges, as is the unexplained scatter in standardized peak luminosity ($\sigma_{\text{unexplained}}^{m_B}$). Unlike the claim of K20, the standardization with age (γ) is only seen at a 1.5σ significance. The outlier fraction (f_{outl}) is consistent with zero, as expected for a small handpicked sample.

age is not statistically significant, nor robust to the removal SN2003ic.

The high χ^2_ν values seen in K20 indicates the possible need for additional uncertainty in the age measurements. To test this, we added an additional parameter to UNITY ($\sigma_{\text{unexplained}}^{\text{age}}$). This parameter is added in quadrature to each quoted age uncertainty term, resulting in a new total age uncertainty and is modeled off the standard method of adding σ_v to the total uncertainty, Equation (1). If $\sigma_{\text{unexplained}}^{\text{age}}$ is consistent with zero, then the reported age uncertainties should be considered re-

alistic. Figure 3 shows a highly significant non-zero additional age uncertainty. On average, the uncertainties on the YEPS (Yonsei Evolutionary Population Synthesis, Chung et al. 2013) ages do not fully explain the variance by 1.6 ± 0.3 Gyr. As expected, with proper uncertainties, the significance of γ falls to 0.5σ . Our fourth conclusion is that the uncertainties reported from YEPS are underestimated, on average, by 1.6 ± 0.3 Gyr, and as a result inflating the significance of any non-zero parameter measurement.

However, we are able to ignore the disputed uncertainties ($\sigma_{\text{unexplained}}^{\text{age}}$, σ_v^2 , and $\sigma_{\text{unexplained}}^{\text{mag}^2}$) and measure a correlation’s significance directly from the scatter in the data. This is done via correlation coefficients. The Pearson correlation coefficient is the most common, but assumes both that the trend is linear and that each data set is normally distributed. There is no expectation that the age values would be normal, in fact, [Chil-dress et al. \(2014\)](#) predicates them to be non-normal, and [Rose et al. \(2019\)](#) and others have observationally confirmed that prediction. The Spearman rank-order correlation coefficient does not have these requirements. When using the final data set of [K20](#), the Spearman correlation coefficient is $r_s = -0.35$, slightly higher than the $r_s = -0.23$ correlation seen in [Rose et al. \(2019\)](#). Uncorrelated variables producing a data set that has a Spearman correlation at least as extreme, is possible at the 2.0σ level, slightly lower than the 2.1σ significance seen in the larger [Rose et al. \(2019\)](#) sample. Bypassing any question about the accuracy of the uncertainties, this trend appears only marginally significant.

Via several alternative analysis methods — both accounting for additional known uncertainties and bypassing them — we have seen the correlation is at most 2σ , but likely less. Three of our previously stated conclusions show a reduced certainty and significance of this trend.

2.3. Propagation to Cosmology

Our second set of concerns are based around how [K20](#) extrapolate a correlation with age to a bias in cosmology. As discussed in Section 2.2, the correlations between Hubble-Lemaître residual and host galaxy age is highly dependent on a unique data set of [K20](#) that is not typical of cosmological samples.

First, [K20](#) use a simple argument using the cosmic star formation rate and the SN Ia delay time distribution to predict the mean progenitor age of a SN Ia as a function of its redshift ([K20](#), Figure 15). We note that the nature of SN Ia progenitors remains highly uncertain and the correct such model will depend on whether SN Ia arise from the merger of two white dwarfs or accretion from a companion star onto a white dwarf. Thus, such a prediction is highly speculative and [K20](#) do not consider any additional uncertainty in their statement of luminosity evolution that the average change in SN Ia progenitor age is “ ~ 5.3 Gyr” between $z = 0$ and $z = 1$. In spite of its precise mathematical derivation, measuring any galaxy physical property is inherently very difficult. Stellar mass is often quoted with an uncertainty of at least 0.3 dex (a factor of 2), and it is the simplest physical property to estimate. When considering stellar

Table 3. Hubble-Lemaître residuals and ages for 254 low redshift Pantheon SN Ia.

SN Ia	HR	uncertainty	Age
	[mag]	[mag]	[Gyr]
2001ah	−0.04	0.13	1.714
2001az	0.14	0.12	1.041
2001bf	−0.10	0.17	1.261
2001da	−0.04	0.14	1.261
2001eh	−0.01	0.16	1.924

NOTE— Hubble-Lemaître residuals are from [Jones et al. \(2018\)](#). Ages (light-weighted) are estimated using ZPEG. We used a fixed 15% uncertainty in this analysis. Table 3 is published in its entirety in the machine-readable format. A portion is shown here for guidance regarding its form and content.

ages, a similar limit in the accuracy of the measurements should be assumed. As discussed previously, Figure 3, it appears that the YEPS ages are under quoted on average by 1.6 ± 0.3 Gyr.

Secondly, modern SN Ia cosmology analyses ([Suzuki et al. 2012](#); [Betoule et al. 2014](#); [Rubin et al. 2015](#); [Scolnic et al. 2018](#); [DES Collaboration et al. 2019](#)), all of which have demonstrated strong evidence for accelerating expansion and dark energy, account for the well-established change in average Hubble-Lemaître residual across a division in host stellar mass. A procedure which reduces the effect from any correlation with age, due to galaxy scaling relationships. In addition, many analyses include a parameter to marginalize over the uncertainty that this change in Hubble-Lemaître residual could be caused by another host galaxy property, such as age. This marginalization would further reducing the effect of a trend with age. In [Rubin et al. \(2015\)](#), this marginalization was done with a the redshift dependent mass step. The $\delta(\infty)$ parameter did slightly prefer an age-like correlation, drastically reduces the maximum bias of any cosmological parameter due to the correlation reported by [K20](#).

To further investigate if standard cosmological analysis, which account for both host and selection effects, may mitigate the effect of [K20](#)’s trend on cosmological parameters, we replaced the Hubble-Lemaître residual with those calculated by the Pantheon analysis ([Scolnic et al. 2018](#)) for the 27 SN Ia where both were available. The associated Hubble-Lemaître residuals and ages are a subset of the full low redshift Parthenon SN Ia, listed

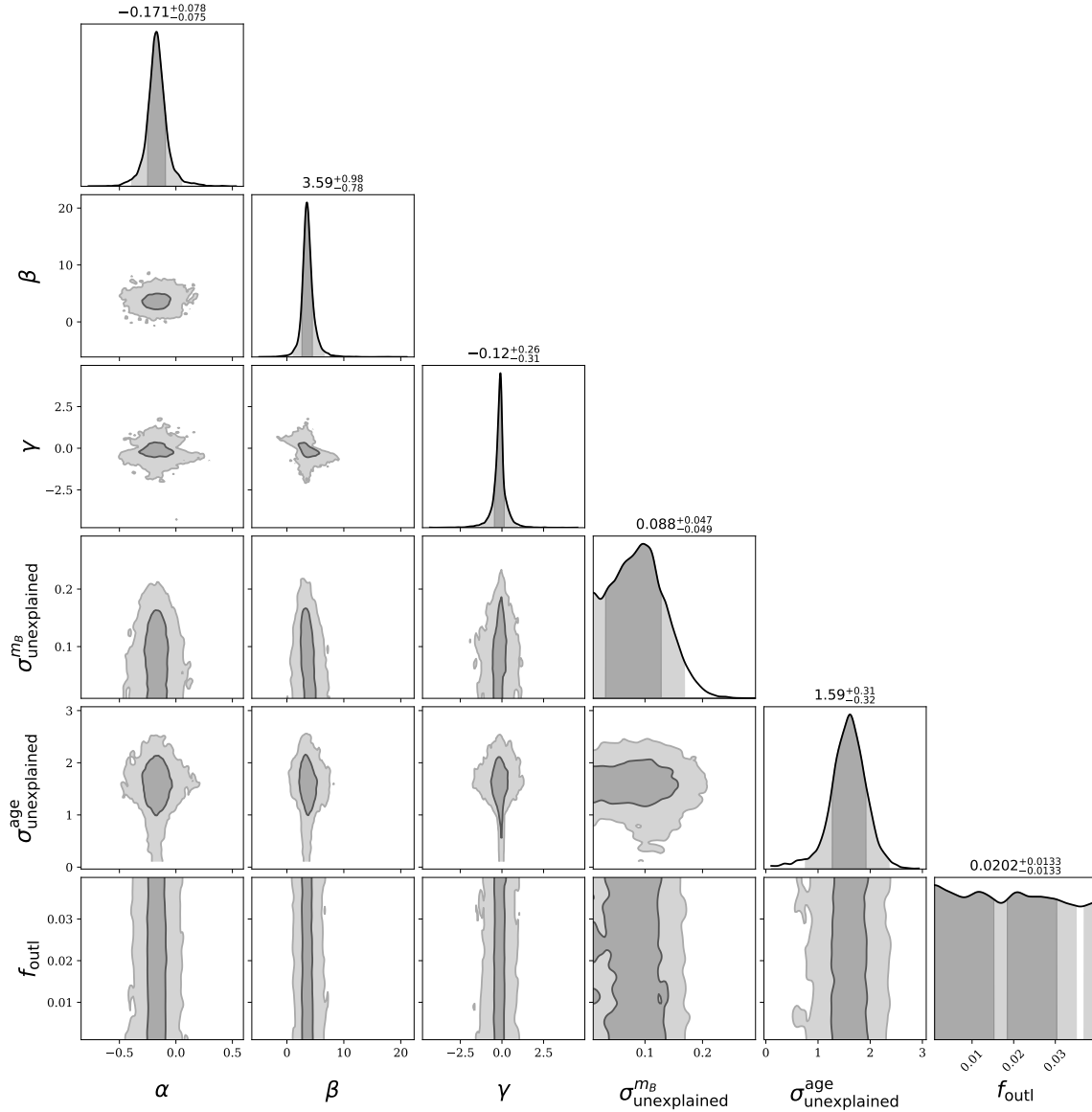


Figure 3. Results for the estimation of SN Ia standardizing parameters, same as Figure 2. A highly significant (3.8σ) non-zero additional age uncertainty ($\sigma_{\text{unexplained}}^{\text{age}}$) is seen. This parameter is added in quadrature to each reported YEPS age uncertainty. If $\sigma_{\text{unexplained}}^{\text{age}}$ is greater than zero, then the reported age uncertainties should be considered underestimated. However, on average, the uncertainties on the YEPS ages appear to be under reported by 1.6 ± 0.3 Gyr.

in Table 3. Using the standard cosmological corrections for the host mass step and observational biases (BBC Kessler & Scolnic 2017), the trend with Hubble-Lemaître residual becomes -0.016 ± 0.031 mag/Gyr or a non-zero significance of 0.5σ . Without SN2003ic, the trend with the Pantheon’s Hubble-Lemaître residuals reverses direction ($+0.008 \pm 0.030$ mag/Gyr). Thus we find no evidence for an age trend from the K20 ages using the distances calculated for cosmological analyses. This is also true for the two other age methods used in K20: going from K20 to Pantheon Hubble-Lemaître residuals decreases the significance of each correlation to $\sim 1.0\sigma$.

Our fifth conclusion is that using Hubble-Lemaître residuals that are fully standardized, including a host galaxy stellar mass term, the trend is small and insignificant and therefore does not propagate to a bias in cosmological estimates.

3. BEYOND THE K20 SAMPLE

K20 ultimately apply their trend to SNe Ia in young hosts (the most common hosts) and to SNe Ia at higher redshifts by assuming it can be extrapolated from old SN Ia to younger stellar populations. This interpretation assumes that the physical mechanism is a smoothly varying process rather than discrete like multiple sub-

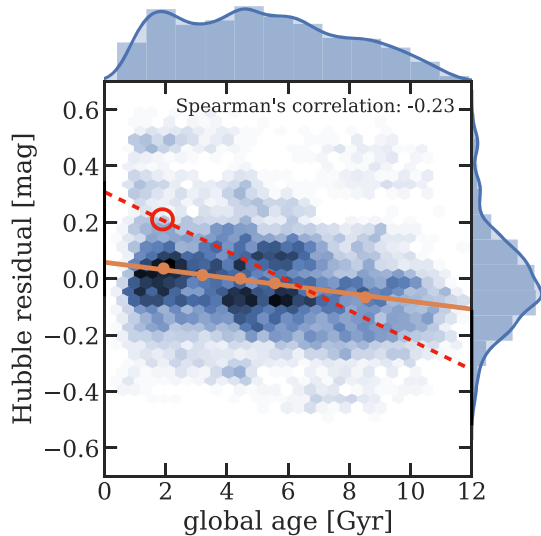


Figure 4. 2D density plot (darker colors indicate a higher density) depicting the probability of finding a SN Ia at a given Hubble-Lemaître residual and average stellar age. A linear fit to the data (orange line) is shown, along with six evenly filled bins (orange points). The extrapolated trend of K20 is shown as a red dashed line. The predicted average Hubble-Lemaître residual for young hosts is shown as the red circle. This prediction (red circle) is highly inconsistent with the measured average (orange point). The original figure is from Rose et al. (2019). We note that like K20, the Hubble-Lemaître residuals in Rose et al. (2019) do not include the mass step correction.

populations seen in Rigault et al. (2013) and Cikota et al. (2019). Indeed, it is quite possible that at all redshifts most SNe Ia are from young progenitors as SNe Ia in early-type hosts galaxies (typically dominated by old stars) make up only a small fraction of cosmological samples.

The interpretation in K20 implies that SN Ia in young hosts will have an average Hubble-Lemaître residual of ~ 0.25 mag. This highly biased average Hubble-Lemaître residual is also ruled out by the analyses of both Gupta et al. (2011) and Rose et al. (2019) who independently looked at data from the Sloan Digital Sky Survey (Sako et al. 2008; Campbell et al. 2013; Sako et al. 2018) using two distinct age estimators. An example of this discrepancy between external data and K20’s predication can be seen in Figure 4. Measurements of the Hubble-Lemaître residuals for SNe Ia from young host galaxies place the prediction of K20 well out in the tail of the distribution. Our sixth conclusion is that the linear extrapolation to young ages is highly inconsistent with external data.

The ages derived from the optical spectral energy distribution (SED) fitting of Rose et al. (2019) are not a

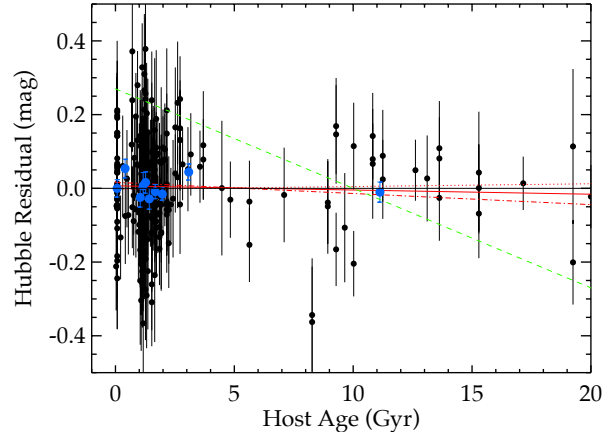


Figure 5. The relationship between the cosmological Pantheon sample’s Hubble-Lemaître residuals and host galaxy age (light-weighted) for the low redshift SNe Ia (black) of Jones et al. (2018). Blue points are bins of 25 SN Ia. The measured correlation is 0.0011 ± 0.0018 mag/Gyr (red line, dashed and dotted red are $\pm 1\sigma$). Light-weighted ages are typically biased by bright young stars, reducing the range of observed ages and increasing the measured slope. The green dashed line is the trend seen in K20. The mass step that is applied to the Pantheon Hubble-Lemaître residuals would not drastically shift the trend from K20 because it was derived using only massive early-type galaxies. However, one can easily imagine concluding a low-significance trend if one only used hosts with ages > 2.5 Gyr (the last two bins), as expected in a passive only sample. The trend seen in the early-type host galaxies does not seem to be present when looking at external cosmologically standardized data.

precise for any one individual host galaxy as the K20’s YEPS ages derived from spectral features. However, when aggregated, SED based ages are statistically powerful. Just like how photometric redshifts are more uncertain for any one object, in aggregate they can be a powerful tool, so are SED based ages.

For a more direct and empirical test of the size and sign of Hubble-Lemaître residuals of SNe Ia with young progenitors we analyzed the full sample of low redshift SNe Ia in the Pantheon sample with host galaxy properties derived by Jones et al. (2018) ($N=254$). We measured the correlation between Hubble-Lemaître residual and host age as in K20. This analysis used light-weighted ages derived from SED fitting via ZPEG (Le Borgne & Rocca-Volmerange 2002), as described in Jones et al. (2018). ZPEG uses 15 star formation histories, the Salpeter initial mass function (Salpeter 1955), 200 stellar age bins, 6 metallicity bins and marginalizes over $E(B - V)$, in order to fit the observed photometry. Figure 5 shows the expected result that the majority of low redshift SNe Ia are seen in young hosts. The Hubble-Lemaître residuals seen in these hosts are

strongly inconsistent with the +0.25 mag residual predicted by extrapolating the trend proposed by K20. Indeed, only a small number of all SNe Ia (at any age) show residuals of $\sim +0.25$ mag, contrary to the prediction that this is the average Hubble-Lemaître residual for SNe Ia in young hosts. Since Pantheon has both young spiral and old elliptical host galaxies, it is not necessary to extrapolate outside the range of the K20 data.

As seen in Figure 5, the age trend in the Pantheon sample is found to be 0.0011 ± 0.0018 mag/Gyr, consistent with no trend at the 0.6σ level. Light-weighted ages are typically biased younger by bright young stars, reducing the range of observed ages and increasing the measured slope. Not surprisingly, by excluding the mass step correction the size of a trend with age more than doubles due to the aforementioned correlation between host mass and age, though this trend is still only significant at the 1.4σ level, consistent with that seen by Rose et al. (2019) for the SDSS data. When using the same light-curve standardization parameters ($\alpha = 0.15$, $\beta = 3.69$) as K20, but including the mass step and BBC corrections, the correlation only has a 0.6σ significance. If we restrict ourselves to the oldest galaxies, as is the sample in K20, a very weak trend is found (1.2σ with only early-type hosts). No method of examining the Pantheon data set, was able to find a significant uncorrected trend with age. Our seventh conclusion is the combination of our fifth and sixth: when looking at a full cosmological data set, we see no significant trend and therefore we see no evidence for a significant and unaccounted for bias in the dark energy signal from SN Ia.

4. CONCLUSIONS

Kang et al. (2020) claim that an empirically-determined dependence of SN Ia host age and luminosity derived from a small sample of early-type host galaxies can be extrapolated to large samples and young ages to account for the majority of the dark energy signal. However, we find that this trend is not robust to reanalysis. The first issue is that 12% of the final sample, would not pass the standard JLA cosmological quality cuts, meaning that a large fraction of the data does not have reliable Hubble-Lemaître residuals.

The inclusion of standard error sources, clearly present in SN Ia residuals, reduces the significance of the dependence to $< 2\sigma$. Bypassing any need for formal uncertainty accounting, the Spearman rank-order coefficient only sees a correlation at 2σ . Further, the removal of the single SN Ia with the oldest host and a poorly sam-

pled light curve, SN2003ic, reduces the significance to 1.5σ . Finally, by doing a full re-fit and analysis that also marginalizes over the correlations in the standardization coefficients with a Bayesian hierarchical model, we find the trend falls to a 1.5σ and 0.6σ significance with and without SN2003ic respectively. The certainty and significance of this trend appear to not be robust.

If this correlation exists, the propagation to a bias in cosmological parameters is not direct or simple. Stellar ages are notoriously difficult to measure, and as such extra uncertainty should be added when propagating any host galaxy age correlation across cosmic time. For example, we find that the YEPS uncertainties appear to be under estimated, on average, by 1.6 ± 0.3 Gyr.

Finally, comparing the claimed trend against large, recent cosmological samples which include young hosts, the trend is strongly ruled out. We see that the standard practice of applying a host galaxy mass correction leaves only a very weak and insignificant relation between Hubble-Lemaître residuals and inferred age. Though there is evidence for a small host galaxy dependence of SN Ia peak luminosity, going back over 10 years, we find no evidence to question the more than 20 years of dark energy measurements.

The recent results of K20, upon re-examination, do not justify calling into question the presence of dark energy. However, we do concur with their closing remarks: the redshift dependence of SN Ia remains an important challenge for future precision dark energy measurements and requires ongoing studies.

ACKNOWLEDGMENTS

The authors would like to thank Peter Garnavich, Young-Wook Lee, and Young-Lo Kim for their time reading and commenting on an early draft. B.M.R., D.R., A.C., S.D., S.D., and A.F. acknowledge support from NASA through grant NNG16PJ311I. D.O.J. is supported by a Gordon and Betty Moore Foundation postdoctoral fellowship at the University of California, Santa Cruz. D.S. is supported by DOE grant de-sc0010007, the David and Lucile Packard Foundation, and in part by NASA through grant NNG17PX03C.

Software: corner.py (Foreman-Mackey 2016), kde.corner, Matplotlib (Hunter 2007), Numpy (van der Walt et al. 2011), Pandas (McKinney 2010), pystan (<https://doi.org/10.5281/zenodo.598257>), python, SciPy (Jones et al. 2001), stan (Carpenter et al. 2017), UNITY (Rubin et al. 2015), ZPEG (Le Borgne & Rocca-Volmerange 2002)

REFERENCES

- Betoule, M., Kessler, R., Guy, J., et al. 2014, *A&A*, 568, A22
- Burns, C. R., Stritzinger, M., Phillips, M. M., et al. 2011, *ApJ*, 141, 19
- Campbell, H., D’Andrea, C. B., Nichol, R. C., et al. 2013, *ApJ*, 763, 88
- Carpenter, B., Gelman, A., Hoffman, M. D., et al. 2017, *Journal of Statistical Software*, 76, 1
- Childress, M. J., Wolf, C., & Zahid, H. J. 2014, *MNRAS*, 445, 1898
- Chung, C., Yoon, S.-J., Lee, S.-Y., & Lee, Y.-W. 2013, *ApJS*, 204, 3
- Cikota, A., Patat, F., Wang, L., et al. 2019, *MNRAS*, 490, 578
- DES Collaboration, Abbott, T. M. C., Allam, S., et al. 2019, *ApJL*, 872, L30
- Foley, R. J., Filippenko, A. V., Aguilera, C., et al. 2008, *ApJ*, 684, 68
- Foley, R. J., Challis, P. J., Chornock, R., et al. 2013, *ApJ*, 767, 57
- Foreman-Mackey, D. 2016, *JOSS*, 24, doi:10.21105/joss.00024
- Freedman, W. L., Madore, B. F., Hatt, D., et al. 2019, *ApJ*, 882, 34
- Gallagher, J. S., Garnavich, P. M., Caldwell, N., et al. 2008, *ApJ*, 685, 752
- Garnavich, P. M., Kirshner, R. P., Challis, P., et al. 1998a, *ApJ*, 493, L53
- Garnavich, P. M., Jha, S., Challis, P., et al. 1998b, *ApJ*, 509, 74
- Garnavich, P. M., Bonanos, A. Z., Krisciunas, K., et al. 2004, *ApJ*, 613, 1120
- Gupta, R. R., D’Andrea, C. B., Sako, M., et al. 2011, *ApJ*, 740, 92
- Guy, J., Sullivan, M., Conley, A., et al. 2010, *A&A*, 523, A7
- Hamuy, M., Phillips, M. M., Maza, J., et al. 1995, *AJ*, 109, 1
- Hamuy, M., Trager, S. C., Pinto, P. A., et al. 2000, *AJ*, 120, 1479
- Hamuy, M., Phillips, M. M., Suntzeff, N. B., et al. 1996, *AJ*, 112, 2408
- Hicken, M., Wood-Vasey, W. M., Blondin, S., et al. 2009, *ApJ*, 700, 1097
- Hunter, J. D. 2007, *Computing in Science & Engineering*, 9, 90
- Jha, S., Riess, A. G., & Kirshner, R. P. 2007, *ApJ*, 659, 122
- Jones, D. O., Riess, A. G., & Scolnic, D. M. 2015, *ApJ*, 812, 31
- Jones, D. O., Riess, A. G., Scolnic, D. M., et al. 2018, *ApJ*, 867, 108
- Jones, D. O., Scolnic, D. M., Foley, R. J., et al. 2019, *ApJ*, 881, 19
- Jones, E., Oliphant, T., Peterson, P., et al. 2001, *arXiv:1907.10121*
- Kang, Y., Kim, Y.-L., Lim, D., Chung, C., & Lee, Y.-W. 2016, *ApJS*, 223
- Kang, Y., Lee, Y.-W., Kim, Y.-L., Chung, C., & Ree, C. H. 2020, *ApJ*, 889, 8
- Kelly, P. L., Hicken, M., Burke, D. L., Mandel, K. S., & Kirshner, R. P. 2010, *ApJ*, 715, 743
- Kessler, R., & Scolnic, D. 2017, *ApJ*, 836, 56
- Kim, Y.-L., Kang, Y., & Lee, Y.-W. 2019, *Journal of Korean Astronomical Society*, 52, 181
- Kim, Y.-L., Smith, M., Sullivan, M., & Lee, Y.-w. 2018, *ApJ*, 854, 24
- Kunz, M., Bassett, B. A., & Hlozek, R. A. 2007, *Phys. Rev. D*, 75, 103508
- Lampeitl, H., Smith, M., Nichol, R. C., et al. 2010, *ApJ*, 722, 566
- Le Borgne, D., & Rocca-Volmerange, B. 2002, *A&A*, 386, 446
- McKinney, W. 2010, *Data Structures for Statistical Computing in Python*
- Moreno-Raya, M. E., López-Sánchez, Á. R., Mollá, M., et al. 2016, *MNRAS*, 462, 1281
- Moshier, J., Guy, J., Kessler, R., et al. 2014, *ApJ*, 793, 16
- Perlmutter, S., Gabi, S., Goldhaber, G., et al. 1997, *ApJ*, 483, 565
- Perlmutter, S., Aldering, G., Goldhaber, G., et al. 1999, *ApJ*, 517, 565
- Phillips, M. M. 1993, *ApJL*, 413, L105
- Pskovskii, Y. P. 1977, *SvA*, 21, 675
- Rest, A., Scolnic, D., Foley, R. J., et al. 2014, *ApJ*, 795, 44
- Riess, A. G., Filippenko, A. V., Challis, P., et al. 1998, *ApJ*, 116, 1009
- Riess, A. G., Macri, L. M., Hoffmann, S. L., et al. 2016, *ApJ*, 826, 56
- Riess, A. G., Casertano, S., Yuan, W., et al. 2018, *ApJ*, 861, 126
- Rigault, M., Copin, Y., Aldering, G., et al. 2013, *A&A*, 560, A66
- Rigault, M., Brinnel, V., Aldering, G., et al. 2018, *arXiv:1806.03849*
- Rose, B. M., Garnavich, P. M., & Berg, M. A. 2019, *ApJ*, 874, 32
- Rose, B. M., Dixon, S., Rubin, D., et al. 2020, *ApJ*, 890, 60

- Rubin, D., Aldering, G., Barbary, K., et al. 2015, *ApJ*, 813, 137
- Sako, M., Bassett, B., Becker, A., et al. 2008, *AJ*, 135, 348
- Sako, M., Bassett, B., Becker, A. C., et al. 2018, *PASP*, 130, 064002
- Salpeter, E. E. 1955, *ApJ*, 121, 161
- Scolnic, D., & Kessler, R. 2016, *ApJL*, 822, L35
- Scolnic, D. M., Jones, D. O., Rest, A., et al. 2018, *ApJ*, 859, 101
- Smith, M., Sullivan, M., Wiseman, P., et al. 2020, *arXiv:2001.11294*
- Sullivan, M., Conley, A., Howell, D. A., et al. 2010, *MNRAS*, 406, 782
- Suzuki, N., Rubin, D., Lidman, C., et al. 2012, *ApJ*, 746, 85
- Tripp, R. 1998, *A&A*, 331, 815
- Uddin, S. A., Mould, J., Lidman, C., Ruhlmann-Kleider, V., & Zhang, B. R. 2017, *ApJ*, 848, 56
- van der Walt, S., Colbert, S. C., & Varoquaux, G. 2011, *CSE*, 13, 22