SUPERNOVA CONSTRAINTS AND SYSTEMATIC UNCERTAINTIES FROM THE FIRST 3 YEARS OF THE SUPERNOVA LEGACY SURVEY[1]

A. Conley[2,3], J. Guy[4], M. Sullivan[5], N. Regnault[4], P. Astier[4], C. Balland[4,6], S. Basa[7], R. G. Carlberg[2], D. Fouchez[8], D. Hardin[4], I. M. Hook[5,9], D. A. Howell[10,11], R. Pain[4], N. Palanque-Delabrouille[12], K. M. Perrett[2,13], C. J. Pritchet[14], J. Rich[12], V. Ruhlmann-Kleider[12], D. Balam[14], S. Baumont[15], R. S. Ellis[5,16], S. Fabbro[14,17], H. K. Fakhouri[18], N. Fourmanoit[4], S. González-Gaitán[2], M. L. Graham[14], M. J. Hudson[19], E. Hsiao[18], T. Kronborg[4], C. Lidman[20], A. M. Mourao[17], J. D. Neill[21], S. Perlmutter[18,22], P. Ripoche[18,4], N. Suzuki[18], E. S. Walker[5,23]

[1] Based on observations obtained with MegaPrime/MegaCam, a joint project of CFHT and CEA/DAPNIA, at the Canada-France-Hawaii Telescope (CFHT) which is operated by the National Research Council (NRC) of Canada, the Institut National des Sciences de l'Univers of the Centre National de la Recherche Scientifique (CNRS) of France, and the University of Hawaii. This work is based in part on data products produced at the Canadian Astronomy Data Centre as part of the Canada-France-Hawaii Telescope Legacy Survey, a collaborative project of NRC and CNRS.

[2]Department of Astronomy and Astrophysics, University of Toronto, 50 St. George Street, Toronto, ON M5S 3H4, Canada

[3]Center for Astrophysics and Space Astronomy, University of Colorado, 593 UCB, Boulder, CO 80309-0593, USA

[4]LPNHE, Université Pierre et Marie Curie Paris 6, Université Paris Diderot Paris 7, CNRS-IN2P3, 4 place Jussieu, 75252 Paris Cedex 05, France

[5]Department of Astrophysics, University of Oxford, Keble Road, Oxford OX1 3RH, UK

[6]Université Paris 1, Orsay, F-91405, France

[7]LAM, CNRS, BP8, Pôle de l'Étoile Site de Château-Gombert, 38, rue Frédéric Joliot-Curie, 13388 Marseille Cedex 13, France

[8]CPPM, CNRS-IN2P3 and Université Aix-Marseille II, Case 907, 13288 Marseille Cedex 9, France

[9]INAF - Osservatorio Astronomico di Roma, via Frascati 33, 00040 Monteporzio (RM), Italy

[10]Las Cumbres Observatory Global Telescope Network, 6740 Cortona Dr., Suite 102, Goleta, CA 93117, USA

[11]Department of Physics, University of California, Santa Barbra, Broida Hall, Mail Code 9530, Santa Barbara, CA 93106-9530, USA

[12]CEA, Centre de Saclay, Irfu/SPP, F-91191 Gif-sur-Yvette, France

[13]Network Information Operations, DRDC Ottawa, 3701 Carling Avenue, Ottawa, ON, K1A 0Z4, Canada

[14]Department of Physics and Astronomy, University of Victoria, PO Box 3055 STN CSC, Victoria BC, V8T 1M8, Canada

[15]LPSC, CNRS-IN2P3, 53 rue des Martyrs, 38026 Grenoble Cedex, France

[16]Department of Astrophysics, California Institute of Technology, MS 105-24, Pasadena, CA 91125, USA

[17]CENTRA-Centro M. de Astrofisica and Department of Physics, IST, Lisbon, Portugal

[18]LBNL, 1 Cyclotron Rd, Berkeley, CA 91125

[19]Department of Physics and Astronomy, University of Waterloo, 200 University Avenue West, Waterloo, Ontario N2L 3G1, Canada

[20]Australian Astronomical Observatory, P.O. Box 296, Epping, NSW 1710, Australia

[21]California Institute of Technology, 1200 E. California Blvd., Pasadena, CA 91125, USA

alexander.conley@colorado.edu

ABSTRACT

We combine high redshift Type Ia supernovae from the first 3 years of the Supernova Legacy Survey (SNLS) with other supernova (SN) samples, primarily at lower redshifts, to form a high-quality joint sample of 472 SNe (123 low-z, 93 SDSS, 242 SNLS, and 14 *Hubble Space Telescope*). SN data alone require cosmic acceleration at $> 99.999\%$ confidence, including systematic effects. For the dark energy equation of state parameter (assumed constant out to at least $z = 1.4$) in a flat universe, we find $w = -0.91^{+0.16}_{-0.20}\,(\mathrm{stat})^{+0.07}_{-0.14}\,(\mathrm{sys})$ from SNe only, consistent with a cosmological constant. Our fits include a correction for the recently discovered relationship between host-galaxy mass and SN absolute brightness. We pay particular attention to systematic uncertainties, characterizing them using a systematics covariance matrix that incorporates the redshift dependence of these effects, as well as the shape-luminosity and color-luminosity relationships. Unlike previous work, we include the effects of systematic terms on the empirical light-curve models. The total systematic uncertainty is dominated by calibration terms. We describe how the systematic uncertainties can be reduced with soon to be available improved nearby and intermediate-redshift samples, particularly those calibrated onto USNO/SDSS-like systems.

Subject headings: cosmological parameters — cosmology: observations — supernovae: general

1. INTRODUCTION

The Supernova Legacy Survey (SNLS) is a 5 year program to measure the expansion history of the universe using Type Ia supernovae (SNe Ia). The goal of this survey is to measure the time averaged equation of state of dark energy w to 0.05 (statistical uncertainties only) in combination with other measurements, and to 0.10 including systematic effects. The fundamental nature of dark energy, which makes up 3/4 of the mass-energy budget of the

---

[22]Department of Physics, University of California, Berkeley, 366 LeConte Hall MC 7300, Berkeley, CA 94720-7300, USA

[23]Scuola Normale Superiore, Piazza dei Cavalieri 7, 56126 Pisa, Italy

universe, remains almost completely mysterious. A solid measurement that $w \neq -1$ (which would rule out the cosmological constant) would have profound implications for cosmology and particle physics. SNLS completed data acquisition in 2008 June; this paper presents SN-only cosmological results from the first 3 years of operation (SNLS3).

All analysis in this paper is in the context of standard cosmological models – i.e., we assume that the universe is homogenous on large scales and that general relativity is correct. SNe are used to measure the cosmological parameters by comparing their apparent brightnesses over a range of redshifts. Hence, it is very useful to include additional SN samples besides SNLS in the analysis, particularly nearby SNe ($z < 0.1$). This paper has two primary goals. The first is to place SNe from the literature on a common framework and demonstrate the resulting constraints. The second is to present the systematic uncertainties on the SN measurements in detail. This is the second in a series of three SNLS cosmology papers based on the first 3 years of data. The SNLS data sample used in this analysis is presented in Guy et al. (2010, hereafter G10), and the SN constraints are combined with other measurements in Sullivan et al. (2011b, hereafter S11), primarily the WMAP7 measurement of the cosmic microwave background (CMB), and baryon acoustic oscillations measured with galaxy redshift surveys. In addition to these three papers, the calibration of SNLS data is discussed in Regnault, Astier, Guy, et al. (2009, hereafter R09), and the simulations used to evaluate our survey selection effects in Perrett et al. (2010).

Searches for additional parameters beyond light-curve shape and color have been ongoing for two decades, and recently, Kelly et al. (2010) have found evidence that SN residuals from the Hubble diagram are correlated with host-galaxy mass. The physical cause of this effect is as yet unknown. Sullivan et al. (2010a) analyze this issue using the SNLS3 sample and confirm this result at higher significance. They find that splitting the sample by host-galaxy mass and allowing the peak absolute magnitudes of the two samples to differ corrects for these effects. We adopt this approach in this paper; see §3.2 for details. This effect has also been confirmed in the Sloan Digital Sky Survey (SDSS) SN sample (Lampeitl et al. 2010).

Here, we give a brief overview of the data sets used in this analysis. SNLS combines photometry from the deep component of the Canada-France-Hawaii Telescope (CFHT) Legacy survey with extensive spectroscopic follow-up from the Keck, Very Large Telescope (VLT), and Gemini telescopes to determine SN types and measure redshifts. The photometry is carried out with MegaCam, a 1 deg^2 imager at the prime focus of CFHT. SNLS is a rolling search, like most other modern high-redshift surveys, which means that the same telescope is used to simultaneously find new SNe candidates and to follow those already discovered, resulting in a large multiplex efficiency advantage. Details of the procedures used to find and prioritize new SNe can be found in Sullivan et al. (2006a), and those used to type candidates

from their spectra in Howell et al. (2005); Balland et al. (2009).

In addition to SNLS, we consider three additional data sets: low-z, SDSS, and high-redshift SNe from the *Hubble Space Telescope* (*HST*). The low-z SNe, which we define here as coming from surveys with the bulk of their SN below $z = 0.1$, come from a heterogeneous combination of non-rolling surveys which generally use different telescopes to find and follow SN candidates. The *HST* SNe (Riess et al. 2007) are at higher redshifts than any of the other samples. They contribute relatively little to measurements of $\langle w \rangle$, but are quite useful when trying to measure any redshift evolution of w. The SDSS SN survey (Holtzman et al. 2008; Kessler et al. 2009, hereafter K09) occupies intermediate redshifts ($0.1 < z < 0.4$) between SNLS and the low-z SN, and is also a rolling search.

As SN samples have grown in size, characterizing and incorporating systematic uncertainties properly has grown in importance. There are many aspects of SNLS which are designed to reduce the effects of systematic uncertainties compared with previous SN projects, but they are still roughly comparable to the statistical uncertainties. A major lesson of this paper is that most of the systematic effects which limit the current analysis are related to the current low-redshift SN sample, particularly in terms of cross-calibration requirements. Absolute calibration is unimportant for our purposes, but the relative calibration of different systems and observations at different wavelengths is critical. The current low-z sample is dominated by SNe largely calibrated to the Landolt system (Landolt 1992), which induces many complications in our calibration (see R09 for more details). As the low-z SN sample is replaced by better calibrated samples in the next few years, it will be possible to cross-calibrate the various samples much more accurately, which will substantially increase the legacy value of the SNLS sample. The dominant uncertainties will probably then relate to the host-galaxy–SN brightness relation and SN modeling, particularly the thorny issue of SN colors.

A critical step in any SN cosmological analysis is light-curve fitting, the conversion of a time series of photometric (and possibly spectroscopic) observations into a set of model parameters for each SN which are used to estimate a relative distance. Because SN physics is sufficiently complicated, theoretical models have so far offered relatively little guidance for this process. As a result, all current models used for light-curve fitting are empirical in nature. In this paper we consider updated versions of two models: SiFTO (Conley et al. 2008) and SALT2 (Guy et al. 2007). While these models share a common overall philosophy, there are many significant differences between them. They are compared in detail in §5 of G10, and we include the differences in the light-curve parameters in our uncertainty budget. Because these models are trained on SN data, they are affected by the same sources of systematic effects as the cosmological analysis. This is the first analysis to include this

effect in the analysis; previous analyses have therefore underestimated their uncertainties by holding the light-curve model fixed when modeling systematics.

There has been some variation in the literature as to how SN systematic effects are treated. Examples of the most common approach, which we will refer to as the quadrature method, can be found in Perlmutter et al. (1999); Astier et al. (2006); Wood-Vasey et al. (2007); K09. This approach has a number of disadvantages (and the advantage that it is relatively easy to understand). The most important disadvantage is that it is difficult for subsequent consumers of SN relative distance moduli to incorporate systematic uncertainties into their analyses. In this paper we rectify these deficiencies by modeling systematic effects using a covariance matrix.

In §2, we describe the data sets included in this analysis and the steps taken to bring them onto a common system with the SNLS data. We then describe the combined data set in §3, presenting the statistical constraints on dark energy from SNe Ia alone, and the combined statistical and systematic uncertainties in §4 along with a description of our systematics methodology. In section §5 we describe individual systematic terms in detail, and in §6 we compare our analysis with previous ones. Finally, in §7 we discuss ways in which the effects discussed in this paper can be improved with the enhanced low- and intermediate-redshift SN data sets which will be available in the near future.

2. DATA SETS

Ideally, all SNe would be observed with a single camera on a single telescope. While this may be possible with future programs such as LSST or Pan-STARRS, at the moment it is not practical because no individual telescope can currently obtain a large sample at both high and low redshifts. To obtain precision cosmological constraints, particularly on w, we must incorporate additional sSN data besides SNLS. Rather than including all available SN data, we have chosen to incorporate only external surveys which cover different redshift ranges than SNLS. Other surveys which cover substantially the same redshift range as SNLS have fewer SN, larger photometric uncertainties, less certain calibration, and have at most 2-band coverage (compared with the 4 bands of SNLS). Including them in our analysis would reduce our statistical uncertainties only marginally, and would introduce additional systematic effects which would have to be analyzed in detail. Nearby SNe are currently the most important addition because the combination of low redshifts and the small redshift

range means that they provide a very good constraint on the absolute magnitudes of SNe Ia[1], which is largely independent of the density parameters and w.

There has been a recent encouraging trend toward providing photometry in the natural system of the detectors used to obtain it – that is, rather than applying linear relations to transform instrumental magnitudes into some standard system, the calibration is transferred from the standard system to the instrumental system, which is much more reliable. Using natural system photometry involves slightly more work in the analysis, but allows for a considerable improvement in accuracy and precision, as long as the natural system response is well measured. We use the natural systems of the SNLS, *HST*, SDSS, Hicken et al. (2009a), and Contreras et al. (2010) samples in our analysis, which together account for $\sim 90\%$ of our sample.

In this section, we present the data samples we include in our analysis, and describe the steps we carry out to bring them to a common system with the SNLS measurements. The most critical items relate to calibration, but in addition it is important to understand and apply a correction for the selection biases of the external samples (i.e., Malmquist bias). The calibration used in this analysis relies heavily on the CALSPEC program based on *HST* observations of pure-hydrogen white dwarfs (Bohlin 1996). In particular, we use BD 17° 4708 as our primary reference standard. We first describe our cuts (§2.1), introduce the data samples (SNLS, nearby, *HST*, and SDSS, §2.2-2.5), and then discuss how we estimate selection effects for these samples (§2.7) and the peculiar velocity corrections we apply to nearby SNe (§2.8). All of these have related systematic uncertainties, which are discussed in §5.

2.1. Selection Requirements

The selection requirements (cuts) applied in our analysis have already been described in §4.5 of G10; these are designed to ensure adequate phase and wavelength coverage to allow accurate parameter measurement. A particularly important requirement is that each SN has data between -8 and $+5$ rest-frame days of peak brightness to avoid biasing the recovered light-curve parameters, which primarily affects the low-z sample – see G10 for further discussion.

As is the case for SNLS, we require spectroscopic confirmation of all SNe. Since each SN

---

[1]More precisely, they constrain a combination of the absolute magnitudes and the Hubble constant, H_0; we refer to this combination as $\mathcal{M}$.

team has its own scheme for spectroscopic classification, it is not clear how to best ensure uniform classification standards without re-examining all of the raw spectra. Based on direct comparison of spectra, the classification requirement applied to SNLS (SN Ia⋆ Balland et al. (2009), or CI=3 in the scheme of Howell et al. (2005)) lies somewhere between the "gold" and "silver" classification scheme of Riess et al. (2007) for spectroscopically typed SN. We have chosen to largely accept the classifications of the original authors with a few caveats. First, as is the case for all SNLS SNe, we do not include candidates whose type is based either purely on the photometric properties of the SN, or the red color or elliptical morphology of its host galaxy; the latter standard is particularly worrisome because some nearby ellipticals show evidence of star formation, and this fraction may increase with redshift. Recently, Kawabata et al. (2010) have reported the discovery of an (unusual) core-collapse SN in an elliptical host, which strengthens our caution.

Because we apply a peculiar velocity correction, we place our minimum redshift cut at $z_{cut} = 0.010$, somewhat lower than the usual value of $0.015 - 0.020$. There has been some recent controversy relating to the minimum allowable redshift, related to the potential for a discontinuous step in the local expansion rate (a so-called Hubble bubble) detected by Jha, Riess, & Kirshner (2007). Riess et al. (2007) use $z_{cut} = 0.023$ in order to avoid this issue, which removes 40% of the nearby sample. Conley et al. (2007) argue that the Hubble bubble is an artifact of the treatment of SNe colors combined with selection effects in the nearby sample, and Hicken et al. (2009b) shows that adding more nearby SNe shifts the position and sign of the putative bubble considerably. K09 take a slightly more agnostic position (§9.1 and 9.2), but find significant variation in the cosmological parameters with the minimum redshift cut, and therefore adopt $z_{cut} = 0.02$ while including the variation as a major source of systematic uncertainty. We also see similar variation with z_{cut} in our sample, but find it to be consistent with shot noise, and see no evidence that such conservatism is warranted (§5.3). SNe with $z < z_{cut}$ are still used to train the light-curve fitters, since we do not use distance information in this process.

We exclude known peculiar SNe by hand, such as SNe 2000cx and 2002cx, rather than using automated quality of fit (χ^2) cuts. In our experience, the uncertainties for low-z photometry are sufficiently inaccurate that such cuts are often misleading. Furthermore, many low-z SNe have the occasional outlying photometric observation which has little to no effect on the derived parameters but which drives the χ^2 of the light-curve fit to large values. A χ^2 based cut will also affect different SN samples very asymmetrically because the signal-to-noise of the photometry varies strongly between samples, and hence may introduce bias. Since SNe Ia are not perfect standard candles, we should include some additional scatter in their corrected peak magnitudes for cosmological purposes. This is often called "intrinsic scatter" (σ_{int}) even though it probably represents the gaps in our understanding

of SN physics and photometry rather than anything truly intrinsic. It is impossible for any fully automated set of cuts to catch all the pathologies of such a large and diverse collection of SNe, so we have also inspected all of the SNe by eye to remove problems not caught by our cuts. In any case, we remove the known spectroscopically peculiar SNe 2000cx (Li et al. 2001), 2001ay (Krisciunas et al., 2011, in preparation), 2002cx (Li et al. 2003), 2005hk (Phillips et al. 2007), 2005gj (Aldering et al. 2006), 03D3bb (Howell et al. 2006), 05D1by (which is spectroscopically similar to SN 2001ay), and SN 2006X because it shows evidence for a light-echo (Wang et al. 2008). Furthermore, we require all SNe to have estimated photometric uncertainties.

We apply Chauvenet's criterion (Taylor 1997) to reject outliers, removing SNe for which we could expect less than half of an event in our full sample (assuming a Gaussian distribution of intrinsic luminosities). This corresponds to a cut at about 3.2σ, which removes six SNe: one low-z (SN 2006cj), one from SDSS (SDSS5635), three from SNLS (03D4au, 04D4gz, 05D2ei), and one from the *HST* sample (McEnroe), many more than the 0.5 objects one would expect if the distribution were truly Gaussian. Of the SNLS SNe, two of these have slightly less secure spectroscopic identifications (SNIa⋆) and are faint relative to the best fit, so may either be non-SN Ia or spectroscopically peculiar SN Ia that our spectra were not high enough signal-to-noise to identify. Note that lensing effects are too small to explain these outliers, as the expected number of 3.2σ or greater outliers due to lensing is much less than 1 for our sample.

We require the Galactic reddening along the line of sight to satisfy $E(B-V)_{MW} < 0.2$ mag because of concerns that the assumed Galactic value of $R_V = 3.1$ might not be appropriate for highly extinguished objects. Next, we require $0.7 < s < 1.3$, where s (stretch) is the light-curve width parameter for SiFTO; neither SiFTO nor SALT2 produce reliable fits for SNe outside that stretch range. Finally, we require $-0.25 < (B-V)_{\mathrm{Bmax}} < 0.25$ because of concerns that very blue SNe are not represented in our training sample, and that the colors of very red SNe may represent a combination of different effects, both affecting the peak magnitudes of the SNe; see Conley et al. (2008) for further discussion of the latter. Histograms for the redshift and color distributions of the input samples are shown in figure 1, and the effects of the cuts are given in table 1. The requirement of good coverage near the peak luminosity has the largest effect, mostly on the low-z sample. We also considered the effects of the sharp color cut because of concerns that a high-redshift SN with poorly measured colors might migrate across the cut boundary, but found that the effect was negligible.

Table 1.   Effects of Cuts on Different Samples

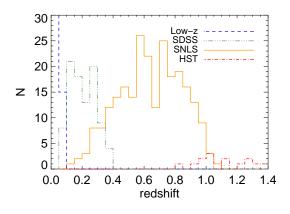| Dataset	Initial[a]	Final[b]	Coverage[c]	Fail[d]	z Cut	s	Color	Outliers
low-z	323	123	99	9	54	32	31	1
SDSS	101	93	6	0	0	4	1	1
SNLS	279	242	16	10	0	6	9	3
HST	35	14	5	12	11	3	0	1

Note. — The number of SN removed by each selection criterion. Many SNe fail multiple cuts.

[a]The initial number of SN, after the removal of known peculiar SNe, SN with clear photometric inconsistencies, and SN which have better photometry from other samples.

[b]The number of SN satisfying all selection criterion.

[c]SNe which did not have data close enough to peak brightness.

[d]SNe for which the fits could not provide reliable light-curve parameters, either because of an insufficient number of epochs of data or insufficient wavelength coverage to measure a color.
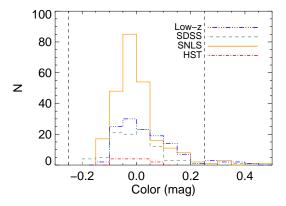

Fig. 1.— Redshift and color histograms of the samples used in this analysis. The left panel shows the redshift histogram after all cuts are applied. The first bin for the low-z SNe contains 108 SNe. The right panel shows the color histogram before the color cut is applied (but removing known peculiar SNe as well as those with insufficient coverage), and the cut is indicated by the vertical dashed lines.

2.2. SNLS SNe

Systematics control is fundamental to the design of SNLS.

- Because SNe are both discovered and photometrically followed with only one telescope, we are able to concentrate our efforts on thoroughly understanding that system (R09). We can also avoid the difficulties associated with combining observations from many telescopes onto a single system.

- The survey is in four passbands ($g_M r_M i_M z_M$), allowing us to measure colors for all SNe in our survey. We measure different rest-frame passbands at different redshifts; at low redshifts we are mostly sensitive to $B - V$, and at high redshifts to $U - B$. These measurements must be used in a consistent fashion at all redshifts to obtain accurate results. We use intermediate-redshift SNLS SNe as a consistency check of this process, since they have high-quality measurements of both $U - B$ and $B - V$ (G10).

- There are four survey fields (D1-D4) distributed in RA to allow year-round coverage. The results from the four fields can be compared with each other (S11).

- The survey is quite deep, which limits the effects of Malmquist bias Perrett et al. (2010) at $z < 0.6$, the sweet spot for measuring a constant w.

- We obtain spectra for all of the SNe used in our analysis, which allows us to limit non-Ia contamination, to search for peculiar SNe such as 03D3bb (Howell et al. 2006), and the comparison of the spectral energy distributions (SEDs) of nearby and distant SNe as a test for evolution (Bronder et al. 2008; Balland et al. 2009; Walker, Hook, Sullivan, et al. 2010).

- We have obtained much higher signal-to-noise spectra of a subset of our SNe which can be used to study the near-UV properties of SNe Ia in detail, and allows more detailed spectroscopic comparisons (Ellis et al. 2008; Sullivan et al. 2009).

- Because SNLS is a rolling search, it is possible to go back after an SN is discovered and study early-time, pre-discovery photometry. Therefore almost all of our SNe have very good early-time coverage, which can be used to test evolutionary models which predict changes in the early-time behavior (Conley et al. 2006b).

- Because the duration of our survey is much longer than the month timescale of SNe Ia, we can construct very deep SN-free image stacks and get accurate colors for the host galaxies. These can be turned into estimates of the mass and star formation history and used to study the relation between SN properties and host galaxy environment to search for host-dependent systematic effects (Sullivan et al. 2006b, 2010a).

- Our photometry allows us to construct improved empirical models of SNe Ia lightcurves and spectra (Guy et al. 2007; Conley et al. 2008), which is particularly important in the near-UV.

SNLS is designed to obtain sufficient quality data to allow us to investigate systematic effects within our own data, as well as enabling a great deal of non-cosmological SN science too extensive to list here.

We take the 3 year SNLS sample from G10 with only one modification: we have updated the light-curve models to account for the 2010 February update to the CALSPEC calibration library. This update has almost no effect on the SNLS, SDSS, and low-z SNe, but does have some on the *HST* SNe, which are observed in the near-IR.

2.3. low-z SNe

The current nearby SN set is dominated by five main samples: Calán/Tololo (Hamuy et al. 1996, 29 SNe), CfAI (Riess et al. 1999, 22 SNe), CfAII (Jha et al. 2006, 44 SNe), CfAIII (Hicken et al. 2009a, 185 SNe), and CSP (Contreras et al. 2010, 35 SNe). In addition, we make some use of 37 SNe with modern photometry from a mixture of other papers. As the numbers above show, the CfAIII sample provides almost half of the nearby sample, and it was tempting to include only these SNe to simplify our analysis. However, the other samples contribute to the training of the light-curve models disproportionately because some have denser phase coverage or sample slightly different wavelengths, and since it is necessary to characterize their properties and systematic uncertainties to include them in the training, there is no point in excluding them from the cosmological analysis. The effects of the cuts of §2.1 are broken down by each nearby sample in table 2. We do not use rest-frame observations in the U-band of the nearby sample in our analysis for the reasons discussed in §2.6.

A handful of the nearby SNe have photometry from multiple sources. This is useful for estimating the uncertainty in the zero points of different samples (§5.1.2), but also forces us to make some decisions about which data to include. Generally, nearby SNe have dense enough light-curve coverage that adding additional points does not improve the uncertainties significantly. Since σ_{int} dominates the uncertainty budget for these SNe[2], including data from multiple sources is not beneficial unless the data samples complement each other in wavelength or light-curve phase. Unless this is the case, or one of the data sets is clearly

---

[2]SN Ia are not perfect standard candles even after correction for the empirical width- and color-luminosity relations. σ_{int} represents the remaining scatter in distance moduli, and is discussed further in §3.1 and 3.4.

superior, we generally prefer to only use data from one of the five large surveys, where we have a better understanding of the systematic uncertainties.

For the Calán/Tololo, CfAI and CfAII samples, the data were transformed by the authors from the natural instrumental system into the the Landolt (1992) system using linear transformations derived from stars in a limited color range. Since SN and stellar spectra are quite different, these linear transformations are incorrect when applied to SNe, and will introduce an error in the reported magnitudes which is correlated across the SN sample. In order to use these samples in a precision analysis, we must determine the effective passbands of the Landolt system, described in Appendix 9.

For the CfAIII and CSP samples, natural system photometry is available, which we use in our analysis. To use the natural systems, we need the magnitudes of our fundamental flux standard in their natural system. For the CfAIII sample we use the linear transformations given in table 2 of Hicken et al. (2009a), after verifying that the arbitrary additive constants are identically zero. Fortunately, our fundamental flux standard (BD 17° 4708) lies in the color range well measured by these transforms so they should be reliable, which would not necessarily be the case for very blue stars such as Vega or the white dwarfs used in the CALSPEC program. For the CSP sample, the magnitudes of BD 17° 4708 are already given in the natural system.

2.4. HST SNe

SNe above $z = 1$ are difficult to observe from the ground. Therefore, the most successful searches in this redshift range have been carried out with HST. SNe at such high redshifts are useful when studying any possible time variation of w, so we include the 2 SNe from Blakeslee et al. (2003), 16 from Riess et al. (2004), and the 22 from Riess et al. (2007) in our sample; the latter also re-reduces the photometry from the previous papers, taking into account the nonlinearity of the NICMOS camera. The HST sample extends from $z = 0.2 - 1.55$. For $z < 0.7$ SNe, rest-frame B is measured by the F606W filter, which has a sufficiently broad response that the U-band is effectively included as well. This is unlike any of the other SNe in our sample, and therefore may introduce additional systematic effects in our light-curve fitting. Because these SNe have virtually no impact on the cosmological parameters when compared with the hundreds of SNe from other samples in this redshift range, we exclude the 10 HST SNe with $z < 0.7$. Requiring spectroscopic type confirmation eliminates five SNe, but ensures a consistent analysis. We also exclude SNe with $z > 1.4$ because we cannot estimate the Malmquist bias above this redshift (§2.7), which eliminates one more SN.

One cut which should be revisited for the *HST* sample is the requirement of having photometry close to the epoch of peak flux. In §4.5 of G10, we quantitatively studied the effects on the recovered fit parameters of not having photometry before a certain epoch, and concluded that an observation near peak between rest-frame epochs -8 and $+5$ was necessary to avoid bias. Unlike any of the other samples we consider, in many cases the *HST* sample has data earlier than rest-frame day -8 but not near peak, because it is the only (psuedo-)classical high-z search we have included (i.e., it is not a rolling search, and therefore there may be a large gap between the reference and search image). We did not consider this possibility in our original simulations. On the other hand, the signal-to-noise of the *HST* photometry is much lower than any of the other samples and the cadence less frequent, which may increase the bias. Therefore, we have revisited this study, this time including a single early-time data point between -20 and -15 with a representative signal-to-noise in F850LP (the filter that the *HST* searches were carried out in). Since it is imperative to use real data in this test to reflect the diversity of light-curve shapes, to simulate the lower signal-to-noise of the *HST* data we use SNLS SNe above $z = 0.65$ and further degrade their uncertainties to match the *HST* data. As for the other samples, only $m_B{}^{3}$ shows significant bias, which is shown in figure 2. Comparing with the results shown in G10, the extra early epoch significantly reduces the bias. Therefore, for *HST* SNe we require either an early observation between -20 and -15 and at least one between -8 and $+9$, or, with no early time point, an observation between -8 and $+5$ (all values rest frame epochs). Altogether, we are left with a sample of 14 *HST* SNe from $z = 0.7 - 1.4$.

The *HST* data set considered here consists of observations obtained with ACS in wide-field mode and camera 2 of the NICMOS instrument. The calibration of the NICMOS data is defined by the solar analog P330e (A .G. Riess, 2009, private communication) having the magnitudes 11.91 and 11.45 in the F110W and F160W filters, respectively. We do not have reliable Landolt system magnitudes for P330e, so it is not suitable as a primary standard. Fortunately it has a CALSPEC SED on the same system as our primary flux standard, so we can use spectrophotometry to align these observations with our calibration. Contrary to convention in the near-IR, this is *not* consistent with Vega having zero magnitude and zero color in these passbands, but instead implies a magnitude of about -0.057 in both. It is not clear if other analyses have taken this into account (K09, for example, do not, while Amanullah et al. (2010) do), but this does not imply that the Riess et al. (2007) analysis is in error. We have also updated the ACS bandpasses and zero points to match Bohlin (2007) and again for the 2009 January and 2010 February CALSPEC updates. Together,

---

[3]m_B is the peak magnitude in the rest-frame B band and is one of the critical parameters we use to estimate relative distances; see § 3.

these imply $2 - 3\%$ zero point changes for ACS[4]. Before these corrections are applied, the *HST* SNe have unusually red colors and are significant outliers in the Hubble diagram.

2.5. SDSS SNe

The SDSS Supernova Survey has recently released light curves for 146 intermediate redshift ($z < 0.4$) SNe observed in their first year of operation (Holtzman et al. 2008). Perhaps the most important aspect of this sample is that the calibration is better understood than will likely ever be possible with the bulk of the current nearby SN sample. Furthermore, because the SDSS filter system is similar to the MegaCam system, many inter-calibration issues with SNLS can be reduced significantly. We have not yet fully carried out this program, but expect to do so in a future collaboration with the SDSS and CSP SN teams. This paper presents the results of a preliminary combination of these data sets with our own which should be at least as good as any previous combinations of different data sets, but which does not take full advantage of the calibration improvements possible with these three samples.

A visual inspection of the SDSS light curves reveals severe systemic problems in the u band, and lesser ones in z. We therefore follow the advice of K09 and only include gri in our fits. We follow Holtzman et al. (2008) in only including points with data quality flag < 1024. Of the 146 SNe present in the full sample, our data quality cuts combined with removing obviously peculiar SNe remove 45 SNe.

The SDSS SN relative calibration is based on the catalog of Ivezić et al. (2007), and tied to an absolute calibration (AB) via observations of three solar-analog stars with high-quality CALSPEC SEDs. Unfortunately, these stars were too bright to observe directly with the SDSS imager. Instead, they were calibrated by transferring observations from the SDSS monitor telescope (Tucker et al. 2006) using linear magnitude and color transformations derived from standard stars. These stars were chosen not because they are solar analogs, but because they have colors similar to those of the standard stars and hence the linear transformations are applicable. Because these data are ultimately compared with data calibrated onto the Landolt system, ideally these three stars would also have Landolt magnitudes. Unfortunately, this is not the case. Since the SEDs, effective bandpasses, and SDSS magnitudes are known, synthetic photometry can be used to compute the offset between SDSS and AB (Fukugita et al. 1996) magnitudes. We can then place the derived light-curve parameters on the Landolt system (i.e., m_B) using our fundamental flux standard (BD 17° 4708), for

---

[4]The ACS calibration involves setting the magnitude of the CALSPEC SED of Vega to zero using synthetic photometry, so updates on the SED or bandpasses change the zero point.
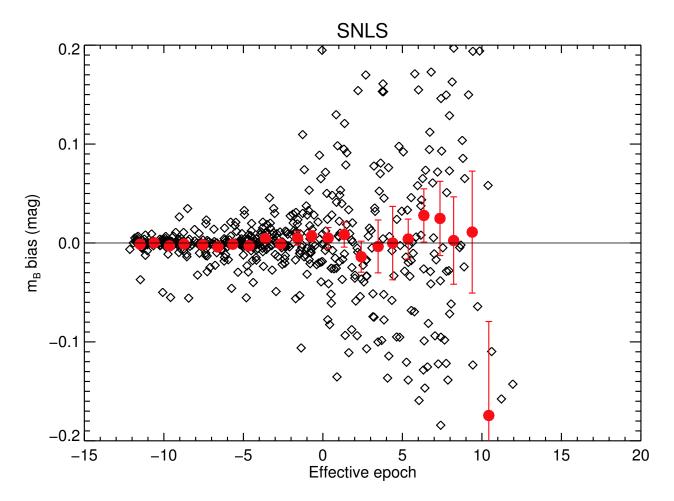
Fig. 2.— Bias in m_B as a function of first epoch relative to maximum after day -15 for *HST*-like data. The diamonds are the individual SNe used in this test, the red circles with errors are the values averaged in 1 day bins.

which we have Landolt magnitudes and whose SED shares the same calibration as the solar analogs. The last step is sensitive to the V magnitude of BD 17° 4708, unlike the rest of our analysis – avoiding this dependence is why the current SNLS calibration is tied to the Landolt system, despite the large differences between the MegaCam and Landolt passbands. Offsets to the AB system are given in Holtzman et al. (2008). As was the case for the *HST* data, we update these offsets for the Feb 2010 CALSPEC update, which changed the SEDs of the solar analogs slightly and thus the SDSS to AB conversion factor by 1% ; when combined with the changes in the SED of BD 17° 4708, the net effect of the CALSPEC update is to change the flux calibration of the SDSS data by about 0.3% in g and less in ri; the corresponding AB offsets $m_{AB} - m_{SDSS}$ are 0.021, 0.005, and 0.018 mag, respectively. Here and elsewhere we use the mean SDSS bandpasses of Doi et al. (2010).

2.6. *U*-band Observations of Nearby SNe

K09 devote considerable attention to discrepancies between low-z rest-frame U-band data and higher redshift surveys (§10.1.3 and 10.2.4), while finding that SNLS and SDSS SNe at $z = 0.3$ were consistent. We will not repeat all of this discussion here, but we independently confirm their result on a larger nearby sample, finding strong evidence for issues with observer-frame U data on several fronts. First, the observer-frame U data shows considerably more scatter in light-curve fits than can be accounted for by the observational uncertainties, which is not the case for higher-redshift samples in rest-frame U. In fact, despite their larger photometric uncertainties, the rms of the rest-frame U-band data for the higher-z samples is much lower than that for the nearby ones. Second, in color-color space (i.e., rest-frame $U - B$ versus $B - V$ at the epoch of maximum B brightness) low-z SNe with observer-frame U observations show much more scatter and a large systematic offset (~ 0.1 mag) compared with the other samples. Third, if observer-frame U-band photometry is included in the cosmological fits, then there is significant ($\sim 2\sigma$) tension in the residuals from the best fit cosmology between the low-z SNe and the others, which all agree quite well with each other, while without the U-band data, there is good agreement (§4.5).

In addition to ~ 90 low-z SNe observed in the U-band, we also have 14 nearby CSP SNe with rest-frame u'-band data. These SNe have virtually identical selection effects as the CfAII and CfAIII samples, since all three samples are primarily composed of KAIT-discovered SNe. Tellingly, the CSP sample displays none of the above pathologies – the scatter around the light-curve template is small, the SNe are completely consistent with the higher redshift samples in color-color space, and this introduces no tension in the residuals from the best-fit cosmology. We can quantify this by looking for offsets in the rest-frame

$U - B$ versus $B - V$ relation. The low-z U band sample is offset from the SNLS+SDSS samples by 0.047 ± 0.005 mag toward redder values of $B - V$ for a fixed $U - B$ – nearly 10σ – while the CSP u' sample is offset by 0.001 ± 0.006 mag; not only are the low-z U-band data inconsistent with the SNLS and SDSS data, they are also quite inconsistent with the low-z u' data. Interpreted in terms of a change in $U - B$, which is natural because the $B - V$ versus $V - R$ relation shows no problems, this corresponds to a blueward shift of ~ 0.1 mag in $U - B$ for the low-z SN relative to the high-z ones, which is consistent in sign and magnitude with that seen by K09.

We are uncertain where the problem lies. U-band data are notoriously difficult to calibrate due to the wide range of detector and filter responses, the fact that the blue end of the bandpass is affected by the atmospheric cutoff, which can be highly variable, and the fact that many U-band filters suffer from red leaks, but none of these seem adequate to fully explain the effect. An interesting possibility may be that the fundamental flux calibration is deficient in the UV, which would only affect the nearby sample. However, this would not explain the large observed scatter. Furthermore, the consistency of the CSP sample argues that if there is any problem it is not with the spectral calibration itself, but rather with the U-band magnitudes of the flux standards. Note that the offset between a Vega-based flux calibration and one based on BD 17° 4708, as used in this analysis, is not nearly large enough to explain these effects. A third possibility is that the response of the effective Landolt U-band is very different than that given by Bessell (1990), which is usually used as a starting point in SN analyses. This is suggested by Maíz Apellániz (2006), who find a substantially different U-band response curve using an *HST* spectroscopic database; however, the passband suggested there does not alleviate these problems. Finally, it is possible that this discrepancy is evidence for evolution in SN properties with redshift, but if so this evolution must be extremely sudden, turning on abruptly around $z = 0.25$, and then there is no additional evolution out to at least $z = 1$, which seems unlikely. Such evolution must also somehow not affect the CSP sample. Ellis et al. (2008) find no evidence for evolution in a spectroscopic study of UV spectra, but this is hindered by the lack of low-z rest-UV spectra. This issue deserves further study, and with the repairs to *HST* it will now be possible to improve this test.

K09 argue that it is likely that this is due to some unidentified issue with the low-z observer frame U-band, and includes this effect as their major source of systematic uncertainty. We concur with the former, but take the more aggressive position of excluding all observer-frame U-band observations from our analysis, while retaining rest-frame UV data for higher redshift SN. Interestingly, two recent spectroscopic studies comparing low- and high-redshift SNe show some evidence for spectral evolution in the near-UV (Cooke et al. 2010; Foley et al. 2010), although there is some concern about selection effects. However, in

both cases the evolution is in the opposite sense to that seen here – that is, the low-z SNe are redder in $U - B$ in these studies, while they appear bluer in our comparison, which only exacerbates the problem.

2.7.   Selection Effects for External Samples

SN samples are expected to suffer from Malmquist bias (Malmquist 1936), which we explicitly correct for in our analysis. In addition to the usual peak flux bias, there should be additional effects associated with the stretch and color of the SNe: for a given peak luminosity, higher stretch, more slowly declining SNe Ia will stay above the detection threshold longer and hence be easier to find. Because our survey has different depths in different passbands, at the high redshift end bluer SNe will have more of their flux in bands we are more sensitive to, and hence will also be easier to detect. The effects of Malmquist bias differ from survey to survey; some of the low-z surveys, which are galaxy-targeted, suffer from little Malmquist bias when searching for SN candidates, while others are traditional flux-limited surveys and are strongly affected. The simulations used to calculate these corrections for SNLS are described in Perrett et al. (2010), and are based on inserting 2.4 million fake SNe into the real images and re-running the extraction pipeline. We are not in a position to apply a similar analysis to the external samples, although we do compute and correct for Malmquist-type effects in these samples as described below. The mean Malmquist bias for the SNLS sample is about 0.03 mag at $z = 1$ for m_B after correction for light-curve shape and color.

For the *HST* and SDSS samples, we make use of the detection and spectroscopic selection models of K09 (§6), which are expressed in terms of the redshift and derived light-curve parameters. These models assume certain distributions of stretch, color, and $\sigma_{\rm int}$. In order to ensure consistency with our SNLS estimates, we generate our own simulated samples using the same distributions as the SNLS simulations and apply the K09 selection models to the results in order to estimate the Malmquist bias.

Our cosmological analysis corrects for both SN light-curve shape and color, which significantly reduces the effects of Malmquist bias. However, the relationship between color, shape, and peak magnitude is not perfect, so it is useful to study residual effects. For example, for an SN at $z = 0.45$ where $r_M \mapsto B$ and $i_M \mapsto V$, because the SNLS i_M images are deeper, an SN which fluctuates red relative to the mean color relation may be harder to detect without affecting the peak rest-frame B magnitude, m_B. In some previous studies (e.g., Perlmutter et al. 1999; Knop et al. 2003) it sufficed to treat Malmquist bias as a simple offset between the nearby and distant SN samples. Because the measurement of w is

more sensitive to the shape of the luminosity-distance relationship within our sample, this approach no longer suffices, and we map out the redshift dependence for all but the nearby sample. Applying selection effects as priors when fitting the SN parameters is unwieldy in the SALT2 and SiFTO frameworks, but is an equally valid approach often used with the MLCS light-curve fitting package (as in Wood-Vasey et al. 2007, K09). Instead, we apply the mean corrections to the relative distance moduli as a function of redshift to each sample. Our tool for calculating these corrections are Monte Carlo simulations of artificial SNe tuned to match the observed properties of the SN samples.

2.7.1. Malmquist Bias in the Nearby Sample

A nice feature of the nearby sample is that we do not need to know the redshift dependence of the Malmquist bias, at least for standard cosmological models, because of the very simple form of the luminosity-distance-redshift relation. Because of the simple nature of the luminosity-redshift relation, if a nearby sample has a sharp cutoff in magnitude, then the mean Malmquist bias is 0.032 mag, independent of the cutoff. In reality, however, any cutoff is unlikely to be perfectly sharp (which tends to decrease the amount of bias), some nearby samples are poorly described by magnitude limits (those based on searching known galaxies), and sample cuts tend to reduce the amount of bias.

We have two tools for analyzing the Malmquist bias in the nearby samples. First, the bias can roughly be measured by examining at what point on the light-curve the SNe were discovered – if most of the SNe were discovered around peak brightness, then the sample is more biased than one in which SNe are mostly discovered before peak. The bias for the Calán/Tololo sample was considered in Perlmutter et al. (1999) using this approach, who found a bias of 0.04 mag. Since this is larger than the value obtained for a sharp flux cutoff (0.032 mag), we adopt the latter. The CfAI sample was partially discovered by searches which targeted specific host galaxies, and therefore should suffer from less Malmquist bias than the Calán/Tololo survey. An examination of the light-curves suggests that this data set is not entirely free from bias, since many of the SNe were discovered near peak. Applying the same technique as was used for the Calán/Tololo sample, we estimate a Malmquist bias of 0.02 mag for this sample. The CfAII sample was almost completely based on galaxy-targeted searches, and hence should suffer from even less Malmquist bias. Indeed, most of the SNe from this sample used in our analysis were discovered prior to peak luminosity. Therefore, we adopt a value of 0 mag here. The CSP is selected from a combination of the same galaxy-targeted survey and the SDSS survey. For the latter, at the redshifts of these SNe, Malmquist bias should be completely negligible, so we also adopt a value of 0 mag.

The size of the CfAIII sample allows us to adopt a more sophisticated treatment. This sample was also largely selected from galaxy-targeted surveys, and therefore the detection efficiency should be fairly flat with redshift. This is confirmed by the fact that the earliest epoch of photometry relative to maximum is mostly uncorrelated with redshift. However, the spectroscopic selection criterion did make use of a magnitude cut of ~ 18.5 mag, which introduces Malmquist bias. This is in turn confirmed by the fact that the mean color of the sample changes with redshift, with few of the faint/red SNe in the upper half of the redshift range, which would not be the case in the absence of selection effects.

We model the spectroscopic selection function of the CfAIII sample by comparing the predicted number of SNe for an unbiased survey to that actually obtained as a function of redshift. We generate 32 million fake SNe using the same color and stretch distributions as the SNLS Malmquist simulations distributed evenly per comoving volume element[5] and then attempt to fit the parameters of a selection function to match the observed redshift distribution. We use a logistics function to model the selection probability P in terms of the peak magnitude: $P = 1/\left(1 + \exp\left[\gamma\left(m - m_0\right)\right]\right)$. Most low-z SN searches are carried out either unfiltered or in very broad bands, so for m we consider both V and R, finding identical parameters. We obtain $\gamma = 2.76 \pm 0.31, m_0 = 16.29 \pm 0.20$ mag. The amount of bias depends weakly on γ, but is independent of m_0. These imply a mean Malmquist bias for this sample of 0.027 mag.

2.7.2. Malmquist Bias in the HST Sample

Strolger et al. (2004) argue that the *HST* searches are sufficiently deep to suffer from little or no Malmquist bias out to the maximum redshift SN discovered. Because of the small size of the *HST* sample, we compute the Malmquist bias for each SN individually by constructing 10,000 constrained realizations of our SN model consistent with the observed properties of that SN given the observational uncertainties, and then apply the models of K09. This only models the search efficiency, and not the spectroscopic selection. Riess et al. (2007) argues that there are no spectroscopic losses below $z = 1.4$, so we cut the *HST* sample at this redshift because we cannot model the Malmquist bias for the single more distant SN. We find that the Malmquist bias for the remaining SNe is small, less than 0.01 mag in all cases.

However, this conflicts with the fact that the majority of the $z > 1$ *HST* SNe were

---

[5]We include a small correction for the evolution in the SN rate with z over this range from the $A + B$ model fits of Perrett et al. 2011, in preparation, although this has negligible effect.

discovered at or near maximum light – if there were no Malmquist bias, we would expect many more of the SNe to be discovered before maximum, as is seen with the lower-redshift *HST* SNe. We do not understand the reason for this discrepancy, but it has almost no effect on our fits to (constant) w.

2.7.3. Malmquist Bias in the SDSS Sample

Our method for modeling the SDSS Malmquist bias is described above. The resulting mean bias per redshift bin is shown in figure 3. The mean amount of bias is relatively small, perhaps because for SDSS it was easy to obtain spectra of even the faintest detected candidates with 8m class telescopes.

2.8. Peculiar Velocities

Most analyses have attempted to include the effects of SN peculiar velocities (relative to the Hubble flow) by assuming a large intrinsic velocity dispersion as an additional source of redshift uncertainty. These are only important for nearby SNe. This approach neglects any correlations in the flows. We instead follow the discussion presented in Neill, Hudson, & Conley (2007) and explicitly correct for peculiar velocities on an SN-by-SN basis. The peculiar velocity model used in this analysis (Hudson et al. 2004) is based on the galaxy density field in the nearby universe ($z < 0.06$) and accounts for infall into nearby superclusters (e.g. Virgo, Hydra-Centaurus and Perseus-Pisces) as well as a large-scale bulk flow. The accuracy of the corrections estimated to be ± 150 km s^{-1} for an individual SN, but this is a random uncertainty whose importance can be reduced by observing multiple SNe. Applying this model adjusts the CMB frame redshift of nearby SNe, but also applies a (much smaller) correction to their peak magnitudes because the standard luminosity-distance relation is not quite correct in the presence of peculiar velocities, although this is only important at low-z – see Hui & Greene (2006) for details.

3. COSMOLOGICAL RESULTS AND STATISTICAL UNCERTAINTIES WITH THE COMBINED SN SAMPLE

After the selection criterion are applied, the combined SN sample for our primary analysis consists of 472 SNe Ia: 123 nearby, 93 from SDSS, 242 from SNLS, and 14 from *HST*. These numbers do not include those nearby SNe that are not in the Hubble flow but which
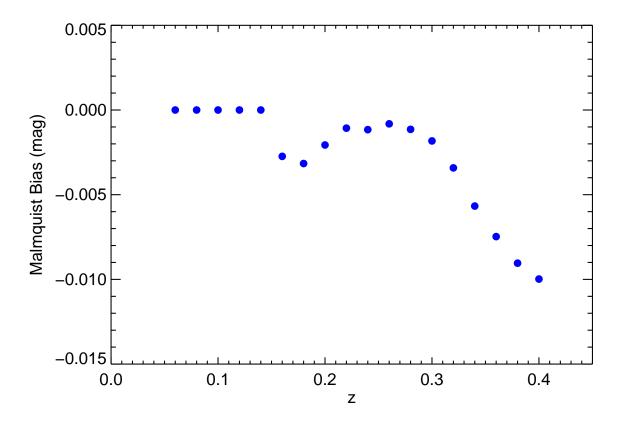
Fig. 3.— Mean Malmquist bias as a function of redshift for the SDSS sample. The sharp feature at $z = 0.15$ is an artifact of the discontinuous spectroscopic efficiency model of K09 and has little effect on the cosmological constraints.

are used for training the light-curve fitters. The parameters of each subsample are given in table 3.

3.1. Model for Relative Distance Moduli

The χ^2 of the data relative to our model is similar to that of A06. In the absence of any covariances between SNe,

$$\chi^2 = \sum_{\text{SNe}} \frac{(m_B - m_{\text{mod}})^2}{\sigma^2}, \tag{1}$$

where m_B is the rest-frame peak B band magnitude of an SN, m_{mod} is the predicted magnitude of the SN given the cosmological model and two other quantities (stretch and color) which describe the light-curve of the particular SN, and σ includes both the uncertainties in m_B and m_{mod}. The model magnitude is given (for SiFTO) by

$$m_{\text{mod}} = 5 \log_{10} \mathcal{D}_L \left(z_{\text{hel}}, z_{\text{cmb}}, w, \Omega_m, \Omega_{DE} \right) - \alpha \left(s - 1 \right) + \beta \mathcal{C} + \mathcal{M}. \tag{2}$$

where $\mathcal{D}_L$ is the Hubble-constant free luminosity distance, z_{cmb} and z_{hel} are the CMB frame and heliocentric redshifts of the SN, s is the stretch (a measure of the shape of the SN light-curve), and $\mathcal{C}$ is color measure for the SN. For SALT2, the expression is similar, with $(s - 1) \mapsto x_1$, where x_1 is a different measure of the light-curve shape. α and β are nuisance parameters which characterize the stretch-luminosity and color-luminosity relationships, reflecting the well-known broader-brighter and bluer-brighter relationships, respectively. $\mathcal{M}$ is another nuisance parameter representing some combination of the absolute magnitude of a fiducial SN Ia and the Hubble constant; even if one of these is known from some other measurement, the other still has to be marginalized over.

m_B, s, and $\mathcal{C}$ will naturally all be correlated for an individual SN because they are determined from the same data. Furthermore, their values are correlated between different SNe in the presence of systematic effects and statistical uncertainties. For example, a systematic uncertainty in the r_M zero point will directly affect m_B for all SNLS SNe between $z = 0.35$ and 0.55 (since rest-frame $B \mapsto r_M$ in this redshift range), and will have indirect effects even for an SN without r_M measurements because it affects our empirical SN models by changing the templates and also changes the measured color-luminosity relationship (β). Introducing a vector of model residuals over the SN sample $\Delta \vec{\mathbf{m}} = \vec{\mathbf{m}}_B - \vec{\mathbf{m}}_{\text{mod}}$ then a better expression for χ^2 is

$$\chi^2 = \Delta \vec{\mathbf{m}}^T \cdot \mathbf{C}^{-1} \cdot \Delta \vec{\mathbf{m}}. \tag{3}$$

Table 2.   Contributions to the Low-z SN Sample

Source	Initial Number[a]	All Uses	Cosmology Fit
Calán/Tololo	29	19	17
CfAI	22	10	7
CfAII	43	18	15
CfAIII	172	60	58
CSP	20	20	14
Other	37	23	12
Total	323	150	123

[a]After the removal of duplicates, known peculiar SNe, and SNe with problematical photometry.

Note. — The relative contributions of various sources to the low-z sample used in this paper.

Table 3.   Summary of SN Samples

Sample	Redshift Range[a]	$N_{\rm SN}$	$\langle N_{\rm points} \rangle$[b]
low-z	$0.01 - 0.10$	123	32
SDSS	$0.06 - 0.4$	93	27
SNLS	$0.08 - 1.05$	242	37
HST	$0.7 - 1.4$	14	10

[a]Redshift range from this sample included in our analysis.

[b]Average number of photometric epochs in the range -25 to $+30$ days, the range used in our light-curve fits.

It is useful, for reasons explained in §4.1, to further factor the total covariance matrix as:

$$\mathbf{C} = \mathbf{D}_{\text{stat}} + \mathbf{C}_{\text{stat}} + \mathbf{C}_{\text{sys}}. \tag{4}$$

The first term is the purely diagonal part of the statistical uncertainty given by

$$\mathbf{D}_{\text{stat}, \, ii} = \sigma^2_{m_B,i} + \alpha^2 \sigma^2_{s,i} + \beta^2 \sigma^2_{\mathcal{C},i} + \sigma^2_{\text{int}} + \left(\frac{5\,(1+z_i)}{z_i\,(1+z_i/2)\log 10} \right)^2 \sigma^2_{z,i} + \sigma_{\text{lensing}} + \sigma_{\text{host correction}} + C_{m_B s \mathcal{C},i}. \tag{5}$$

$D_{\text{stat}, \, ii}$, or something similar, is the only term that has been included in most analyses. Here $\sigma_{m_B,i}, \sigma_{s,i}, \sigma_{\mathcal{C},i}$ are the errors on the fitted light-curve parameters of the ith SN, and σ_{int} represents the intrinsic scatter of SNe Ia. $C_{m_B s \mathcal{C},i}$ represents the covariance terms between m_B, s, and $\mathcal{C}$ for each SN, and is a function of α and β. We will discuss how to construct the off-diagonal parts of the covariance matrix ($\mathbf{C}_{\text{stat}}$ and $\mathbf{C}_{\text{sys}}$) in §4.1.

We use the empty-universe approximation for the relation between redshift uncertainty and the associated magnitude uncertainty (with z_i being the redshift in the rest frame of the CMB). This is appropriate because σ_z is only important for the lowest redshift SNe in our sample. Future surveys which make use of more uncertain photometric redshifts may have to use the full form for the redshift uncertainty, which depends on the cosmological parameters. The lensing term represents the statistical uncertainty caused by gravitational lensing (§3.4), and the host correction term relates to the statistical uncertainty in the host-galaxy correction (§3.2).

3.2. Correcting for Host-galaxy Properties

Sullivan et al. (2010a) explore various approaches for including host-galaxy information in the cosmological fits to correct for the dependence of SN residual from the Hubble relationship; see that paper for the details of how we measure host-galaxy parameters using broad band photometry and the Z-PEG package (Le Borgne & Rocca-Volmerange 2002). Without a clear physical understanding of the cause, or even which descriptive host parameter best accounts for the variation (metallicity, host mass, or star formation rate), they find that simply splitting the sample on one of these parameters and allowing the absolute magnitude ($\mathcal{M}$ in our analysis) to be different between the two samples is sufficient to describe the observations. It seems likely that the true relationship is both more complex and more continuous, but the current data do not require a more sophisticated model. Therefore, we adopt this approach here. This is the first analysis to incorporate these corrections.

Since all that is available for the vast majority of the SN hosts in our combined sample is broad-band photometry, the host galaxy parameters can have considerable measurement

correlations, and there is little evidence to favor one over the others. Therefore, we split our sample based on host-galaxy stellar mass at $10^{10} M_\odot$, since it is the most directly constrained parameter. The effects of changing this selection are included as a systematic as described in §5.8. For systematic tests, instead of allowing $\mathcal{M}$ to vary based on host type, α and β might be allowed to differ between subsamples, as explored in Sullivan et al. (2010a); Lampeitl et al. (2010). Because the SNe light-curve properties are physically correlated with host type (i.e., higher-stretch SNe are preferentially found in star-forming hosts), these approaches are almost entirely degenerate with the $\mathcal{M}$ model for our purposes, although they have different physical implications.

Not including this correction has a substantial effect on the measured cosmological parameters. The difference in $\mathcal{M}$ between the two samples is ~ 0.075 mag (Sullivan et al. 2010a). In addition to the systematic uncertainties discussed in §5.8, we include the statistical uncertainties in the measured host masses by multiplying this magnitude difference by the probability that a given SN host is assigned to the wrong subsample due to measurement uncertainties. The overall mass scale does not affect our correction since it will affect all host masses identically.

3.3. Combining SALT2 and SiFTO

We fit all SNe using both SiFTO and SALT2. Since our analysis gives us no clear reason to choose either fitter, we combine results from both in our final analysis, although we show the results from each. The details are discussed in §5 of G10. We also include the difference between the two fitters in our systematic uncertainty budget (§5.2). We give the parameters from each fitter an equal weight in the combination. It would be incorrect to treat the parameters from each as independent since they are fit to the same data; this would unjustifiably reduce our uncertainties by $\sim 1/\sqrt{2}$. One approach would be to treat the relative distance moduli from each fitter as perfectly correlated. However, this has the side effect of making it impossible to ask questions like what the uncertainty on the combined peak magnitude is, and is complicated to extend to the systematics analysis. We have therefore adopted a simpler approach, which is to average the covariance matrices for each statistical and systematic term derived for SiFTO and SALT2. This satisfies our requirement that the final uncertainty should be larger than the minimum of the SALT2 and SiFTO uncertainties for each SN for every individual term.

3.4. Statistical Results

As mentioned previously, we add some additional, intrinsic scatter (σ_{int}) to the peak magnitudes of SNe Ia to match observations. This value will include both true intrinsic scatter, related to our imperfect understanding of SN physics, and any mis-estimates of the photometric uncertainties, uncorrected selection biases, etc. We have no way to distinguish between these contributions. Therefore, we allow different values of σ_{int} for each sample (e.g., SNLS, low-z, SDSS), as is supported by the data. The σ_{int} values are derived by requiring the χ^2 of the best fitting Ω_m, w cosmological fit to a flat universe to be one per degree of freedom for each sample, including systematic effects, and are given in table 4. The cost is that we weaken our ability to detect any deviations from our cosmological parameterization that are much smaller than the intrinsic scatter – but such deviations would be difficult to detect in any case. The σ_{int} are held fixed at these values for all other fits (e.g., Ω_m, Ω_Λ). If the data turn out to be inconsistent with a flat universe, fixing these values unduly penalize some data sets – see S11 for fits that allow for spatial curvature.

All SNe are corrected for Galactic extinction using the maps of Schlegel, Finkbeiner, & Davis (1998), including the estimated 10% random uncertainty for each SN (there is also a correlated systematic uncertainty discussed in §5.6). We correct for peculiar velocities in the nearby sample (§2.8), Malmquist bias effects (§2.7) for all samples, and assign a random peculiar velocity uncertainty of 150 km s^{-1}, as is appropriate after the peculiar velocity correction. Our statistical uncertainties also include the random uncertainty in the SN model, as described in Appendix A.3 of G10; this means that the statistical covariance matrix between SNe is not diagonal: $\mathbf{C}_{\text{stat}} \neq \mathbf{0}$. We include the measurement uncertainties in the host-galaxy masses as described in §3.2, and random, uncorrelated scatter due to lensing following the prescription of Jönsson et al. (2010): $\sigma_{\text{lens}} = 0.055\,z$.

Table 4: σ_{int} and rms values for various samples

Sample	σ_{int}	rms
low-z	0.113	0.153
SDSS	0.099	0.143
SNLS	0.068	0.156
HST	0.082	0.242

Note. — The uncertainty in each value is about 0.005 mag, and the RMS is around the best fit cosmology.

We perform two types of fits: one in which we compute probabilities over a grid and then report the mean value of the marginalized parameters, and a χ^2 minimization routine that reports the best fit. We should not expect the two to agree as they have different meaning, but as such it is useful to provide both – see Appendix 10 for further details. The light-curve parameters for the combined sample are given in table 5. The luminosity distance integral does not converge for $\Omega_m < 0$, so both fits effectively have the (very reasonable) prior of $\Omega_m \geq 0$. Our statistical constraints on Ω_m, w in a flat universe for a constant dark energy equation of state are shown in figure 4 using the marginalization approach, and are summarized in table 6. We find $w = -0.90^{+0.16}_{-0.20}$, consistent with a cosmological constant ($w = -1$). The Hubble diagram is shown in figure 5. The error introduced by using a simplified treatment of the nuisance parameters α and β (such as holding them fixed at their best fit values) is described in §4.6.

The σ_{int} and rms values for each sample are summarized in table 4. The rms residuals are similar for the low-z, SNLS, and SDSS samples, which is impressive considering the range in flux densities from $z = 0.01$ to $z = 1$. That of the *HST* sample is larger, but this is unsurprising due to the larger photometric uncertainties for such distant and difficult to observe SNe. K09 carry out the same analysis (tables 11 and 15) and find a similar rms for the SDSS sample as we do (0.15 versus 0.16 mag), but a smaller σ_{int} (0.08 versus 0.10 mag). However, they find significantly larger values of both measures for the other samples. Note that the SNLS data in this paper benefits from improved calibration and photometry, and both the low-z and SNLS samples are not the same as those in K09.

Comparing the individual light-curve fitters, for SiFTO we find $\Omega_m = 0.173^{+0.095}_{-0.098}$ and $w = -0.85^{+0.14}_{-0.20}$ and for SALT2 $\Omega_m = 0.214^{+0.072}_{-0.097}$ and $w = -0.95^{+0.17}_{-0.19}$ (all uncertainties statistical only). The contours are directly compared in figure 6. We include the difference between the light-curve parameters from the two fitters in our systematic uncertainty budget as described in §5.2.

The value for β is larger than that found by A06 (who found $\beta = 1.57 \pm 0.15$), although it remains inconsistent with a value of 4.1 expected if SN colors are primarily caused by MW-like dust in the host galaxies. There are several reasons for the increase. First, our current light-curve fitting frameworks handle the intrinsic variation between SNe Ia much better than in A06, particularly the observed scatter in the relations between different rest-frame wavelengths – see Appendix A.2 of G10 for details. This causes a significant increase in the measurement uncertainties for SN colors, and deweights high-z SNe which are primarily measured in the near-UV. These SNe, coupled with a simpler near-UV model, were driving the fits in A06 to low values of β. Modeling the uncertainty in the intrinsic uncertainty in the SN color relations is a complicated subject which can also present itself as apparent

Table 5. Light-curve Parameters of the Combined SN Sample

Name	z_{cmb}[a]	m_B[b]	s	$\mathcal{C}$	$\log_{10} M_{\mathrm{host}}$[c]	$\mathrm{MJD}_{\mathrm{max}}$	Filters	Reference
sn2004s	0.010(0.000)	14.183(0.042)	0.973(0.026)	0.035(0.025)	12.07	53039.56(0.60)	BVR	1
sn1999ac	0.010(0.000)	14.130(0.030)	0.987(0.009)	0.056(0.018)	9.92	51249.82(0.07)	BVR	2
sn1997do	0.011(0.000)	14.317(0.036)	0.983(0.023)	0.056(0.025)	12.07	50765.81(0.14)	BVR	2
sn2006bh	0.011(0.001)	14.347(0.021)	0.814(0.008)	−0.045(0.019)	10.91	53833.01(0.06)	$BgVr$	3
sn2002dp	0.011(0.000)	14.597(0.030)	0.973(0.029)	0.113(0.024)	10.47	52450.45(0.20)	BVR	4

Note. — Combined SiFTO and SALT2 light-curve parameters for full data sample, with uncertainties in parentheses. Note that the individual variables are correlated, as are the values for different SNe due to systematic and light-curve fitter training uncertainties.

[a]Uncertainty does not include residual peculiar velocity uncertainty.

[b]Includes the lensing term and the effects of the statistical uncertainty in $\log_{10} M_{\mathrm{host}}$.
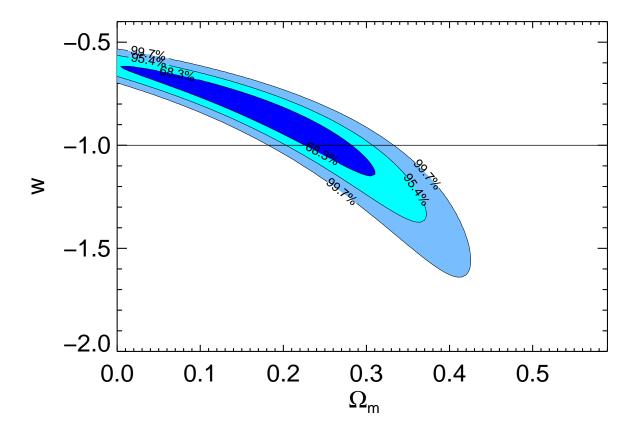
[c]Host stellar mass in solar masses.

References. — (1) Krisciunas et al. (2007), (2) Jha et al. (2006), (3) Contreras et al. (2010), (4) Hicken et al. (2009a), (5) Altavilla et al. (2004), (6) Krisciunas et al. (2000), (7) Hamuy et al. (1996), (8) Krisciunas et al. (2004b), (9) Krisciunas et al. (2001), (10) Riess et al. (1999), (11) Pastorello et al. (2007), (12) Krisciunas et al. (2004a), (13) Leonard et al. (2005), (14) Krisciunas et al. (2006), (15) Strolger et al. (2002), (16) Kowalski et al. (2008), (17) Holtzman et al. (2008), (18) G10, (19) Riess et al. (2007)

Fig. 4.— Statistical SN only constraints on Ω_m, w assuming a flat universe and constant dark energy equation of state.

Table 6: Results from SN-only fits

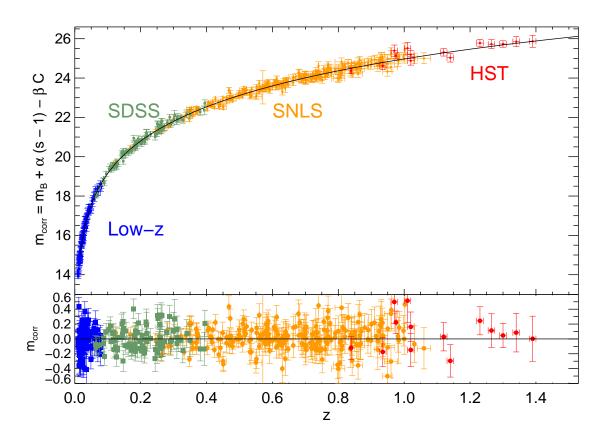| Uncertainties	Ω_m	w	α	β
Marginalization fits				
Stat Only	$0.19^{+0.08}_{-0.10}$	$-0.90^{+0.16}_{-0.20}$	$1.45^{+0.12}_{-0.10}$	$3.16^{+0.10}_{-0.09}$
Stat plus Sys	0.18 ± 0.10	$-0.91^{+0.17}_{-0.24}$	$1.43^{+0.12}_{-0.10}$	$3.26^{+0.12}_{-0.10}$
χ^2 minimization fits				
Stat Only	$0.19^{+0.09}_{-0.12}$	$-0.86^{+0.17}_{-0.19}$	$1.397^{+0.085}_{-0.083}$	$3.152^{+0.095}_{-0.093}$
Stat plus Sys	$0.17^{+0.10}_{-0.15}$	$-0.86^{+0.22}_{-0.23}$	$1.371^{+0.086}_{-0.084}$	3.18 ± 0.10

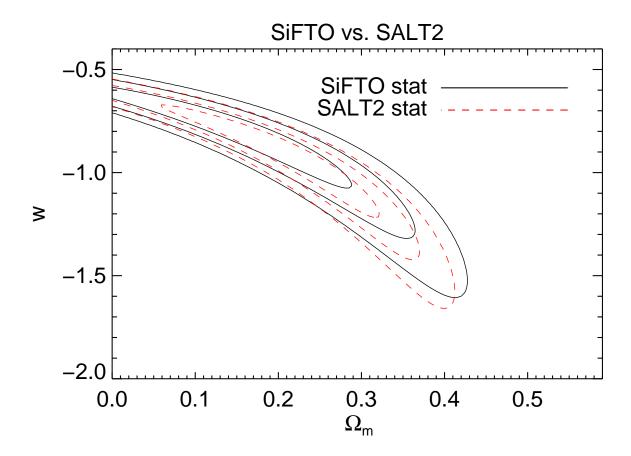Fig. 5.— Hubble diagram of the combined sample. The residuals from the best fit are shown in the bottom panel.

Fig. 6.— Comparison of SN-only statistical constraints for SiFTO and SALT2 Ω_m, w assuming a flat universe and constant dark energy equation of state.

evolution in β; see §5.7 for more details. Second, the method used to marginalize over α and β in A06 was intrinsically biased toward low values, as shown in Kowalski et al. (2008).

4. COSMOLOGICAL RESULTS INCLUDING SYSTEMATIC UNCERTAINTIES

The SNLS search and analysis proceeds through two independent pipelines, one in France and one in Canada, which can then be checked against each other. This has resulted in many improvements to both pipelines. In this section we explain the methodology we use for analyzing systematic uncertainties and how they impact our measurements (§4.1), present the combined results including both statistical and systematic effects (§4.4), explore tension between subsamples (§4.5), and discuss the consequences of simplified treatments (§4.6). The detailed descriptions of individual systematic terms are given in §5.

A difficulty in estimating systematic uncertainties is how to handle effects for which there is no clear physical model. The obvious example for SNe Ia is the possibility of undetected and uncorrected evolution in the SN population. While considerable effort and ingenuity have been devoted to making specific theoretical predictions for what form evolution might take and what signatures it might produce, currently these predictions frequently disagree with each other even as to the sign of possible effects. In the absence of a physical model, it will *always* be possible to imagine an evolutionary scenario which will pass all available tests, yet which will bias the cosmology by an arbitrary amount. We could artificially add some uncertainty in w to try to take this into account, but this is highly unsatisfying because the adopted value would be essentially arbitrary. In this paper we will concentrate on systematic effects for which we have some sort of model, which includes some simple models of evolution (§5.7).

4.1. Systematics Methodology

The difference between statistical and systematic uncertainties is not always entirely clear – what one author labels a systematic another may label a statistical uncertainty. Here we (mostly) adopt an effective definition which is that terms whose effects on our final uncertainty budget could be simply reduced by increasing the SN sample size are statistical uncertainties. Thus, for example, the uncertainty in the redshift of a particular SNLS SN is a statistical uncertainty, while the uncertainty in our zero points is a systematic uncertainty, since it affects multiple SNe in a correlated fashion and its importance cannot be reduced

by simply observing more SNe.

There is no standard method for handling SN systematic effects, but the most common approach works as follows: first, a fit without any systematic effects is performed. Then, the effect under consideration (e.g., a zero point offset) is applied to the SN sample and the cosmological parameters are re-measured. The difference in each parameter is taken as the systematic uncertainty in that parameter. Finally, the values for all of the known effects are then added in quadrature to form the final systematic uncertainty; we shall therefore refer to this shift-and-add approach as the quadrature method. Not all SN papers follow this approach, but with a few exceptions noted below, the alternatives have generally been even less sophisticated.

The main advantage of this method is that it is simple to implement once one has a list of systematic effects and estimates of their size (in reality, of course, this is by far the most difficult part of the analysis), but it has several disadvantages. One is that individual effects can be shown graphically, but the combined effects of multiple terms and their interrelations are difficult to visualize, relying on many-panel plots or graphical distortions that may not capture the full effects (see Knop et al. (2003); Conley et al. (2006a, K09) for examples). A more serious disadvantage is that it is difficult for others to include these systematics in their analyses. Even were the individual modified distance moduli made available, to apply this information to a new cosmological parameter space, or to include additional datasets, an entirely new cosmological fit has to be carried out for each systematic term, which can be prohibitively expensive. The result has been that SN papers themselves include systematic uncertainties, but most subsequent analyses ignore them; for example, the WMAP7 analysis of Komatsu et al. (2010) provides constraints that do not include SN systematics because they had no simple way to do so. Because systematic effects induce correlations between different SNe, in principle the best fit point should be modified by their presence, so ignoring these effects not only underestimates the uncertainties, but can lead to biased results. Additional criticism of the quadrature approach can be found in Barlow (2003).

One way to overcome both problems is to marginalize over all of the systematic terms during the fit. This marginalization can be carried out explicitly (using the uncertainty estimates for each term of the following sections as priors), or by adding a systematics covariance matrix to the statistical covariance matrix as in equation 4. We have used both methods in this analysis, but only the results of the latter are presented here. This separation of the covariance matrix has been used before in the SN literature, although in abbreviated form. In Perlmutter et al. (1999); Knop et al. (2003) a similar approach was used for the zero point uncertainties and their effects on m_B, with all other systematic effects handled via the quadrature method.

The statistical and systematic covariance matrices, $\mathbf{C}_{\text{stat}}$ and $\mathbf{C}_{\text{sys}}$ of equation 4, are generally not diagonal, which have also been neglected in most SN analyses. We compute them using standard techniques (e.g., Aharmim et al. 2005):

$$\mathbf{C}_{\text{sys},\,ij} = \sum_{k=1}^{K} \left(\frac{\partial m_{\text{mod}\,i}}{\partial S_k} \right) \left(\frac{\partial m_{\text{mod}\,j}}{\partial S_k} \right) (\Delta S_k)^2 \tag{6}$$

where the sum is over the K systematics S_k, ΔS_k is the size of each term (for example, the uncertainty in the zero point), and m_{mod} is defined as in equation 2. A similar expression is used for $\mathbf{C}_{\text{stat}}$. An example of such a term in $\mathbf{C}_{\text{stat}}$ is the uncertainty in the SN model. Since these uncertainties could be reduced with more SNe, they are a statistical uncertainty rather than a systematic one, but are correlated between different SNe and therefore cannot be included in $\mathbf{D}_{\text{stat}}$. Note that all of the components of $\mathbf{C}$ are functions of the nuisance parameters α and β, but not $\mathcal{M}$ because the relative distance moduli are independent of its value. Since $\mathbf{D}_{\text{stat}}$ is diagonal and the off-diagonal pieces of $\mathbf{C}$ have the right form, we can take advantage of the Sherman-Morrison-Woodbury formula (Golub & van Loan 1996) to compute $\mathbf{C}^{-1}$ in N^2 steps.

This approach must be modified slightly for effects which cannot be smoothly parameterized. An example is contamination by non-SNe Ia, which has the effect of removing an SN from the sample rather than changing m_B, s or $\mathcal{C}$. In such cases, we resort to studying SNe in redshift bins, where there are enough objects in each bin that the effect is sufficiently continuous for our purposes, and then reproject the systematic onto the individual SNe (e.g., non-Ia contamination §5.5) by determining what systematic uncertainty, assumed perfectly correlated between all SNe in that bin, would produce the same uncertainty in the mean bin parameters as the specified effect and then assigning it to those SNe.

Kowalski et al. (2008) have recently presented an alternative to the standard scheme for handling systematic effects which shares some characteristics with that adopted here. They treat all systematic effects as offsets between the peak magnitudes of nearby and distant SNe on a sample-by-sample basis. This could be replicated in our scheme by forcing $\mathbf{C}_{\text{sys}}$ to consist of blocks of identical values when ordered by SN sample, and in fact this transformation was used in Amanullah et al. (2010). This is a good approximation when each SN sample covers a very small range in redshift, since it does not allow for any effects within a sample, but this is not true of current high-z SN samples. Both papers also hold α and β fixed while calculating $\mathbf{C}_{\text{sys}}$, which both biases their values, and more importantly underestimates the uncertainty on w, as discussed in §4.6.

4.2. Light-curve Training

All current light-curve analysis frameworks are trained on SN data which share systematics in common with the data used to derive the cosmological constraints. While many previous analyses have included some estimates of the uncertainty in the light-curve models, none have properly considered the interplay of systematic effects with the training process, and hence have underestimated their systematic uncertainties. A unique feature of the current analysis is that both of the frameworks used for SNLS3 (SiFTO and SALT2) are trained on high-redshift SN data. This is only practical because the training process for both does not assume any relationship between redshift and distance (and, in fact, makes no use of distance information), and so are completely independent of the cosmological parameters. The benefit of including the high-redshift data in the training process is that it allows us to probe further into the blue, where the SNLS calibration of rest-frame near-UV data is much more secure than the corresponding low-z data. Because SNLS systematics affect the light-curve model, they will affect the derived parameters of even nearby SNe, and vice-versa. Excluding high-z data from the training sample would only increase the overall uncertainty budget.

We proceed in two steps for each systematic effect. First, we calculate the effect on the light-curve model. We then apply the modified model to all of the data, deriving a new set of light-curve parameters. A similar procedure could easily be applied to other light-curve techniques (e.g., MLCS2k2 or Δm_{15}).

4.3. Method for Presenting Systematic Effects

There are 134 individual systematic terms considered in this paper. Attempting to compare the importance of each term is difficult, and we have not found any entirely satisfactory method. In the parameter space we are most interested in for this paper (Ω_m, w for a flat universe with a constant dark energy equation of state) the size of the uncertainties on the cosmological parameters depends strongly on the values of those parameters. Therefore, introducing a correlated uncertainty which shifts the results can actually reduce the uncertainties. This effect is mitigated when we combine SN data with external constraints such as those from the Wilkinson Microwave Anisotropy Probe (*WMAP*) because they are nicely orthogonal to the SN ones, and therefore prevent the constraints from shifting too much. Therefore, some effects which seem very large when only SN are considered are minor when combined with *WMAP* and BAO results (see S11). Also, some systematic effects appear minor simply because the SNe they affect have little weight in the fits considered in this paper, which may not be the case for a different cosmological model.

With those caveats in mind, we present the effects of our systematics in three ways: first, in terms of the effects on Ω_m, w for SNe only, and second in terms of the effects on w with Ω_m fixed. The latter approximates the effects of including the BAO and *WMAP* constraints, which mostly improve the measurement by constraining Ω_m. In both cases, we compare the uncertainties on the cosmological parameters from a fit that only includes statistical uncertainties with one that includes statistical uncertainties plus only that systematic term. We also provide the relative size of the contour that encloses 68.3% of the probability, compared with the statistical only contours. This is similar in spirit to the Dark Energy Task Force figure of merit (Albrecht et al. 2006), and is perhaps the simplest way of expressing the importance of each term. Because of the curvature of the SN-only constraints, the area of the inner contour can actually increase while the marginalized uncertainties decrease. We caution against the practice of comparing the shifts in the best fit as a useful method of measuring systematic effects, as it can be misleading. An updated version of the effects on the cosmological parameters is given in S11 combined with external constraints such as baryon acoustic oscillations and CMB measurements. An example of the effects of one systematic (the i_M SNLS zero point, discussed in §5.1.1) on the light-curve parameters as a function of redshift is shown in figure 7.

4.4. Combined Statistical and Systematic Results

Including all identified systematic effects, the results for an Ω_m, Ω_Λ fit are shown in figure 8. The SN data alone require acceleration at high significance. The results for a flat universe with a constant dark energy equation of state are summarized in table 6, and the contours are shown in figure 9. We find $w = -0.91^{+0.16}_{-0.20} \, (\text{stat})^{+0.07}_{-0.14} \, (\text{sys})$, again consistent with a cosmological constant. An overview of the importance of each class of systematic effect is given in table 7 – calibration effects are by far the dominant type of identified systematic uncertainty. Overall, the systematic uncertainties degrade the area of the uncertainty ellipse by a bit less than a factor of 2 relative to the statistical-only constraints. Excluding the SDSS and CSP SNe (calibrated to a USNO system) increases the area of the uncertainty ellipse by about 10%; the improvement from including this data should be increased once the full benefits of cross calibrating these two samples with SNLS are realized.

It is interesting to compare the constraining power of the low-z, SDSS, and SNLS samples. This was also explored in K09, but with a smaller low-z sample (33 versus 123 SNe). The resulting constraints without each sample are shown in figure 10; as can be seen, the first year of SDSS data is not a good replacement for the nearby sample, although the full 3 year sample may alter this situation. Excluding the SNLS sample has a significant
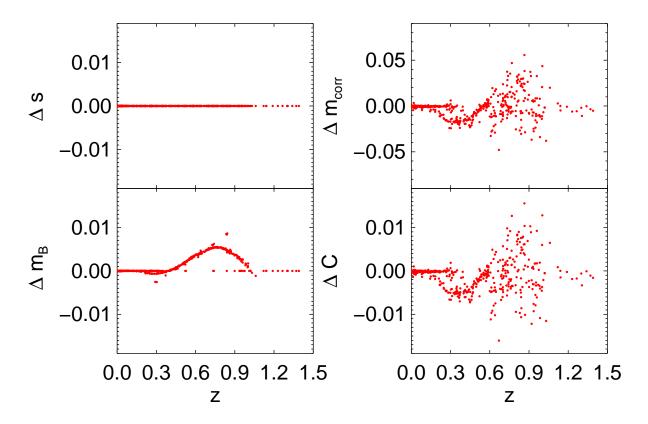
Fig. 7.— The effects of changing the i_M zero point by 6.1 mmag on various light-curve parameters and the corrected peak magnitude $m_{\rm corr} = m_B + \alpha\,(s-1) - \beta\mathcal{C}$, as a function of redshift. Note that because this affects our SN models (SALT2 and SiFTO), as the training sample includes SNLS data, this alters the corrected peak magnitudes of all SNe, not just those in the SNLS sample. Furthermore, because of the changes in the light-curve model, the offset in the derived color can actually be larger than the shift in the zero point.

Table 7: Identified systematic uncertainties

Description	Ω_m	w	Rel. Area [a]	w for Ω_m=0.27	Section
Stat only	$0.19^{+0.08}_{-0.10}$	$-0.90^{+0.16}_{-0.20}$	1	-1.031 ± 0.058	
All systematics	0.18 ± 0.10	$-0.91^{+0.17}_{-0.24}$	1.85	$-1.08^{+0.10}_{-0.11}$	§4.4
Calibration	$0.191^{+0.095}_{-0.104}$	$-0.92^{+0.17}_{-0.23}$	1.79	-1.06 ± 0.10	§5.1
SN model	$0.195^{+0.086}_{-0.101}$	$-0.90^{+0.16}_{-0.20}$	1.02	-1.027 ± 0.059	§5.2
Peculiar velocities	$0.197^{+0.084}_{-0.100}$	$-0.91^{+0.16}_{-0.20}$	1.03	-1.034 ± 0.059	§5.3
Malmquist bias	$0.198^{+0.084}_{-0.100}$	$-0.91^{+0.16}_{-0.20}$	1.07	-1.037 ± 0.060	§5.4
non-Ia contamination	$0.19^{+0.08}_{-0.10}$	$-0.90^{+0.16}_{-0.20}$	1	-1.031 ± 0.058	§5.5
MW extinction correction	$0.196^{+0.084}_{-0.100}$	$-0.90^{+0.16}_{-0.20}$	1.05	-1.032 ± 0.060	§5.6
SN evolution	$0.185^{+0.088}_{-0.099}$	$-0.88^{+0.15}_{-0.20}$	1.02	-1.028 ± 0.059	§5.7
Host relation	$0.198^{+0.085}_{-0.102}$	$-0.91^{+0.16}_{-0.21}$	1.08	-1.034 ± 0.061	§5.8

[a]Area relative to statistical only fit of the contour enclosing 68.3% of the total probability.

Note. — Results including statistical and identified systematic uncertainties broken down into categories. In each case the constraints are given including the statistical uncertainties and only the stated systematic contribution. The importance of each class of systematic uncertainties can be judged by the relative area compared with the statistical-only fit.

negative impact on the cosmological constraints.

4.5. Tension Between Data Sets

As a test of whether our estimates for systematic effects are reasonable, we compute the mean offsets in the residuals from the cosmological fit between different samples. We compute the weighted mean residual for each sample, including the statistical and systematic covariance matrices and assuming the best fit values of α and β. The results are summarized in table 8, and show no significant evidence for any disagreement between samples (the apparent increase with z is not statistically significant). Note that our estimates for each systematic term were constructed before this test was carried out. We have also compared the rms around the best fit for different sources of the low-z sample. Generally, these are consistent, with the exception of the CfAII sample (which has an rms of 0.20 mag, compared with 0.153 mag for the other samples). This is mostly due to a single SN which just barely passes our outlier rejection, SN 1999dg. Without this SN, the rms of the CfAII sample is 0.163 mag. The CfAI, CfAIII, and CSP samples show slightly below average rms (about 0.14 mag), but this is not statistically significant. Note that the relative weights of the samples in this comparison should not be taken too seriously because of the somewhat arbitrary way in which systematics were attributed to particular samples, although test for tension is meaningful. Specifically, the systematic uncertainties in the cross-calibration between the low-z and other samples are always assigned to the latter (SDSS, SNLS, etc.) which decreases their apparent weight; it would be just as valid to assign the cross-calibration uncertainties to only the low-z sample, which would make the other samples appear to have much more weight. The actual weights should be judged in terms of the consequences of removing each sample, as in figure 10.

4.6. The Consequences of Simplified Treatments

In this section we describe the consequences of various simplifying assumptions in the analysis. Figure 11 shows the effects of not including uncertainties related to the light-curve models, specifically omitting the model statistical uncertainty and the effects of the systematic effects on the model training. The consequences are not as severe as the effects of fixing the nuisance parameters, but still will obviously underestimate the uncertainties. The general effects are to underestimate the total uncertainty budget, although the size of the effect depends on the simplification.
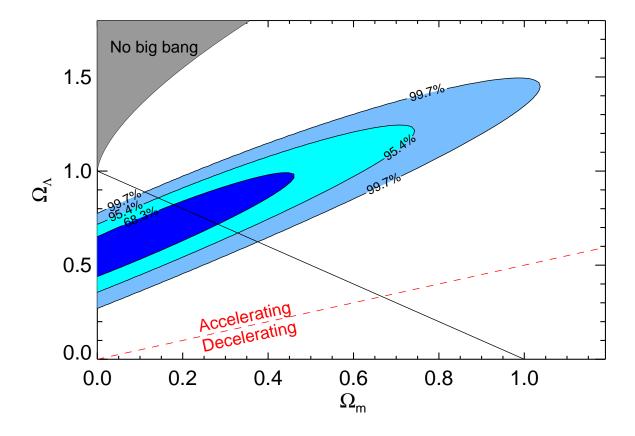
Fig. 8.— Ω_m, Ω_Λ (i.e., $w = -1$, but allowing for non-zero spatial curvature) contours including all identified systematic uncertainties.

Table 8: Tension between different SN samples

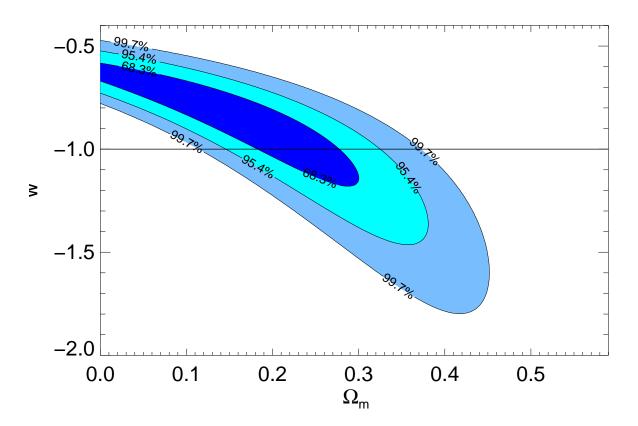| Sample	Mean offset (mag)	Uncertainty	N
low-z	-0.027	0.024	123
SDSS	0.020	0.027	93
SNLS	0.023	0.023	242
HST	0.043	0.072	14
Calán/Tololo	-0.027	0.046	17
CfAI	0.064	0.062	7
CfAII	0.051	0.049	15
CfAIII	-0.047	0.034	58
CSP	0.052	0.057	14
Other	0.052	0.057	12

Fig. 9.— Constraints on Ω_m, w in a flat universe including all identified systematic uncertainties.
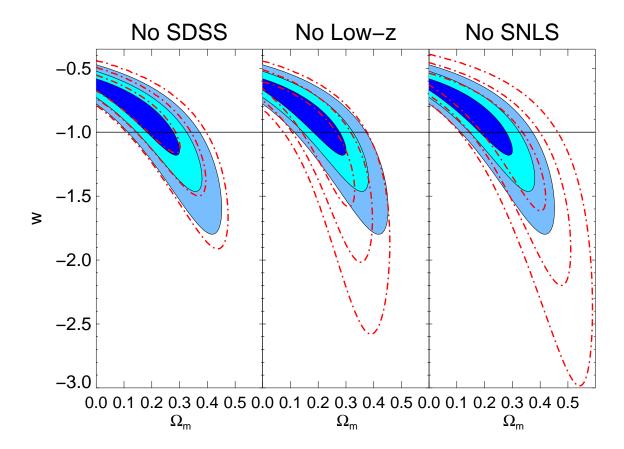
Fig. 10.— Comparison of the constraints on Ω_m, w in a flat universe excluding various samples. First, in the left panel, we exclude the SDSS sample (so the included samples are SNLS, *HST*, and low-z), in the middle the low-z sample, and on the right the SNLS data. The fits include all identified systematic uncertainties. The filled contours are the constraints with all samples, and the dashed contours exclude the labeled sample.

α and β are correlated with the cosmological parameters, with correlation coefficients of about 0.2 for the Ω_m, w fit; dropping the assumption of flatness or investigating time varying w generally increases these correlations. Therefore, the treatment sometimes found in the literature of fixing α and β at their best fit values and not fitting for them explicitly both underestimates the uncertainties and results in biased parameter estimates. A related simplification is allowing α and β to vary for the statistical uncertainty ($\mathbf{D}_{\mathrm{stat}}$ of equation 5), but holding it fixed when computing the systematics, as in Kowalski et al. (2008); Amanullah et al. (2010). This simplification also biases α and β, and therefore the cosmological parameters, as shown in the right hand panel of figure 11, and underestimates the uncertainties; for this sample, it amounts to underestimating the size of the inner uncertainty contour by $\sim 40\%$, although the effects on the marginalized uncertainties are modest.

5. INDIVIDUAL SYSTEMATIC TERMS

In this section we discuss the individual systematic terms in detail, also presenting their effect on the parameter estimates. Different categories of systematic effects are summarized in table 7. The systematic correlation matrix is shown in figure 12 to convey some of the sample- and redshift-dependent structure of the systematic effects. When summarizing individual terms, for brevity we only provide the effects on a fit for w assuming a flat universe and $\Omega_m = 0.27$ and the area relative to statistical only fit of the contour enclosing 68.3% of the total probability.

5.1. Calibration

The calibration of the SNLS data onto a standard photometric system is described in detail in R09. Achieving the target accuracy of $\approx 1\%$ has proven quite difficult, and remains the major source of systematic uncertainty in our measurement. In addition to the SNLS calibration, we also consider the calibration of the external samples included in our analysis. It is useful to separate the effects into a number of categories. Calibration can be considered to consist of two steps: first, the observations are standardized onto some photometric system, and second, they are converted from the standard system into (relative) fluxes in order to compare SNe at different redshifts. Items related to the first step are hereafter referred to as zero-point uncertainties, and those related to the second as flux calibration uncertainties, which includes both bandpass uncertainties and uncertainties in the magnitudes and SED of our fundamental flux reference, BD 17° 4708. In many other treatments these two terms are lumped together into a single "calibration error" term.
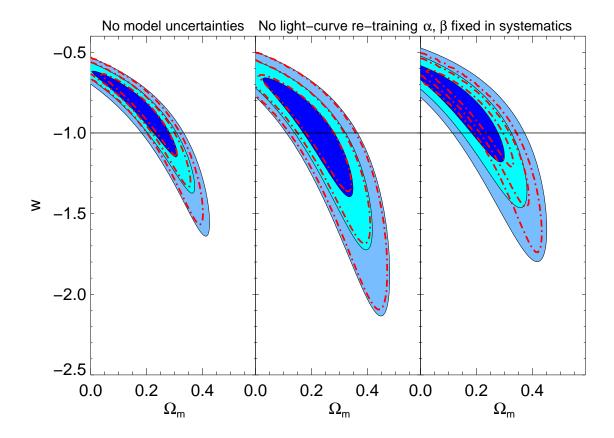
Fig. 11.— Effects of simplified systematic and statistical treatments of SN data. In the left panel, the statistical uncertainties with (filled contours) and without (dashed contours) the model statistical uncertainties are shown. In the center panel, the results of not including the effects of systematics on the light-curve model are shown for the combined statistical and systematic errors using the same labeling scheme. Finally, the right panel shows the consequences of holding α and β fixed when computing the systematics covariance matrix. The center panel is for SiFTO only, since it is easier to separate the retraining effects, but the other two panels show combined SALT2 and SiFTO results. The dashed contours are smaller than the filled contours, indicating that each simplification will underestimate the errors by some amount.
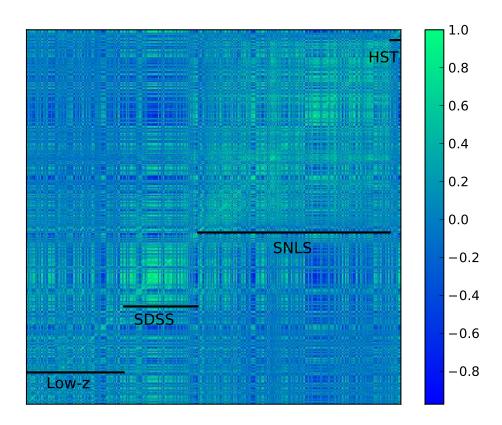
Fig. 12.— Combined correlation matrix ($\mathbf{C}_{\mathrm{stat}} + \mathbf{C}_{\mathrm{sys}}$ of §4.1) for the best-fit values of α, β sorted by sample, and by reshift within each sample. The purely diagonal statistical uncertainty term of equation 5 is not included. The redshift and sample dependent structure of the covariance information is quite complicated, and is not well represented by simple $(1 + z)^n$ type models.

The details of how most of these uncertainties are derived are given in §12 of R09. Here we present them in a somewhat different (but equivalent) form, breaking them down into various contributions in order to highlight which particular items cause the largest effect on the cosmological parameters. In addition to the effects discussed in R09, we also present calibration uncertainties for the external samples and uncertainties in the bandpasses. The various contributions are summarized in table 9. Calibration is the largest component of our systematics uncertainty budget. This is not due to the fact that we train our light-curve fitters on our own data, as can be inferred from the center panel of figure 11. The most important sub-terms are the colors and SED of BD 17° 4708 and the SNLS zero points.

5.1.1. SNLS Zero Points

Since SNLS observes in four filters and there are four fields which each have their own zero point, there are 16 zero-point terms to consider. However, most of the important terms are strongly correlated between different fields, and in fact in some cases between different filters. These are described in the first section of table 12 of R09 in combination with the uncorrelated random uncertainties of table 8 of that paper. The net effect is an uncertainty of around $2.5, 2.4, 5.9$, and 3.2 mmag in $g_M r_M i_M z_M$, with correlations. These uncertainties are fairly small, as benefits the considerable effort involved in the SNLS calibration; this is an advantage of large, single instrument surveys, which can devote considerably more effort to internal calibration than is practical for small surveys or those which make use of many instruments. In addition to these uncertainties, there is a small additional uncertainty of about 2 mmag in each filter arising from the differences between the point-spread function and aperture photometry used on the SNe and stars, as discussed in §3.2 of G10.

The net effect on the measured parameters is given in table 10. Because of the correlations, the breakdown by individual filter should only be considered approximate, but the largest effect stems from the i_M filter, which maps to rest-frame B near the redshift range where SNe are most sensitive to w.

5.1.2. Low-z Data Set Zero Points

The zero-point uncertainties of the nearby SN samples are generally not as well understood as those of SNLS. These samples are generally less homogenous than the SNLS sample, and are frequently not reported in the natural system of the telescope/detector. On the other hand, since the data come from many sources, both in terms of the telescope/detector sys-

Table 9: Calibration Systematics

Description	w for Ω_m=0.27	Rel area	Section
Stat only	-1.031 ± 0.058	1	
All calibration	-1.06 ± 0.10	1.79	§5.1
Colors of BD 17° 4708	-1.075 ± 0.075	1.31	§5.1.7
SED of BD 17° 4708	-1.026 ± 0.073	1.23	§5.1.8
SNLS Zero Points	-1.030 ± 0.069	1.21	§5.1.1
low-z Zero Points	-1.044 ± 0.065	1.13	§5.1.2
SDSS Zero Points	-1.028 ± 0.060	1.02	§5.1.4
MegaCam Bandpasses	-1.017 ± 0.066	1.20	§5.1.5
low-z Bandpasses	-1.027 ± 0.059	1.04	§5.1.6
SDSS Bandpasses	-1.026 ± 0.059	1.02	§5.1.6
HST Zero Points	-1.027 ± 0.058	1.03	§5.1.3
NICMOS Nonlinearity	-1.029 ± 0.059	1.05	§5.1.3

Note. — Individual calibration systematics.

Table 10: Zero Point Uncertainty Systematics

Description	Uncertainty (mag)	w for Ω_m=0.27	Rel Area
Stat only	$\cdots$	-1.031 ± 0.058	1
SNLS zero points	$\cdots$	-1.031 ± 0.070	1.21
g_M only	0.0028	-1.030 ± 0.059	1.00
r_M only	0.0028	-1.012 ± 0.060	1.12
i_M only	0.0061	-1.034 ± 0.064	1.10
z_M only	0.0035	-1.049 ± 0.060	1.02
low-z zero points	$\cdots$	-1.044 ± 0.065	1.13
CfAIII only	$\cdots$	-1.047 ± 0.061	1.08
CSP only	$\cdots$	-1.033 ± 0.059	1.01
B	0.015	-1.027 ± 0.060	1.03
V	0.015	-1.024 ± 0.060	1.01
R	0.015	-1.030 ± 0.059	1.00
SDSS zps	$\cdots$	-1.028 ± 0.060	1.02
g only	0.003	-1.029 ± 0.058	1.00
r only	0.004	-1.029 ± 0.059	1.01
i only	0.007	-1.030 ± 0.059	1.00
Scale from V	0.0062	-1.030 ± 0.059	1.00
HST zero points	$\cdots$	-1.027 ± 0.058	1.03

Note. — The SNLS zero points do not include the transformational uncertainties for BD 17° 4708, which are given in table 12. The combined uncertainties are given in table D.2 of R09. SDSS values do not include the effects of the filter mean wavelength uncertainties on the conversion to AB magnitudes, given in table 11. The HST zero-point uncertainties do include the effects of passband uncertainties. The zero-point uncertainties for the CSP sample are not broken down into individual terms because it constitutes a small fraction of the overall sample.

tems and the techniques applied to reduce the data, some of the calibration errors may partially average out across multiple samples, resulting in a smaller uncertainty. We are not in a position to model the zero-point uncertainties of the low-z data at the same level of sophistication as the SNLS zero points as we do not have access to the raw data, and the inhomogeneity would make this quite difficult in any case.

As discussed previously, the nearby SN sample is dominated by five large datasets. The observations for each of these sets are generally from 1-3 instruments at a single site, with a few exceptions. In addition to the raw measurement uncertainties in the calibrations, we expect uncertainties from at least two other sources. First, three of the five samples present photometry in the Landolt system rather than the natural system by applying linear transformations to the natural system magnitudes. These linear transformations are derived using stars, and therefore are not entirely appropriate for SNe. This introduces an error in the resulting magnitudes, one that is determined by the mismatch between the actual instrumental passbands and the Landolt system. Since the passbands are shared across many observations, and because these surveys cover a fairly limited redshift range, these uncertainties will be correlated between different SN. Second, and less importantly, uncertainties in the modeling of atmospheric extinction will be shared between most observations for a given survey.

However, in contrast to the earliest low-z samples, the CfAIII and CSP samples provide data in their natural systems. Since what we are really doing is comparing SN fluxes observed through low-z natural systems with SN fluxes in the MegaCam natural system, both of which are much more accurately measured, using the natural system frees us of the uncertainty related to the effective passbands of the Landolt system, which is a significant advantage.

Hamuy et al. (1993) discuss the photometric transformations used to calibrate the Calán/Tololo sample, giving a typical uncertainty of ~ 0.02 mag. For the CfAI sample, Riess et al. (1999) quotes an rms of 0.02 mags for the typical residuals for the calibration stars around the mean. No specific zero-point uncertainty is given by Jha et al. (2006) for the CfAII sample, but uncertainties in the color terms are estimated as $0.5 - 2\%$. Together, these estimates suggest that the zero points of the nearby data are quite well known. However, the treatments in these papers are generally focused on the contribution of zero-point uncertainties to individual observations, not the overall zero-point uncertainty. To estimate this value, we consider photometry of the same SNe from different sources. Riess et al. (1999) performs this comparison for two SNe, while Jha et al. (2006) carry out a similar analysis for 10 SNe and Hicken et al. (2009a) do so for 7. However, the latter is done in the Landolt system, and hence does not directly apply to the natural system magnitudes we use in this analysis. Some of these SNe do not have extensive enough coverage to allow a

detailed comparison. Considering these results, we estimate a global zero-point uncertainty of about 0.015 mag in BVR. We therefore adopt these values for the three samples on the Landolt system.

Hicken et al. (2009a) quote zero-point systematic uncertainties of 0.03 mag in $BVRIr'i'$, and 0.07 mag in U for the CfAIII sample. However, this refers to the Landolt transformed magnitudes, not the natural system magnitudes, which removes a substantial part of the uncertainty. The second major contribution comes from a comparison of the pipeline reductions with those of an earlier version used in Jha et al. (2006). They note that the scatter in the reduced magnitudes from object to object is typically about 0.02 mag, and adopt this as their systematic uncertainty. However, the pipeline offset varies randomly from object to object, and therefore is properly a statistical uncertainty for our purposes. The mean offsets between the two pipelines, however (0.005 mag), are systematic uncertainties, since their effects do not average down by combining distances to many SNe. The appropriate systematic uncertainty from this term is therefore 0.005 mag in each filter. In addition, we include the transformational uncertainties for the magnitude of our fundamental flux standard. While it is likely that these are actually smaller at the color of this standard than at color zero, we take the conservative approach of assuming that the measurement uncertainties are smallest for zero color and extrapolating. Putting these contributions together, we arrive at zero-point uncertainties of 0.011, 0.007, and 0.007 mag for BVR on the 4Shooter camera, and essentially identical values for BVr' on Keplercam.

The zero-point uncertainties of the CSP sample have not been analyzed in detail. Comparison with photometry of SNe in common is complicated because of the different photometric systems, but suggests the uncertainties are on the order of 1% or less. The CSP quotes zero-point uncertainties including the effects of bandpass uncertainties, which we treat separately; removing these results in 0.8% zero-point uncertainties. There are 7 well observed SNe in common between the CSP and CfAIII sample, so we can compare the photometry. Because they are in different systems, this comparison must be mediated by an SN model. We perform SiFTO fits to each SN and use the resulting model SEDs to K-correct the photometry from one system to another. We are interested in mean offsets between the two systems, so calculate the weighted mean difference and uncertainty, including the estimated calibration correlations. We find good agreement in $Bg'Vi'$, and somewhat poorer in r' (with differences of 0.8, 0.9, 1.5, 1.0, and 2.2σ, respectively), after outlier rejection (which actually increases these values), although the precision is only about 2%. The i value should be treated with some caution, as at the low redshift of the comparison objects it lies beyond the wavelength range we normally consider in SiFTO and SALT2. It is not clear how seriously to take the disagreement in r', but fortunately the r' zero point has virtually no impact on our results.

The effects are broken down into individual terms in table 10. The B and V filters can be seen to have the largest effects, but overall the low-z sample zero points are not a major contributor to our final uncertainty budget. Unsurprisingly, including the SDSS SNe significantly reduces the importance of these terms by providing an independent absolute magnitude anchor – without the SDSS sample, the combined statistical and systematic uncertainty from the low-z zero-point uncertainties is ± 0.074 rather than ± 0.064.

5.1.3. HST Calibration

In addition to the zero-point uncertainties, NICMOS suffers from multiple types of non-linearity: an expected exposure length non-linearity probably due to charge trapping, and a poorly understood non-linearity related to the count rate of the observed object (plus sky). The NICMOS pipeline attempts to correct for both non-linearities. However, these corrections, particularly for the latter effect, are not perfect, and therefore we include related terms in our uncertainty budget. The second nonlinearity is described in de Jong et al. (2006a,b); Shaw & de Jong (2008) and the scheme used to correct the SN data applied by Riess et al. (2007) is described in de Jong (2006).

The correction model is parameterized by a power law such that count rate $\propto$ flux$^{\alpha_N}$ with measured values (for NICMOS chip 2) of $\alpha_N = 1.025 \pm 0.002$ for the F110W filter and $\alpha_N = 1.012 \pm 0.006$ for F160W. Here the flux should include both the object and background, which is important for faint sources such as high-z SN. It is interesting that this effect seems to be wavelength dependent. Note that this correction is not based on a physical model, and so may not be the most appropriate parameterization. Furthermore, this nonlinearity has only been tested down to count rates of a few per second, while $z > 1$ SN plus background observations are typically an order of magnitude fainter.

The count rate non-linearity affects the observations of bright standard stars differently than faint SNe, which means that uncertainty in each α_N affects how the NICMOS calibration propagates to SN observations, and hence introduces a systematic uncertainty in the light-curve parameters. The NICMOS zero points are based on observations of P330e, which is ~ 11.5 mag in F110W and F160W. The proper way to measure the impact of this effect would be to go back to the original SN and standard star images and re-extract the photometry for different values of α_N. However, we can obtain a decent estimate simply by converting the magnitudes of the standard stars and SNe to counts rates and comparing the flux ratio for different values of α_N.

Carrying out this program (and including the appropriate typical background levels for

the two passbands), the average uncertainty for observations in the F110W and F160W filters induced by the uncertainties in α_N are 0.022 and 0.060 mag, respectively, for *HST* SN with data in those filters; the exact values depend on the flux level of a particular photometry point, and so vary both from SN to SN and with epoch for a particular SN. Because the effects of the uncertainties in the α_N are correlated across all of the SNe observed with NICMOS, they have a quite substantial impact on the weight of these SNe in the cosmological fits. Furthermore, these uncertainties may be optimistic, since they assume the power law model is valid at such low flux levels. Because of the high background levels, these numbers are fairly constant as a function of epoch, do not change much for different SNe in the *HST* sample, nor are they strongly affected by the possible presence of a host galaxy.

Changes in α_N make SNe either simultaneously fainter and redder or bluer and brighter, so that correction for the brighter-bluer relation partially cancels the effects of this systematic term. However, the residual effect is still substantial. The effect on w is small because the *HST* SNe have relatively little weight in this fit, but this is not necessarily true of fits for other cosmological parameters such as time-varying w. We can roughly estimate the importance of this effect by calculating the change in the inverse variance of the mean corrected magnitude of the *HST* sample with and without this systematic – this is essentially the weight of the sample. This shows that the uncertainty in the nonlinearity correction reduces the weight of the *HST* sample by 17%, approximately equal to removing three SNe; reducing the uncertainties in α_N by a factor of two would be equivalent to adding two more SNe. Fortunately, only the four highest redshift SNe are strongly dependent on NICMOS F160W observations, where the nonlinearity correction is the most uncertain.

In addition to the uncertainty in the NICMOS nonlinearity corrections, we assign zero-point uncertainties of 2% for the ACS bands longward of 7000 Å following Bohlin (2007), and 1% below that. For NICMOS, we additionally include an uncertainty of 0.009 and 0.0145 mag for F110W and F160W following the information on the NICMOS web site. The above terms nominally include the uncertainty in the bandpasses, so we do not include separate terms for those.

5.1.4. SDSS Zero Points

The calibration adopted by the SDSS SN team is based on observations of solar analogs with CALSPEC *HST* measured SEDs. There are no Landolt magnitudes available for these stars, so we take a slightly less direct route by comparing their SEDs with that of BD 17° 4708 (which, fortunately, is calibrated onto the same system). For the uncertainty in each zero point, we adopt values given in §3.2 of K09, which are 0.4%, 0.4%, and 0.7% in

gri[6]. In addition to the individual filter terms, because this calibration is not linked directly to the Landolt system, the inter-calibration is subject to an additional "gray" uncertainty of 0.6% arising from the uncertainty in the V magnitude of BD 17° 4708 on the Landolt system, including transformational uncertainties. This additional term is the price paid for not calibrating the SDSS (and CSP) observations to the same system as the low-z SNe, although the net calibration uncertainty is still small compared to some other samples.

There are a small number of well-observed SNe in common between the SDSS and low-z samples, in particular two with the CSP and one from the CfAIII sample. Comparing the photometry as between the CSP and CfAIII samples, we find good agreement in gr but not in i, with mean offsets of 1.2, 0.5, and 2.1σ , respectively. It is even less clear how to interpret these numbers due to the very small sample size. Fortunately, there should be many more objects in common between years 2-3 of the CSP and SDSS samples, so it will be possible to carry out much more stringent comparisons in the future.

5.1.5. System Bandpasses: SNLS

We next turn to uncertainties in the SNLS system bandpasses. The four MegaCam filters used by SNLS for the first 3 years of operation were manufactured by SAGEM. The filters scans show clear evidence of variation with radial position in the array, which we account for in our analysis (R09 §7). The largest systematic effect on our measurements arises from shifts in the mean wavelengths of the bandpasses, since this changes which portion of the SN SED is sampled at different redshifts. Alternative prescriptions, such as adding white noise to the bandpasses (as in Wood-Vasey et al. 2007), have little effect on the cosmological parameters and therefore underestimate the importance of accurate bandpass measurements. The mean bandpass can shift both from measurement uncertainties, and from the effects of the dry environment on Mauna Kea on the interference films[7]. Therefore, we parameterize our bandpass uncertainties here and in §5.1.6 in terms of shifts in the mean wavelengths. Other effects, such as changing the width of the bandpasses, have a much smaller effect. There have been several filter scans made of the MegaCam filters. These are consistent, once the radial and angular dependence of the filters is taken into account.

Two of the SNLS fields overlap with the footprint of SDSS, so we can calculate accurate

---

[6]These do not include the effects of the uncertainty in the SDSS bandpasses on the conversion from SDSS to AB magnitudes which we include in §5.1.6

[7]Fortunately, our filters do not sit in a vacuum, and hence should not experience the large shifts seen in the SDSS camera.

linear transformations between the two systems as discussed in Appendix G of R09. We can then compare these transformations against the results of synthetic photometry to estimate the size of any shifts in the mean bandpasses. We perform synthetic photometry on a library of stellar SEDs and try to reproduce the slope of the relation, but not the offset. The latter is important because by ignoring the offset these results are independent of effects such as zero-point uncertainties.

There is some variation in the results depending on which spectral library is used; we considered Oke & Gunn (1983); Pickles (1998); Stritzinger et al. (2005). The uncertainties in the transformation coefficients are of roughly equal importance. For g_M we find a best fit shift of $3 \pm 4 \pm 4$ Å, where the first uncertainty derives from the spread of results from the different SED libraries, and the second from the uncertainty in the measured color term between SDSS and MegaCam. For r_M we derive a shift of $1 \pm 1.5 \pm 7$ Å, and for i_M $12 \pm 6 \pm 5$ Å. The SED libraries do not extend far enough to the red to allow us to test z_M. While the different SED libraries agree best in r_M, this bandpass is particularly sensitive to uncertainties in the color term. Taken together, the probability of finding these shifts or larger due to measurement uncertainties in the case of no shifts is 46% . We therefore interpret the results of the above test as being consistent with the SAGEM filter scans having the correct mean wavelengths.

These limits are conservative for several reasons. First, they do not distinguish between uncertainties in the SDSS bandpasses and the SNLS ones. Furthermore, the SAGEM filter scans are thought to be much more accurate than these uncertainties, and the other contributions to the system bandpass all vary fairly slowly with wavelength. We therefore roughly average the above tests and adopt uncertainties of 7 Å in $g_M r_M i_M$. For z_M, where we have not been able to carry out the above test, we assume a bandpass uncertainty of 15 Å by scaling up the value from the other filters to include variations in the CCD response, which set the red-side cutoff for this filter. The consequences are summarized in table 11. We also considered varying the width of the bandpasses, and the effects of using the mean air mass when computing our bandpasses instead of the air mass for a particular observation. We find little effect from either, so do not include them in our uncertainty budget. At an air mass of 2 the mean air mass effects are less than 0.001 in the width of the light-curve and 0.003 in color in the most extreme case. Since SNLS doesn't observe above air mass 1.6 in practice, this effect is completely negligible.

5.1.6. System Bandpasses: Other Samples

As described in Appendix 9, we have applied small shifts to the Bessell (1990) passbands to represent the Landolt system. There are uncertainties associated with these shifts, as

Table 11: Bandpass Systematics

Description	Uncertainty (Å)	w for Ω_m=0.27	Rel Area
Stat only	$\cdots$	-1.031 ± 0.058	1
MegaCam Bandpasses	$\cdots$	-1.017 ± 0.066	1.20
g_M only	7	-1.030 ± 0.058	1.01
r_M only	7	-1.037 ± 0.059	1.01
i_M only	7	-1.030 ± 0.059	1.02
z_M only	15	-1.032 ± 0.060	1.04
low-z Bandpasses	$\cdots$	-1.027 ± 0.059	1.04
CfAIII only	7	-1.031 ± 0.058	1.00
CSP only	$\cdots$	-1.031 ± 0.058	1.00
B only	12	-1.029 ± 0.059	1.02
V only	12	-1.028 ± 0.059	1.01
R only	12	-1.031 ± 0.058	1.00
SDSS Bandpasses	$\cdots$	-1.026 ± 0.059	1.02
g only	7	-1.030 ± 0.059	1.00
r only	16	-1.026 ± 0.058	1.00
i only	25	-1.030 ± 0.059	1.01

described in Appendix A, and here we include them in our uncertainty budget. We adopt mean wavelength uncertainties of 12 Å in the mean wavelength for B and V and 25 Å for R. In addition to the Landolt system, we also include terms representing the uncertainties in the CfAIII natural system passbands. As we did for the SNLS filters, we adopt uncertainties of 7 Å for these passbands, which again are probably conservative. The uncertainties in the mean wavelength of the CSP filters are described in Contreras et al. (2010). These are limited by the number of spectrophotometric standards available, and are probably much larger than the real uncertainties. Due to the way in which the calibration is constructed, the CSP bandpasses affect the zero points strongly, but because this is currently such a small sample, the cosmological effects are negligible.

For the SDSS bandpass uncertainties, we use the values given in table 1 of K09. These systematic effects have very little effect in fits where the SDSS SNe are included, although they do also affect the SDSS to AB conversion and therefore the effective zero points. The effects of these shifts on the final uncertainty budget are given in table 11.

5.1.7. The Colors of the Flux Reference on the Landolt System

As part of our analysis, we convert between magnitudes on the Landolt system and fluxes. This is necessary because we are comparing SNe at a range of redshifts, and so the same regions of the rest frame SN SEDs are sampled by different filters for different objects. Fortunately, our measurement is not sensitive to the absolute flux scale, so more precisely we convert the magnitudes to relative fluxes. A major influence in our decision to calibrate MegaCam to the Landolt system was the insensitivity to absolute calibration. Were we to, for example, calibrate our observations onto an AB system, we would have included the uncertainty in the absolute flux scales of both systems when comparing them. This is the case for the SDSS sample.

Contrary to some statements found in the literature, the Landolt system is not *defined* in terms of any particular magnitudes for any particular star (i.e., Vega), nor is the Johnson & Morgan (1953) system on which it is based. Instead, both are defined based on particular observations of multiple stars – Vega is one of the stars contributing to $U-V$ and $B-V$ for the Johnson & Morgan (1953) system. We are free to choose any star (or set of stars) with a known SED as long as we know its magnitudes in the Landolt system and on the Landolt calibrated MegaCam system. Ideally, we would be able to observe this star directly with MegaCam. We have not yet been able to fully realize this goal, and have instead settled on the star BD 17° 4708 as our fundamental flux standard. This star has a high-quality CALSPEC STIS SED (Bohlin & Gilliland 2004) and Landolt magnitudes from Landolt &

Uomoto (2007). Because it has colors similar to the "average" Landolt star we have high confidence that its magnitudes have been accurately transferred to the Landolt (1992) catalog system , which most nearby observations are calibrated onto. This is probably not true of the CALSPEC white-dwarfs also observed by Landolt & Uomoto (2007) – due to the steepness of their SEDs, their magnitudes are offset from the Landolt (1992) system by a non-negligible but not perfectly known amount, rendering them unsuitable as fundamental flux standards for the Landolt system. For a more extensive discussion of all of these issues, see §8 and S11 of R09. We use the2010 February update to the BD 17° 4708 SED available on the CALSPEC web site[8].

The uncertainties in the magnitudes of BD 17° 4708 in both the Landolt system and the SNLS measurements are a significant contribution to the final uncertainty budget, as seen in table 12. The largest effects arise from the magnitudes of BD 17° 4708 in the SNLS filters, which is partially due to the fact that BD 17° 4708 is a subdwarf, and hence may transform slightly differently between photometric systems than the average Landolt standard star. This transformational uncertainty is the most significant for the z_M filter, which is reflected in this term having the single largest effect on w.

Most previous SN analyses make use of Vega as their fundamental flux standard. This is a poor choice because the magnitudes of Vega on the Landolt system are in fact poorly known, and it is too bright to observe directly with almost any modern imager. The uncertainties in the Landolt magnitudes of BD 17° 4708 are the largest single identified systematic uncertainty in our current analysis, and they would be several times larger were we to use Vega instead. These uncertainties have been significantly underestimated in previous analyses that use Vega. If we were able to replace the nearby SN sample with one calibrated onto a system similar to the natural MegaCam one (e.g., the USNO or SDSS system), these uncertainties would be much smaller. Since BD 17° 4708 effectively defines the USNO/SDSS system, this should be practical in the future, and significantly reduce this contribution to the final uncertainty budget.

5.1.8. The SED of BD 17° 4708

In addition to the uncertainty in the magnitudes of BD 17° 4708, we also consider the uncertainties in the SED itself. These are essentially set by the quality of the reductions and the repeatability of STIS in Bohlin & Gilliland (2004). This is discussed in R09 (§12.8 and table D.3), and the cosmological effects are given in table 12.

---

[8]http://www.stsci.edu/hst/observatory/cdbs/calspec.html

Table 12: Flux standard (BD 17° 4708) uncertainties

Description	Uncertainty (mag)	w for Ω_m=0.27	Rel Area
Stat only	$\cdots$	-1.031 ± 0.058	1
Colors	$\cdots$	-1.075 ± 0.075	1.31
$U - V$ only	0.0026	-1.031 ± 0.058	1.00
$B - V$ only	0.0015	-1.030 ± 0.059	1.01
$R - V$ only	0.0011	-1.030 ± 0.058	1.00
$I - V$ only	0.0013	-1.031 ± 0.060	1.05
$g_M - V$ only	0.0022	-1.030 ± 0.059	1.01
$r_M - V$ only	0.0042	-1.006 ± 0.062	1.13
$i_M - V$ only	0.0022	-1.039 ± 0.062	1.06
$z_M - V$ only	0.0178	-1.094 ± 0.071	1.15
SED	$\cdots$	-1.027 ± 0.073	1.23

Note. — The $UBVRI$ values have a correlated effect on $g_M r_M i_M z_M$ via the linear transformation equations of R09, which are not included in the above numbers, but are included in the overall uncertainty budget.

5.2. Comparisons of Different Light-Curve Fitters

§5 of G10 discusses comparisons between different light-curve fitting packages, and shows that when SALT2 and SiFTO are trained using the same incidental settings (such as the Landolt filter functions, colors of BD 17° 4708, etc. that are discussed elsewhere in this paper), the resulting fits on the same data are very similar. We consider this to be a very encouraging result. However, differences remain, which should not be surprising given the very different approaches taken to train the two models. In essence, the SN data are not yet high-enough quality to clearly prefer a particular modeling approach. We therefore include the differences in the derived light-curve parameters between the two fitters for each SN as a systematic in our analysis. In addition, there is some uncertainty in how to parameterize the model for the intrinsic uncertainty in the color model. We also include the difference between two parameterization choices (the sigmoid and exppol models of §4.4 in G10) in our systematic uncertainty; the effects on the cosmological parameter uncertainties are given in table 13.

We do not carry out any comparisons with the other publicly available light-curve fitting package, MLCS2k2 (Jha, Riess, & Kirshner 2007), as discussed in §5 of G10. Unlike SALT2 and SiFTO, MLCS2k2 explicitly attempts to separate intrinsic and extrinsic SN colors from photometric data, and then assumes that the extrinsic colors arise purely from dust, and that the remaining intrinsic color not related to the shape of the light-curve does not affect SN luminosity. SiFTO and SALT2 do not make this distinction. The merits of the two approaches depend critically on how well this separation can be performed, and how well SN intrinsic color can be predicted by the light-curve shape. The former depends on accurate models of the distribution of extinction with redshift and how they combine with selection effects (Wood-Vasey et al. 2007).

If the two conditions are met, then MLCS, by incorporating additional information beyond SN photometry, may be able to give tighter statistical constraints on SN relative

Table 13: Light-Curve Fitter Systematics

Description	w for Ω_m=0.27	Rel Area
Stat only	-1.031 ± 0.058	1
Combined	-1.027 ± 0.059	1.02
SALT2 vs. SiFTO	-1.027 ± 0.059	1.01
Color Uncertainty Model	-1.030 ± 0.058	1.00

distances. A test of how well this procedure works is to check if the MLCS2k2 predictions of A_V and Δ (the MLCS2k2 light-curve shape parameter) correlate with residual from the Hubble diagram, since in principle the MLCS2k2 model should already apply these corrections based on training with low-z SNe. Taking the MLCS2k2 fits from K09, we find that the Hubble residuals exhibit 7σ (statistical uncertainties only) evidence for a linear relation with A_V and 6σ with Δ. This suggests either a problem with the MCLS2k2 training sample, or with the detailed extinction prior used. Additionally, the issues with observer-frame U-band data of low-z SNe are a serious problem for MLCS2k2 because it is currently not feasible to train it on high-z data; therefore, we would effectively have to abandon most of our data for SN at $z > 0.7$ where the rest-frame U is measured by r_M and i_M. While we fully expect that this situation will be addressed, especially in light of the improved low-z data sets that are becoming available, this means that comparing MLCS2k2 fits to SiFTO and SALT2 at the current time is not productive. Note that the SALT2 and SiFTO models are capable of reproducing the MLCS2k2 assumptions if SN colors are dominated by normal dust extinction. The fact that the SiFTO and SALT2 derived color laws do not match the Galactic color law, assumed in MLCS2k2, for any value of R_V suggests that this assumption is incorrect.

5.3. Peculiar Velocities, $z_{\rm cut}$, and the "Hubble Bubble"

Hui & Greene (2006) discuss methods to calculate the correlated uncertainty in SN luminosity distances due to peculiar velocities given a survey geometry and radial selection function. This method is applicable as long as the survey geometry does not contain a large number of holes. Unfortunately, for the current low-z sample, the survey geometries consist almost entirely of holes, so this formalism is not currently practical.

The flow model used in this analysis (§2.8) is parameterized by a quantity $\beta_I = \Omega_m^{0.55}/b_I$, where b_I is a "biasing" parameter for *IRAS*-selected galaxies. β_I should not be confused with the color-luminosity nuisance parameter for SNe, β. Pike & Hudson (2005) gives $\beta_I = 0.49 \pm 0.04$ (statistical uncertainties only). We adopt this value, but conservatively take the uncertainty in β_I to be five times larger, $\sigma_{\beta_I} = 0.2$ to reflect the spread in values derived from other surveys and to account for simplifications in the modeling, and therefore compare models with $\beta_I = 0.3, 0.5, 0.7$ to estimate the systematic effect on our sample. The results are given in table 7. This is not a major contributor to our uncertainty budget.

A related issue is the possibility of a monopole term in the local expansion – a so-called Hubble bubble. Recently, Jha, Riess, & Kirshner (2007) found evidence for such an effect using light-curve fits to nearby SNe Ia. As discussed in Conley et al. (2007), this is related

to the interpretation of SN colors. The result is that when the same data are analyzed in a framework in which the relation between SN color and luminosity is determined from SN data – as is used here – we see no evidence for a Hubble bubble, and therefore we do not include this in our analysis.

Still, we can investigate the dependence of our results on the minimum redshift. K09 study this issue with a smaller nearby sample (§9), and include the variation as a systematic uncertainty. We find a similar trend, as shown for w in figure 13. In order to determine if this trend is significant, or is caused by shot noise from removing SNe from the sample, we have carried out a set of Monte-Carlo simulations where we note the number of nearby SN removed by each redshift cut, and try randomly removing the same number from the sample and refitting the cosmological parameters. Neighboring points are extremely correlated because $z_{\rm cut} = 0.02$ will also remove all of the SNe that, for example, $z_{\rm cut} = 0.016$ does. Including these correlations in our simulation, we find no evidence for a significant trend (a χ^2 of 10.16 for 11 degrees of freedom). Therefore, we find that the observed effect is entirely consistent with shot-noise, and therefore is already included in our statistical uncertainty budget.

5.4. Malmquist Bias

As noted previously, we apply a redshift-dependent correction for Malmquist bias to each SN sample. Therefore, the uncertainty in this correction should be included in our total uncertainty budget. There are a large number of individual terms, although only a few have significant effects. The SNLS Malmquist bias simulations and their systematic uncertainties are described in Perrett et al. (2010). The systematic uncertainty on the Malmquist bias for SNLS is about 20% of the correction, with some redshift dependence. The dominant terms are the uncertainty in the input $\sigma_{\rm int}$ and the value of β used in the calculation.

For SDSS, we follow the prescription of §9 of K09 in varying the spectroscopic selection model. For the Calán/Tololo, CfAI, and CfAII samples, we assign 0.01 mag uncertainty in the mean Malmquist bias, while for the CfAIII sample, we double the estimated uncertainties for our model of the spectroscopic selection function ($\sigma_\gamma = 0.62$, see §2.7.1) in order to compensate for the crudity of our model. Within the redshift limits we have set, the *HST* sample suffers from very small amounts of Malmquist bias, so we do not include this as a systematic.

Some of the terms in the Malmquist bias corrections affect multiple samples simultaneously. The bias estimates for the SNLS, SDSS, and CfAIII samples depend on the assumed values of α, β, and $\mathcal{M}$. We assign uncertainties of 0.22, 0.22, and 0.05 mag to these, which are
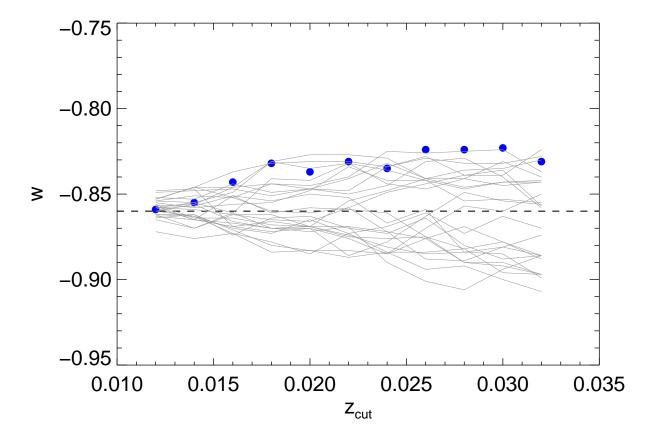
Fig. 13.— Variation of w with z_{cut}, the minimum redshift low-z SN included. The dashed line is the result when $z_{\mathrm{cut}} = 0.01$, the value used in our analysis. The points are highly correlated. Taking this into account via Monte-Carlo simulation, we find that this trend is not statistically significant, and is consistent with shot noise. The gray lines are a sample of 30 independent Monte-Carlos simulations illustrating the shot noise as discussed in the text.

twice the formal uncertainties (table 6), and affect all three samples in a correlated fashion. Each simulation also depends on σ_{int}, and this is varied separately for each sample by 0.01. In total there are 22 individual terms in our Malmquist uncertainty; the combined effect is given in table 7. Malmquist correction is one of the larger terms in our overall uncertainty budget, ranking only behind calibration and the relation between host type and SN absolute magnitudes. About 80% of the effect on the cosmological parameters comes from the low-z sample, which is not surprising because the selection effects of the higher-redshift samples are much better characterized.

5.5. Contamination by Other Types of SNe in the SNLS Sample

All of the SNe used in our cosmological analysis are spectroscopically typed, so contamination by non-SNe Ia should be fairly minimal. However, some is inevitable, particularly at the highest redshifts where the signal to noise of the spectra can be quite low. The principle contaminants are expected to be SN Ib/c, which are brighter than other core-collapse SNe and whose spectra can be difficult to distinguish from SN Ia at certain phases with low signal-to-noise. Unfortunately, we know relatively little about the demographics of this population, which makes contamination effects difficult to estimate. Non-Ia contamination is also discussed in Balland et al. (2009).

Our spectroscopic typing scheme is described in detail in Howell et al. (2005). In this scheme, SN candidates are typed into various classes. Those that concern this paper are SNIa, which represent objects which are essentially certain SNe Ia, and SNIa⋆, which probably are, but other types can't be conclusively ruled out. The false positive rate for SNIa is expected to be negligible, but it is difficult to estimate the rate for SNIa⋆. Approximately 20% of the SNLS sample presented here are SNIa⋆ (50 out of 242), with the fraction increasing with redshift.

It is tempting to compare the corrected peak magnitudes of the SNIa⋆ sample to the SNIa sample, and to use this comparison to estimate the systematic uncertainty on the cosmological parameters. However, this test is severely flawed because luminosity affects our ability to type spectra, so SNIa⋆ will be fainter even in the absence of contamination. Ideally, fainter SNe would be compensated for with longer spectroscopic exposure times, but in practice this was rarely realized at the 0.1-0.2 mag precision required due to weather, the amount of time remaining in the queue, and the fact that exposure times were estimated from a few pre-maximum points using preliminary, real-time photometry. Furthermore, the intrinsic luminosity of the SN is correlated with the relative amount of host-galaxy contamination in the spectra (the %-increase of the SN+host relative to the host), which is

not as easy to compensate for, especially in poorer seeing conditions. We do not have a good model for predicting the odds of a given SN candidate getting an inconclusive spectroscopic type, and attempts to develop one based on the observed demographics of the SNIa versus SNIa⋆ were somewhat circular, naturally explaining away any observed difference in the apparent luminosity. Note that this implies that simply excluding all SNIa⋆ from our analysis would bias our measurement in a fashion we would not know how to correct for very well.

Therefore, we adopt a simple model to estimate the contamination bias. The critical factors in such a model are the number of contaminants that can be confused with SN Ia and the mean difference in magnitude between the contaminants and SNe Ia at that redshift. There are a few aspects of SNLS that help reduce the contamination. First, we have high-quality spectroscopic redshifts for all objects considered in these fits. Second, the majority of our SN have firm spectroscopic types. Combined, these two features mean that the outlier rejection we apply (described in §2.1) is highly effective at removing any contaminants with a different magnitude distribution than SN Ia – and if the magnitude distributions were identical (after the shape and color corrections are applied), there would be no bias even if the contamination rate were high. In addition, we reject SN that are very poorly fit by the light-curve template, and preselect candidates for spectroscopic follow up based on early-time light-curve points (Sullivan et al. 2006a). Finally, recall that SNIa⋆ are probably SNe Ia – spectra which are truly uninformative are not used in this analysis. While the final three items have considerable power at rejecting candidates, for our simple model we will conservatively assume that they are completely ineffective, significantly increasing the bias.

The basic ingredients for our model are the relative rate of SNe Ia and SNe Ib/c and the SN Ib/c luminosity distribution. We assume that the SN Ib/c rate is proportional to the star-formation rate, which we base on Hopkins & Beacom (2006); this is a good approximation because the progenitor systems are massive stars, and the delay time between formation and explosion should be negligible for our purposes. We take the SN Ia rate from K. Perrett et al. 2011, in preparation. The most critical, but most uncertain, ingredient is the assumed luminosity function of SN Ib/c. Recently, Li et al. (2010) have studied the luminosity function of various SN types using the galaxy targeted KAIT survey. They find that SNe Ib/c are, on average, 2.4 mag fainter than SNe Ia, and that the SN Ib/c rate is about 80% of the SN Ia rate at low redshift. If we take their luminosity function at face value, the resulting bias is utterly negligible because they essentially find no bright SN Ib/c that could be confused with SN Ia; that is, all SNe from their sample would be rejected as outliers. In contrast, Richardson et al. (2002) suggest the existence a bright component of the SN Ib/c population which would be more easily confused with SN Ia. It is difficult to be sure if this disagreement is significant given the sample sizes, but we note that the KAIT result is based on a galaxy targed survey which avoids low-luminosity host galaxies, so if there is any relation between

SN Ib/c luminosity and host type, as there is for SN Ia, it could miss this portion of the sample. We will therefore conservatively assume that there is an additional bright SN Ib/c population missed by KAIT. Combined with the KAIT sample, the bright sample is $< 15\%$ of the total SN Ib/c population. The magnitude distribution of core-collapse SN has also been investigated photometrically by Bazin et al. (2009), although this paper includes other types of core-collapse SNe which are easier to distinguish from SNe Ia.

The method is then to generate a simulated sample of a few hundred thousand SN of both types, include observational uncertainties representative of SNLS, apply the outlier rejection, assume that all SN of any type which survive the outlier rejection are included in the sample, and finally compute the difference between the mean magnitude and the assumed cosmology in redshift bins as the bias. This is then multiplied by the actual fraction of SNIa$\star$ in the corresponding real sample to estimate the contamination bias (really, a conservative upper limit on the bias).

To describe this bright population, we use a Gaussian distribution characterized by its mean offset from the SN Ia population and its width, Δ_{bc} and σ_{bc}. Richardson et al. (2002) estimates $\Delta_{\mathrm{bc}} = 0.7$ and $\sigma_{\mathrm{bc}} = 0.33$ mag, after color correction. However, these parameters are quite uncertain. The amount of bias is maximized for $\Delta_{\mathrm{bc}} = \sigma_{\mathrm{bc}}$ mag (for smaller values of Δ_{bc} the amount of contamination increases, but the bias decreases, and for larger values the contamination rapidly falls), increases with the fraction of bright SN Ib/c, and weakly increases as σ_{bc} decreases. We therefore tune these parameters to maximize the bias while remaining vaguely plausible, adopting $\Delta_{\mathrm{bc}} = 0.25$, $\sigma_{\mathrm{bc}} = 0.25$, and also conservatively assume that the bright component is 25% of the total SN Ib/c population. The results, including the effects of multiplying by the actual fraction of SNIa$\star$ in each bin, are given in table 14. The effects on the cosmology are smaller than we can measure. The bias increases with redshift, both because the larger observational uncertainties make it harder for sigma clipping to reject contaminants, and because the fraction of SNe with SNIa$\star$ increases. Because our outlier rejection is based on the cosmological fit, the fact that the SNIa$\star$ fraction is large in the highest redshift bin does not affect this process, since these SNe have little influence on the cosmological models considered here. However, if this data set is used to investigate models where the Hubble relation changes sharply above $z = 0.9$, these results should be treated with some caution.

The predicted contamination is 0% at low redshift, rising to $\sim 10\%$ above $z = 0.9$. However, this model is deliberately conservative, and does not match the observed properties of the SNLS data very well. For example, it predicts that $> 60\%$ of the candidates sent for spectroscopic typing would either be identified as SN Ib/c or SNIa$\star$, while we actually find $< 25\%$, and that is only because we deliberately targeted some objects not expected to be

SN Ia. It also predicts that $> 20\%$ of the candidates sent for spectroscopy would be identified as bright SN Ib/c, whereas there are only a handful in reality.

While we find that for the SNLS3 sample the effects are small, this may not be true for future surveys that will predominantly rely on photometric rather than spectroscopic typing. In this case, where contamination may be large, the effects of sigma clipping are no longer as simple or as beneficial, since the mean magnitude in a bin may be severely affected when the types of most the SNe are potentially uncertain. This becomes even more complicated if only photometric redshifts are available. Such efforts will require a more precise understanding of the properties and demographics of SNe Ib/c if they are to compete with spectroscopically typed surveys.

5.6.   Milky-Way Extinction Correction

We correct all SNe for Milky-Way extinction. In addition to the random uncertainty in the $E(B - V)$ values, there is a correlated uncertainty in the conversion from dust column density to extinction of 10%, as discussed in Schlegel, Finkbeiner, & Davis (1998). This affects the nearby and distant SNe differently because distant SN searches generally target regions of low galactic extinction and observe at longer rest-frame wavelengths. The overall effect is given in table 7, and is one of the larger contributions to our final uncertainty budget.

5.7.   Evolution

The general lack of firm theoretical predictions makes it difficult to quantify the effects of potential SN evolution. Evolution in the absolute magnitude of SNe Ia with redshift is not constrainable without a detailed physical model because it can in principle mimic any cosmology. Here we consider only models that we can put some constraints on, in particular the possibility of evolution in the stretch-luminosity and color-luminosity relations with redshift. There is considerable evolution in the demographics of the SN population between $z = 0$ and $z = 1$ (Howell et al. 2007), but this does not bias the results as long as the demographic evolution is in parameters that are corrected for. If there are any differences in the width- and color-luminosity corrections between different sub-groups, we would expect there to be some evolution in the mean values of α and β with redshift. In particular, a change in the relative mixture of SN intrinsic colors and external effects such as dust would affect β.

Recently K09 (§10.2.3) have presented evidence for a strong decrease in β with redshift

Table 14: Bias from non-Ia contamination

z	Raw Bias (mag)	SNIa⋆%	Effective Bias[a] (mag)
0.1	0.015	0%	0.0
0.26	0.024	6%	0.001
0.41	0.024	14%	0.003
0.57	0.024	17%	0.004
0.72	0.023	24%	0.006
0.89	0.026	50%	0.013
1.04	0.025	NA	NA

[a]Includes the actual fraction of SNIa⋆ in each bin. There are no SNLS SNe above $z = 1.04$.

Table 15: Evolutionary Systematics

Description	dX/dz	w for Ω_m=0.27	Rel Area
Stat only	$\cdots$	-1.031 ± 0.058	1
Combined	$\cdots$	$-1.028^{+0.059}_{-0.058}$	1.02
α evolution	0.07	-1.030 ± 0.058	1.00
β evolution	1.0	$-1.028^{+0.059}_{-0.058}$	1.02

using SALT2 (although not the version used here – see Appendix A of G10 for details). They find that β decreases from ~ 2.5 at $z = 0$ to 1 by $z = 0.7$, but this evolution is only seen in ESSENCE and first-year SNLS samples. They also carry out MLCS2k2 fits, but do not analyze them in the same fashion. The relation between light-curve shape, color, and distance modulus for MLCS2k2 has already been discussed in §5.2. If we extend this model to allow for evolution in the width- and color-corrections with redshift, we find $d\alpha/dz = 0.25 \pm 0.07$ and $d\beta/dz = -1.86 \pm 0.43$, where the modified distance modulus is $\mu^\star = \mu + \alpha\Delta - \beta A_V$. Hence, for the K09 sample, both SALT2 and MLCS2k2 show strong evidence for evolving β with redshift in the SDSS analysis, with MLCS2k2 showing more evolution.

As shown in §5.5 of G10, we find much weaker evidence for such a trend in our data with SiFTO and an updated version of SALT2. This is probably related to the more sophisticated color-uncertainty modeling compared with that used in K09. Here we extend this analysis to include external data sets and the combined SiFTO/SALT2 distances (figure 14). We find $d\alpha/dz = 0.021 \pm 0.07$ and $d\beta/dz = 0.588 \pm 0.40$, marginal evidence of evolution. Also note that the β evolution has the opposite sign than the SDSS results, increasing slightly with redshift. Furthermore, SALT2 and SiFTO show small and opposite trends, with SiFTO increasing while SALT2 decreases.

It is unlikely that the observed evolution in β is real. §5.5 of G10 presents several lines of evidence to support this: the opposite trends for SALT2 and SiFTO, the fact that artificially adding selection effects at a different redshift moves the evolution to that redshift, and the dependence on the assumed parameterization of uncertainties in the light-curve intrinsic color model. However, while we can construct a plausible model that explains these trends, we cannot completely rule out α, β evolution. Therefore, the conservative approach is to treat this effect as real and include it in our systematic uncertainty budget. Since we find no evidence for α evolution, we adopt the uncertainty in the slope as our systematic. As in G10, for β we conservatively adopt $d\beta/dz = 1.0$. The effects of doing so are given in table 15.

Based on a model of inter-galactic dust, Ménard, Kilbinger, & Scranton (2010) predict a slow change in mean SN color with redshift without affecting the absolute magnitudes that appears as a change in β over a long enough redshift interval. Unfortunately, over a small redshift range, such as one of the redshift bins we use above, their model manifests as a change in the offset in the color-luminosity relation but does not change the slope (β), and therefore our test is not sensitive to their model. However, the $d\beta/dz$ value assumed in our systematic is larger than that predicted by their model for our data, so it is effectively included in our uncertainty budget.
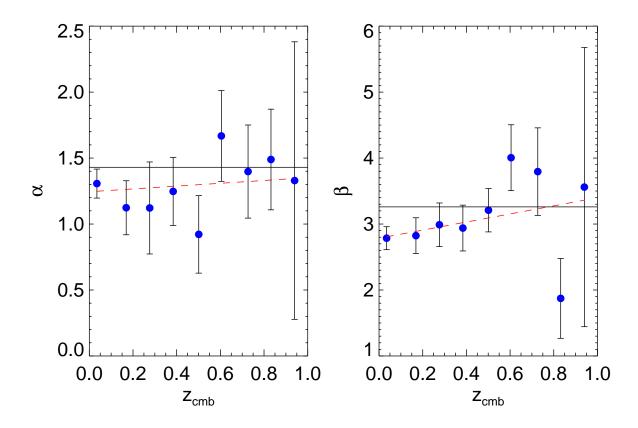
Fig. 14.— Evolution of α and β with redshift in the SiFTO/SALT2 combined light-curve fits. The solid lines are the best-fit values assuming no evolution, including systematic effects, and simultaneously fitting the cosmological parameters, while the point in each redshift bin is evaluated with the same fixed cosmology. The dashed line is the best linear fit to the points.

5.8. Environmental Dependence of SN Properties

As discussed in §3.2, in our analysis we correct for the recently discovered relationship between host-galaxy properties and peak magnitude by splitting the sample on host-galaxy mass and allowing $\mathcal{M}$ to be different for the two samples. Because we do not have a physical model for this phenomenon, the cut point is not well understood. Furthermore, we are not certain if mass is the best parameter, or if metallicity or star-formation rate are more relevant. Therefore, we consider systematic effects related to this split. Sullivan et al. (2010a) investigate the effects of changing the split mass, and find that the effects on the cosmological parameters are largest between 10^9 and $10^{11} M_\odot$. We therefore estimate this systematic by comparing which SNe change subsamples as we move the cut to these extremes from our default cut at $10^{10} M_\odot$.

In addition, we consider changing variables to metallicity, and similarly compute which SNe change subsamples. Clearly it is unreasonable to assume that both the appropriate cut differs from our nominal mass cut and that mass is not the right variable, so we equally weight each term (changing the cut and the two other host-variables) so that the total weight is 1. The final effects are given in table 7, and are one of the larger contributions to our final uncertainty budget (much less than calibration, however). If we do *not* apply this correction, as in all previous analyses, but apply the difference purely as a systematic uncertainty, the effect is approximately the same size as all other uncertainties (statistical and systematic) combined.

5.9. Gravitational Lensing

Gravitational lensing is expected to cause increased dispersion in the Hubble diagram of high redshift SNe. While the mean amount of magnification is unity at all redshifts, when there are a small number of SNe in each redshift bin, the asymmetric nature of the lensing probability (there is a long tail extending to high magnifications, but a lower limit on the de-magnification), coupled with selection effects, can produce biases in the peak luminosities. For surveys over very small areas, lensing will also induce correlations between different SNe. These issues are studied in Holz & Linder (2005), who find that for surveys such as SNLS, the number of SNe in each redshift range and the survey area are large enough that these issues are minor. Lensing does induce additional, almost uncorrelated, scatter in the peak magnitudes, which we already include in our statistical uncertainty budget using the model of Jönsson et al. (2010), which gives $\sigma = 0.055z$. This term is largely degenerate with σ_{int}, but we keep track of it separately because it has a redshift dependence while σ_{int} does not. Substituting a different model for the intrinsic scatter, such as Holz & Linder (2005)

$(\sigma = 0.088z)$ has almost no effect because $\sigma_{\rm int}$ changes to compensate. Therefore, besides including the statistical term, we do not include any additional lensing systematics in our uncertainty budget.

6. COMPARISON OF UNCERTAINTY BUDGET WITH OTHER ANALYSES

It is somewhat difficult to compare our systematic uncertainties directly with those of previous analyses because, in general, they do not take into account that systematic uncertainties will usually shift the measured cosmological parameters, which in turn affects the marginalized uncertainties on each parameter. For example, the potential α, β evolution actually shrinks the marginalized uncertainties slightly for fixed Ω_m, although the area of the Ω_m, w contour increases. We shall first present two examples of the classic form of systematics analysis discussed in §4.1. Note that we do not restrict ourselves to the same data samples as the other papers in these comparisons. As noted previously, all previous papers have not included the effects of systematic uncertainties on the training of light-curve fitters. The principal effect of including these effects is to increase the importance of any systematic uncertainty that affects SN colors.

The systematic contributions from the first year SNLS analysis are given in table 6 of A06 in terms of the effects on each cosmological parameter while holding all other parameters fixed at their best fit value. We can use the same technique to compare our results. For fixed Ω_m, we find $\sigma_w = 0.067$ (sys), which is in reasonable agreement with the 0.07 value estimated by A06. Examining individual terms in detail, the fact that the values are similar in fact represents considerable progress, as the A06 values were somewhat optimistic, particularly regarding calibration. We include a large number of terms not considered in the first year analysis, although most of these are sub-dominant. Our Malmquist bias uncertainty is much smaller, as benefits the much more detailed simulations carried out for the 3-season sample. Many of the individual uncertainty terms have gone down, such as the SNLS zero-point uncertainty which has shrunk from $\sigma_w = 0.05$ to 0.02. However, A06 significantly underestimated the uncertainties related to the flux standard, giving $\sigma_w = 0.03$ while we have 0.045. This is particularly significant because the A06 analysis was based on Vega, which has far less certain magnitudes on the Landolt and MegaCam systems than the standard adopted here, BD 17° 4708.

Wood-Vasey et al. (2007) discuss the systematics of the 3 year ESSENCE sample. Hicken et al. (2009b) uses a very similar analysis and obtain similar results. They find a total systematic uncertainty in w for ESSENCE of 0.13, compared with 0.11 here. However, their

value is combined with the BAO constraint of Eisenstein et al. (2005), which reduces the importance of systematic effects that move the SN contours along their longest axis. When this combination is carried out, the equivalent value for SNLS3 is ~ 0.06, about a factor of 2 smaller – see S11 for further discussion. This reduction is despite the fact that we again include many additional systematic terms not considered by these authors, such as the magnitudes of the flux standard, the low-z zero points and filters, shifts in the mean passbands of our survey, Malmquist uncertainties of the nearby sample, etc. The major difference arises from two terms: first, Wood-Vasey et al. (2007) includes a systematic related to the "Hubble bubble" discussed in §5.3 which we find no evidence for, and second, they have significant systematic effects related to the prior on extrinsic SN colors used in their MLC2k2 fits, which does not apply to SiFTO or SALT2. Without these two terms, their systematic uncertainty on w (with the BAO measurement) is 0.054, similar to our value.

As discussed previously, MLCS2k2 trades sensitivity to an assumed prior on the distribution of extinction for reduced statistical uncertainties. A major portion of the Wood-Vasey et al. (2007) systematic uncertainty budget arises from the uncertainty in this prior, and the fact that survey selection effects cause it to evolve strongly with redshift. This tradeoff makes comparing our systematic uncertainty budgets difficult. SALT2 and SiFTO do not use such priors, and Monte-Carlo studies of our selection biases indicate that these have a very small effect on our color modeling.

We next consider the analysis of Kowalski et al. (2008). As noted in §4.1, these authors use a more sophisticated framework for including systematic uncertainties than the classic quadrature method. Their method is a good approximation to the one used here when there is no structure in the covariances within a given data set, and no significant redshift-dependent correlations between datasets. As shown in the right panel of figure 12, the first is a reasonable assumption in some cases, but not in others, particularly for surveys which cover a large redshift range such as SNLS. The second assumption neglects the highly redshift-dependent correlations caused by calibration uncertainties related to the fundamental flux standard, which are not treated. One advantage of Kowalski et al. (2008) is that, unlike this paper, it is a blind analysis, which helps limit observer bias. Combining SN data with the BAO measurement as above, they find a systematic uncertainty in w of 0.085, significantly larger than our value of 0.067). This difference mostly arises from the extremely heterogeneous nature of their high-z SN sample and the large uncertainties in relative zero point assigned between them.

The analysis of Kowalski et al. (2008) was extended to a larger sample and refined to allow for more subtle redshift dependence within each sample in Amanullah et al. (2010). In both cases, the authors fix α and β at their best fit values when computing systematic

uncertainties, which will underestimate the final uncertainties because of the correlation between the cosmological and nuisance parameters, as discussed in §4.6. Their calibration is based on Vega rather than BD 17° 4708, adopting a magnitude uncertainty of 0.01 mag in each of 6 bands. As discussed in §10 of R09, the uncertainties in the magnitude of Vega on the Landolt system are much larger, especially in $V - R$, so it seems likely that the effects of calibration on their uncertainty budget are underestimated by at least a factor of 2 to 3. In general, the systematic uncertainty as a function of redshift of this paper has a much simpler structure than that presented here, although it is still a large improvement over most treatments.

K09 is probably the most extensive treatment of systematic uncertainties in the literature, although the authors use the classical quadrature method. They include many of the terms considered here, although several are omitted (e.g., bandpass uncertainties for SNLS, correlated peculiar velocities, the NICMOS nonlinearity). They present results for both MLCS2k2 and SALT2 (although not the improved version used here), but do not explicitly include the comparison in their uncertainty budget. Their largest source of uncertainty is due to the rest-frame low-z U band, while we exclude this data entirely based on the evidence presented in §2.3. They also find evidence for a large variation of β with redshift, but only include this term for SALT2. Our analysis shows that this is present in their data for both fitters, and more strongly for MLCS2k2. As shown in §5.7, our analysis does not support evolution in β. Finally, they also include a large uncertainty for the minimum z_{cut}, which we have shown is consistent with shot noise and therefore implicitly included in our statistical uncertainty. If we exclude these three terms, which we do not find evidence for, the total systematic uncertainty budget for K09 is about 0.07 in w when combined with non-SN constraints, similar to the value for our sample.

7. HOW THESE SYSTEMATICS WILL BE REDUCED

The dominant systematic uncertainties in our analysis are due to calibration, specifically the colors of BD 17° 4708. Other important terms include the relationship between SN peak magnitudes and host galaxy properties and corrections for Malmquist bias. A careful reading of R09 will reveal that many of the important calibration systematic effects are related to the necessity of cross-calibrating SNLS data to the current low-z SN sample on the Landolt system. We expect that a number of significantly improved low- and intermediate-z samples should become available in the next few years (a partial listing includes the the SNfactory sample, the KAIT sample, and the other 2 years of SDSS data), and so it is worth examining exactly what aspects of these samples will help us reduce these uncertainties.

The uncertainties related to the color of BD 17° 4708 include large terms related to how well the Landolt magnitudes of this star can be transferred to the MegaCam system; see §10.2 of R09. Because it is a slightly unusual star (a subdwarf), and because the MegaCam bandpasses are very unlike the Landolt ones, these uncertainties are large. Furthermore, the effective bandpasses of the Landolt system are simply not well understood, and likely never will be, putting a fundamental limit on how well we can characterize this uncertainty. Were the low-z sample replaced with one observed on a better understood system more similar to the MegaCam one, such as the USNO or SDSS systems, these transfer effects would be reduced substantially reducing our calibration uncertainties.

Improved lower-z samples may allow us to understand various degeneracies between intrinsic and extrinsic SN color, possibly by studying SNe in a variety of environments or incorporating near-IR observations. Such samples may also lead to better luminosity predictors beyond stretch and color, such as the spectral flux ratio feature of Bailey et al. (2009). Either possibility could lead to qualitative improvements and not simply quantitative ones in SN models.

The differences between light-curve models will also be reduced with better training samples, or else it will become clear which model is a good description of SN properties. While including high-z SN data in the training has been very useful, especially in the near-UV, improved low-z samples can improve the situation even further. A particular weakness of the current high-z data is the lack of multiple epochs of spectroscopy at different phases of the light-curve, especially in the near-UV. This deficiency can be remedied with STIS or COS spectroscopy of nearby SNe. The low-z sample zero points can obviously be improved by better nearby samples. The effects on the MegaCam zero points are more subtle, but if SNLS is calibrated directly against a similar system, they will be also be improved.

Since so many of these considerations relate directly to the calibration of the nearby sample, we expect to obtain the best results with low-z SNe calibrated to an observing system that is closer to the MegaCam one. Therefore, we expect the best results in the near term will be obtained when SNLS is properly combined with the SDSS SN sample and the CSP samples, both of which are approximate implementations of the USNO system. In particular, due to the large number of faint standard stars that have been observed with the SDSS telescope (Ivezić et al. 2007), the inter-calibration of SNLS and SDSS should be extremely powerful.

We also hope that the CALSPEC calibration can be refined and extended to more stars, particularly faint ones that can be observed directly by survey telescopes (BD 17° 4708 is too bright for both MegaCam and SDSS), and to stars of normal spectral types in the color range well sampled by standard star catalogs, which will significantly improve the reliability

of the transformations between natural systems. This program should be carried out while the natural USNO, and MegaCam systems still exist (although it is too late for the i_M filter used in this paper, which was destroyed in an unfortunate filter jukebox malfunction, and for the SDSS imager, which has been decommissioned, but there were many stars already observed by both); a major problem with the Landolt system is that the original system no longer exists, so there will always be limitations as to how accurately any given flux standard can be tied to this system. This finding has implications for future absolute calibration programs. From the standpoint of SN observations, how well the calibrated flux standard can be tied to the magnitude system in use is currently the limiting factor rather than how well the SED itself is calibrated, so good quality spectroscopic observations of fainter (and hence directly observable) standards are vastly preferable to improved observations of very bright standards such as Vega.

For $z > 0.7$ SNe, SNLS essentially observes in rest-frame U and B, and then relies on the relationship between $U - B$ and $B - V$ to compare with lower-redshift SNe. A survey which measures rest-frame $B - V$ out to higher redshifts using the near-IR should be able to reduce the importance of some of the calibration systematic effects discussed here for high-z SNe. This would be particularly useful for testing time-varying w models. However, such a survey then becomes sensitive to the problems of inter-calibrating near-IR and optical data, which are non-trivial.

8. CONCLUSIONS

We have presented measurements of the cosmological parameters from the first 3 years of the SNLS, combining them with external SN data sets at higher and lower redshift and including a correction for host-galaxy mass. This is the largest, highest-quality moderate- to high-redshift SN Ia sample to date. From SN data only, we find $w = -0.91^{+0.16}_{-0.20}\,(\text{stat})^{+0.07}_{-0.14}\,(\text{sys})$, consistent with a cosmological constant. The data requires an accelerating universe at $> 99.999\%$ probability including all identified systematics. The combination with non-SN constraints is given in S11, and yields considerable additional precision as well as allowing us to put constraints on time variations in w.

We model systematic uncertainties using a covariance matrix, including their effects on our empirical SN models. This approach should allow these uncertainties to be properly included in future analyses[9]. For SNe only, our current uncertainty budget is statistics dom-

---

[9] The statistical and systematic covariance matrices used in this paper, as well as the light-curve parameters, are available at `https://tspace.library.utoronto.ca/snls`

inated; when external, non-SN constraints are included, which we simulate by fixing Ω_m, statistical and systematic uncertainties are comparable. The dominant identified systematic uncertainties are related to calibration. This limitation can best be addressed with improved nearby SN data sets, particularly if they are calibrated onto more modern, better understood photometric systems such as USNO/SDSS. In this analysis, we have included external SNe samples calibrated onto both the Landolt and USNO-like systems without taking full advantage of the inter-calibration possibilities. Future papers in collaboration with the SDSS and CSP teams will address this point. The next most important systematic term, the relation between host mass and SN peak magnitudes can also be addressed with improved nearby samples to ascertain whether the peak magnitude subsample effect correlates best with metallicity or some other variable. Solid-angle, rather than galaxy-targeted searches, will be the most useful for this purpose because they will sample the underlying host galaxy population in a less biased fashion.

In agreement with K09, we find that there are significant problems with current low-z observer-frame U-band observations, although we are unsure where the issue lies; we have therefore excluded these observations in our fits. Interestingly, u' observations of low-z SNe, which are calibrated to the USNO instead of the Landolt system, but probe the same wavelength range, do not show any problematical behavior. We also note that the sign of the effect we observe is opposite that predicted by the spectral studies of Cooke et al. (2010); Foley et al. (2010).

In general, the systematic uncertainties of the SNLS3 sample are smaller than those of previous analyses, even after correcting for the fact that we do not find evidence for a number of pathologies found by other authors, such as β evolution of the Hubble bubble, and therefore do not include them in our analysis. The improvement in the final uncertainty budget, despite the many additional effects we consider, represents a significant advance, and makes clear the benefits of large, homogenous SN samples.

9. APPENDIX A: PASSBANDS FOR THE LOW-Z SAMPLE

Determining the natural system response of the Landolt system, which is the system of much of the low-z data, is non-trivial. Fortunately, this affects less than half of the nearby sample, and has no effect on the SNLS, SDSS, or HST samples except indirectly through the light-curve training[10]. There have been many attempts to reconstruct the Landolt passbands,

---

[10]Although we report the peak B band magnitude, because we work in SED space for both fitters, we are free to define this filter to be anything we want. Therefore, any uncertainty on the Landolt filters only

which are nominally a realization of the Johnson-Morgan and Cousins systems. These are natural systems, so it useful to attempt to reconstruct them. The standard work on this subject is Bessell (1990), which has been used in many previous SN cosmological analyses. Notable exceptions are A06, who used these filters responses but shifted them by 41, 27, 21 and 25 Å blueward (for $BVRI$, respectively), and K09 (Appendix B) who use shifts of 15, -12, -7, 45 Å (blueward)[11]. The former was based on a comparison of synthetic photometry of Landolt standard stars from Hamuy et al. (1992, 1994) to their actual magnitudes, and the latter on CALSPEC $STIS$ standards using magnitudes from Landolt & Uomoto (2007).

We have carried out a similar analysis to that of A06 for this paper, but using a more recent spectrophotometric library of Landolt stars (Stritzinger et al. 2005). The CALSPEC library is attractive, but unfortunately, as discussed in §10.1 of R09, the magnitudes of these stars from Landolt & Uomoto (2007) are probably not on the Landolt system to the required accuracy because the linear transformations derived from normal stars were not appropriate to the extremely blue CALSPEC white dwarfs. The Stritzinger et al. (2005) compilation is essentially an extension of that of Hamuy et al. (1992, 1994) to a much larger data set by many of the same authors.

Using this sample, we find the best fits if we shift the Bessell $BVRI$ by 9, 3, 21, and 14 Å to the *red*. Using bootstrap with replacement, and folding in the uncertainty in the colors of the fundamental flux standard, the uncertainties are about 12 Å. Except possibly in R, we find shifts which are consistent with zero. Our results are in agreement with the analysis presented in Stritzinger et al. (2005) using the same data, and those of Maíz Apellániz (2006), who performed a similar test using a HST-based spectrophotometric library. Repeating this test with the Hamuy et al. (1992, 1994) libraries, we find general agreement with the values used in A06. Note that neither the Hamuy nor the Stritzinger libraries have enough wavelength coverage to enable this test for U.

A fraction of the Landolt standard stars also have magnitudes on the USNO system (Smith et al. 2002). Since the bandpasses of the USNO system have been determined much more accurately, we can use the spectrophotometry of these stars to test the Stritzinger et al. (2005) library. We find shifts of 4, -8, and 26 Å for $g'r'i'$, with uncertainties of about ± 16 Å. Except for i', this library does an excellent job reproducing the USNO magnitudes. We can't apply this test to the Hamuy et al. (1992, 1994) libraries because there is no overlap with

---

affects SNe whose observations are reported directly in that system.

[11]In fact, filter shifts were not used, but instead linear transforms were applied to bring Landolt system photometry onto the Bessell system. These are equivalent because the implied shifts are small. For larger shifts, using linear transforms is inadvisable.

the USNO sample. Again, we find marginal evidence for problems in the red, which may explain the shift found for R, since the Landolt/Cousins R filter has a much redder response than r'. This is particularly interesting because the synthetic photometry tests agree well with the other libraries in our tests of the MegaCam filters (§5.1.5). Fortunately, the R filter plays a very minor role in our analysis. We therefore adopt the best fit values as shifts applied to the Bessell filters in our analysis, and treat the uncertainties in §5.1.6.

In order to construct natural system bandpasses for the Hicken et al. (2009a) sample, which combines Landolt-like and USNO-like filters, we multiply the telescope/detector responses by an assumed atmospheric transmission. We were unable to obtain an atmospheric absorption curve specific to Mt. Graham, so substituted that of Kitt Peak, which is at a similar altitude and shares a similar climate. This substitution is included in the filter bandpass systematics discussed in §5.1.6.

10. Appendix B: COSMOLOGICAL FITTING TECHNIQUES

We consider two different techniques for fitting the cosmological parameters[12]. In the first, we calculate the relative probability of each value of the parameters over a grid and report the expectation value of the resulting distribution for each parameter (the marginalization approach). In the second, we attempt to find the value of the parameters that minimizes the χ^2(the minimization approach). In the first case, the uncertainties are derived by directly computing the bounds that contain the desired fraction of the total probability on each marginalized parameter. In the second, we make standard assumptions about the relation between the χ^2 and the parameters (namely, that the uncertainties are Gaussian and the model is approximately linear in the parameters over the uncertainties) to estimate the confidence limits. The Markov-Chain Monte-Carlo results now used in most CMB analyses are ideologically equivalent to the first approach.

The results reported from these two approaches will generally not agree because they do not have the same meaning mathematically. This does not imply that either approach is incorrect; while it would be comforting to be able to clearly choose one method as more desirable, reality is not so kind. Further discussion of the differences between these two approaches can be found in Upadhye, Ishak, & Steinhardt (2005).

In the grid based approach, the χ^2 is converted into a relative probability via $P \propto \exp\left(-1/2\,\chi^2\right)$. An evenly spaced grid is used, which is equivalent to assuming a flat prior on

---

[12]Both codes are available from `http://casa.colorado.edu/~aaconley/Software.html`

all parameters. Properly speaking, since $\mathbf{C}$ (the total covariance matrix of equation 3) is a function of α, β, we should have $P \propto \exp\left(-1/2\,\chi^2\right)/\sqrt{\det \mathbf{C}}$, where det is the determinant operator. However, including this term results in large negative biases on α and β that are worse for larger samples, whereas if it is omitted the biases are negligible for data sets similar to ours. See Kelly (2007) for more discussion of the det $\mathbf{C}$ factor.

A weakness of the χ^2 minimization approach is that the usual method of estimating uncertainties (by searching for the boundary where the χ^2 increases by a certain amount) depends on the assumption that the model is close to linear in the parameters over the size of the uncertainties. While this is true for the fits presented in this paper, it is not necessarily accurate for more poorly constrained parameters, such as the derivative of w, and can lead to significant underestimates of the uncertainties.

11.   APPENDIX C: ANALYTIC MARGINALIZATION OVER $\mathcal{M}$

Because the uncertainties of individual SN do not depend on $\mathcal{M}$, it is possible to remove this parameter from our fits by analytically marginalizing over it, following the technique described in Goliath et al. (2001). Here we trivially extend this treatment to the case of non-diagonal uncertainties and to the case of two values of $\mathcal{M}$ split by some additional variable (such as host galaxy stellar masses).

We marginalize by converting the χ^2 to a relative probability, integrating it over a prior π, and then converting back to a χ^2. For the case of a single $\mathcal{M}$:

$$\chi^2_{\mathcal{M}\,marg} = -2\log\left[\int_{-\infty}^{\infty} d\mathcal{M} \, \exp\left(-\frac{1}{2}\chi^2\right) \pi\left(\mathcal{M}\right)\right].$$

In our code we assume a flat prior on $\mathcal{M}$ (as we do implicitly on all other parameters). Following the discussion in §4.1, if we define the vector of residuals between the model magnitudes and the observed magnitudes $\Delta\vec{\mathbf{m}}$, but this time omitting $\mathcal{M}$ from m_{mod}, then

$$\chi^2 = \left(\Delta\vec{\mathbf{m}} - \mathcal{M}\vec{\mathbf{1}}\right)^T \cdot \mathbf{C}^{-1} \cdot \left(\Delta\vec{\mathbf{m}} - \mathcal{M}\vec{\mathbf{1}}\right)$$

where $\mathbf{C}^{-1}$ is the inverse covariance matrix. Carrying out the integral,

$$\chi^2_{\mathcal{M}\,\mathrm{marg}} = a + \log\frac{e}{2\pi} - \frac{b^2}{e}. \tag{7}$$

where $a \equiv \Delta\vec{\mathbf{m}}^T \cdot \mathbf{C}^{-1} \cdot \Delta\vec{\mathbf{m}}$, $b \equiv \Delta\vec{\mathbf{m}}^T \cdot \mathbf{C}^{-1} \cdot \vec{\mathbf{1}}$, and $e \equiv \vec{\mathbf{1}}^T \cdot \mathbf{C}^{-1} \cdot \vec{\mathbf{1}}$. Note that the determinant of $\mathbf{C}$ does not appear in these relations. We do not analytically marginalize over α and β

because the uncertainties of each SN depend on their values; such a procedure was carried out in A06 by holding their values fixed in $\mathbf{C}$, but this approach causes a significant bias in the recovered values. Note that a, b, and c all depend on parameters that we are fitting.

The formula is more complicated in the presence of two values of $\mathcal{M}$. If we define two vectors $\vec{K}_1$ and $\vec{K}_2$ so that $\vec{K}_1$ is one where the SN is in set 1 and 0 in set 2, and vice-versa for $\vec{K}_2$, then a remains the same, but the expressions for b and e are modified to $b \equiv \Delta\vec{\mathbf{m}}^T \cdot \mathbf{C}^{-1} \cdot \vec{K}_1$ and $e \equiv \vec{K}_1^T \cdot \mathbf{C}^{-1} \cdot \vec{K}_1$. Further introducing $c \equiv \Delta\vec{\mathbf{m}}^T \cdot \mathbf{C}^{-1} \cdot \vec{K}_2$, $d \equiv \vec{K}_1^T \cdot \mathbf{C}^{-1} \cdot \vec{K}_2$, $f \equiv \vec{K}_2^T \cdot \mathbf{C}^{-1} \cdot \vec{K}_2$, and $g \equiv ef - d^2$ then

$$\chi^2_{\mathcal{M}\,\mathrm{marg}} = a + \log\frac{e}{2\pi} + \log\frac{g}{2\pi e} - \frac{b^2 f}{g} - \frac{c^2 e}{g} + 2\frac{bcd}{g}. \tag{8}$$

Estimates for the two values of $\mathcal{M}$ are given by $\mathcal{M}_1 = (bf - cd)/g$ and $\mathcal{M}_2 = (ce - bd)/g$.

Often there is a desire to evaluate the likelihood of some cosmological parameters without fitting the nuisance parameters. One technique for doing this is to find the value of the nuisance parameters that minimize the χ^2 for a given set of cosmological parameters and then evaluate the likelihood by substituting those nuisance parameters. So, for example, the $\mathcal{M}$ that minimizes the χ^2 is $\mathcal{M} = b/e$. The resulting χ^2 on substituting this value into equation 3 is equal to that of equation 7 to within an offset that depends on α and β, but not $\mathcal{M}$. This trick is commonly used (e.g., in `CosmoMC`, Lewis & Bridle 2002), but unfortunately significantly biases the cosmological results because the χ^2 for different values of α and β can no longer be compared. Attempting to avoid fitting explicitly for α and β following this approach is also quite biased, because the uncertainties depend on the nuisance parameters. For the χ^2 minimization fits, the results are not biased, but the uncertainty estimates are incorrect, while for the marginalization fits both the results and the uncertainties are biased. Therefore it is important to explicitly fit α and β along with the cosmological parameters.

Facilities: CFHT (MegaCam), Gemini:Gillett (GMOS-N), Gemini:South (GMOS-S) Keck:I (LRIS), VLT:Antu (FORS2), VLT: Kueyen (FORS1)

REFERENCES

Aharmim, B. et al. 2005, Phys. Rev. C, 72, 055502

Albrecht, A. et al. 2006, ArXiv Astrophysics e-prints

Aldering, G. et al. 2006, ApJ, 650, 510

Altavilla, G. et al. 2004, MNRAS, 349, 1344

Amanullah, R. et al. 2010, ApJ, 716, 712

Astier, P. et al. 2006, A&A, 447, 31

Bailey, S. et al. 2009, A&A, 500, L17

Balland, C. et al. 2009, A&A, 507, 85

Barlow, R. 2003, ArXiv Physics e-prints

Bazin, G. et al. 2009, A&A, 499, 653

Bessell, M. S. 1990, PASP, 102, 1181

Blakeslee, J. P. et al. 2003, ApJ, 589, 693

Bohlin, R. C. 1996, AJ, 111, 1743

—. 2007, Photometric Calibration of the ACS CCD Cameras, Tech. rep.

Bohlin, R. C. & Gilliland, R. L. 2004, AJ, 128, 3053

Bronder, T. J. et al. 2008, A&A, 477, 717

Conley, A., Carlberg, R. G., Guy, J., Howell, D. A., Jha, S., Riess, A. G., & Sullivan, M. 2007, ApJ, 664, L13

Conley, A. et al. 2006a, ApJ, 644, 1

Conley, A. et al. 2006b, AJ, 132, 1707

Conley, A. et al. 2008, ApJ, 681, 482

Contreras, C. et al. 2010, AJ, 139, 519

Cooke, J. et al. 2010, ArXiv Astrophysics e-prints

de Jong, R. S. 2006, Correcting the NICMOS count-rate dependent non-linearity, Tech. rep.

de Jong, R. S. et al. 2006a, in The 2005 HST Calibration Workshop: Hubble After the Transition to Two-Gyro Mode, ed. A. M. Koekemoer, P. Goudfrooij, & L. L. Dressel, 121–+

de Jong, R. S., Bergeron, E., Riess, A., & Bohlin, R. 2006b, NICMOS count-rate dependent nonlinearity tests using flatfield lamps, Tech. rep.

Doi, M. et al. 2010, AJ, 139, 1628

Eisenstein, D. J. et al. 2005, ApJ, 633, 560

Ellis, R. S. et al. 2008, ApJ, 674, 51

Foley, R. J. et al. 2010, ArXiv Astrophysics e-prints

Fukugita, M., Ichikawa, T., Gunn, J. E., Doi, M., Shimasaku, K., & Schneider, D. P. 1996, AJ, 111, 1748

Goliath, M., Amanullah, R., Astier, P., Goobar, A., & Pain, R. 2001, A&A, 380, 6

Golub, G. H. & van Loan, C. F. 1996, Matrix computations (Matrix computations / Gene H. Golub, Charles F. Van Loan. Baltimore : Johns Hopkins University Press, 1996. (Johns Hopkins studies in the mathematical sciences))

Guy, J. et al. 2007, A&A, 466, 11

Guy, J. 2010, A&A, 523, 7

Hamuy, M. et al. 1993, AJ, 106, 2392

Hamuy, M. et al. 1996, AJ, 112, 2408

Hamuy, M., Suntzeff, N. B., Heathcote, S. R., Walker, A. R., Gigoux, P., & Phillips, M. M. 1994, PASP, 106, 566

Hamuy, M., Walker, A. R., Suntzeff, N. B., Gigoux, P., Heathcote, S. R., & Phillips, M. M. 1992, PASP, 104, 533

Hicken, M. et al. 2009a, ApJ, 700, 331

Hicken, M., Wood-Vasey, W. M., Blondin, S., Challis, P., Jha, S., Kelly, P. L., Rest, A., & Kirshner, R. P. 2009b, ApJ, 700, 1097

Holtzman, J. A. et al. 2008, AJ, 136, 2306

Holz, D. E. & Linder, E. V. 2005, ApJ, 631, 678

Hopkins, A. M. & Beacom, J. F. 2006, ApJ, 651, 142

Howell, D. A., Sullivan, M., Conley, A., & Carlberg, R. 2007, ApJ, 667, 37

Howell, D. A. et al. 2006, Nature, 443, 308

Howell, D. A. et al. 2005, ApJ, 634, 1190

Hudson, M. J., Smith, R. J., Lucey, J. R., & Branchini, E. 2004, MNRAS, 352, 61

Hui, L. & Greene, P. B. 2006, Phys. Rev. D, 73, 123526

Ivezić, Ž. et al. 2007, AJ, 134, 973

Jha, S. et al. 2006, AJ, 131, 527

Jha, S., Riess, A. G., & Kirshner, R. P. 2007, ApJ, 659, 122

Johnson, H. L. & Morgan, W. W. 1953, ApJ, 117, 313

Jönsson, J. et al. 2010, MNRAS, 405, 535

Kawabata, K. S. et al. 2010, Nature, 465, 326

Kelly, B. C. 2007, ApJ, 665, 1489

Kelly, P. L., Hicken, M., Burke, D. L., Mandel, K. S., & Kirshner, R. P. 2010, ApJ, 715, 743

Kessler, R. et al. 2009, ApJS, 185, 32

Knop, R. A. et al. 2003, ApJ, 598, 102

Komatsu, E. et al. 2010, ArXiv e-prints

Kowalski, M. et al. 2008, ApJ, 686, 749

Krisciunas, K. et al. 2007, AJ, 133, 58

Krisciunas, K., Hastings, N. C., Loomis, K., McMillan, R., Rest, A., Riess, A. G., & Stubbs, C. 2000, ApJ, 539, 658

Krisciunas, K. et al. 2001, AJ, 122, 1616

Krisciunas, K. et al. 2004a, AJ, 127, 1664

Krisciunas, K., Prieto, J. L., Garnavich, P. M., Riley, J.-L. G., Rest, A., Stubbs, C., & McMillan, R. 2006, AJ, 131, 1639

Krisciunas, K. et al. 2004b, AJ, 128, 3034

Lampeitl, H. et al. 2010, ApJ, 722, 566

Landolt, A. U. 1992, AJ, 104, 340

Landolt, A. U. & Uomoto, A. K. 2007, AJ, 133, 768

Le Borgne, D. & Rocca-Volmerange, B. 2002, A&A, 386, 446

Leonard, D. C., Li, W., Filippenko, A. V., Foley, R. J., & Chornock, R. 2005, ApJ, 632, 450

Lewis, A. & Bridle, S. 2002, Phys. Rev. D, 66, 103511

Li, W. et al. 2003, PASP, 115, 453

Li, W. et al. 2001, PASP, 113, 1178

Li, W. et al. 2010, ArXiv e-prints

Maíz Apellániz, J. 2006, AJ, 131, 1184

Malmquist, K. G. 1936, Medd. Stockholmes Obs., 26

Ménard, B., Kilbinger, M., & Scranton, R. 2010, MNRAS, 1064

Neill, J. D., Hudson, M. J., & Conley, A. 2007, ApJ, 661, L123

Oke, J. B. & Gunn, J. E. 1983, ApJ, 266, 713

Pastorello, A. et al. 2007, MNRAS, 377, 1531

Perlmutter, S. et al. 1999, ApJ, 517, 565

Perrett, K. et al. 2010, AJ, 140, 518

Phillips, M. M. et al. 2007, PASP, 119, 360

Pickles, A. J. 1998, PASP, 110, 863

Pike, R. W. & Hudson, M. J. 2005, ApJ, 635, 11

Regnault, N., Astier, P., & Guy, J. 2009, in A&A, submitted

Richardson, D., Branch, D., Casebeer, D., Millard, J., Thomas, R. C., & Baron, E. 2002, AJ, 123, 745

Riess, A. G. et al. 1999, AJ, 117, 707

Riess, A. G. et al. 2004, ApJ, 607, 665

Riess, A. G. et al. 2007, ApJ, 659, 98

Schlegel, D. J., Finkbeiner, D. P., & Davis, M. 1998, ApJ, 500, 525

Shaw, B. & de Jong, R. S. 2008, Verifying the NICMOS Count Dependent Non-Linearity Correction, Tech. rep.

Smith, J. A. et al. 2002, AJ, 123, 2121

Stritzinger, M., Suntzeff, N. B., Hamuy, M., Challis, P., Demarco, R., Germany, L., & Soderberg, A. M. 2005, PASP, 117, 810

Strolger, L.-G. et al. 2004, ApJ, 613, 200

Strolger, L.-G. et al. 2002, AJ, 124, 2905

Sullivan, M. et al. 2010a, MNRAS, 406, 782

Sullivan, M., Ellis, R. S., Howell, D. A., Riess, A., Nugent, P. E., & Gal-Yam, A. 2009, ApJ, 693, L76

Sullivan, M. et al. 2006a, AJ, 131, 960

Sullivan, M. et al. 2006b, ApJ, 648, 868

Sullivan, M. 2011b, in ApJ, in preparation

Taylor, J. 1997, Introduction to Error Analysis, the Study of Uncertainties in Physical Measurements, 2nd Edition (University Science Books)

Tucker, D. L. et al. 2006, Astronomische Nachrichten, 327, 821

Upadhye, A., Ishak, M., & Steinhardt, P. J. 2005, Phys. Rev. D, 72, 063501

Walker, E. S., Hook, I. M., & Sullivan, M. 2010, in MNRAS, submitted

Wang, X., Li, W., Filippenko, A. V., Foley, R. J., Smith, N., & Wang, L. 2008, ApJ, 677, 1060

Wood-Vasey, W. M. et al. 2007, ApJ, 666, 694