epl draft

Cosmological Tests with the Joint Lightcurve Analysis

F. Melia[1] [(a)], J.-J. Wei[2] , R. S. Maier[3] , and X.-F. Wu[4]

[1] *Depts. of Physics and Astronomy, and The Applied Math Program, The University of Arizona, AZ 85721 USA*
`fmelia@email.arizona.edu`
[2] *Purple Mountain Observatory, Chinese Academy of Sciences, Nanjing 210008, China*
`jjwei@pmo.ac.cn`
[3] *Depts. of Mathematics and Physics, and The Statistics Program, The University of Arizona, AZ 85721 USA*
`rsm@math.arizona.edu`
[4] *Purple Mountain Observatory, Nanjing 210008, China;* `xfwu@pmo.ac.cn`

Abstract –We examine whether a comparison between wCDM and $R_{\rm h} = ct$ using merged Type Ia SN catalogs produces results consistent with those based on a single homogeneous sample. Using the Betoule et al. [1] joint lightcurve analysis (JLA) of a combined sample of 613 events from SNLS and SDSS-II, we estimate the parameters of the two models and compare them. We find that the improved statistics can alter the model selection in some cases, but not others. In addition, based on the model fits, we find that there appears to be a lingering systematic offset of ~ 0.04–0.08 mag between the SNLS and SDSS-II sources, in spite of the cross-calibration in the JLA. Treating wCDM, ΛCDM and $R_{\rm h} = ct$ as separate models, we find in an unbiased pairwise statistical comparison that the Bayes Information Criterion (BIC) favors the $R_{\rm h} = ct$ Universe with a likelihood of 82.8% versus 17.2% for wCDM, but the ratio of likelihoods is reversed (16.2% versus 83.8%) when $w_{\rm de} = -1$ (i.e., ΛCDM) and strongly reversed (1.0% versus 99.0%) if in addition $k = 0$ (i.e., flat ΛCDM). We point out, however, that the value of k is a measure of the net energy (kinetic plus gravitational) in the Universe and is not constrained theoretically, though some models of inflation would drive $k \to 0$ due to an expansion-enforced dilution. Since we here consider only the basic ΛCDM model, the value of k needs to be measured and, therefore, the pre-assumption of flatness introduces a significant bias into the BIC.

Introduction. –

A study of the Type Ia SN Hubble diagram has allowed us to measure the expansion history of the Universe [2–4]. But as successful as this program has been, it relies on the use of integrated quantities that are not independent of the assumed dynamics. The use of SN measurements in conducting unbiased, comparative studies of alternative expansion histories is therefore intricate, because at least three 'nuisance' parameters characterizing the standard candle must be optimized simultaneously with any cosmological model's free parameters [5,6].

It has therefore been useful to seek additional methods of probing the cosmic spacetime, including the analysis of cosmic chronometers [7], gamma-ray bursts [8], high-z quasars [9], the cosmic microwave background [10], and Baryon Acoustic Oscillations (BAO) [11]. However, the

results of these studies differ from the Type Ia SN perception that either wCDM (with dynamical dark energy) or ΛCDM (with a cosmological constant) is the optimal cosmological model. Instead, they tend to favor another Friedmann–Robertson–Walker (FRW) cosmology known as the $R_{\rm h} = ct$ Universe [13–16]. For example, the most recent application of the Alcock–Paczyński test to model-independent BAO data [11,12] has favored $R_{\rm h} = ct$ over wCDM at better than a 99.34% confidence level [12].

It is therefore desirable to compare wCDM, ΛCDM and $R_{\rm h} = ct$ directly, using the SN measurements themselves. In our previous paper [17], we carried out such a comparative analysis based on a single, homogeneous Type Ia SN sample—the Supernova Legacy Survey (SNLS; ref. [18]). It is well known now that merging different subsamples, each with its own set of possibly unknown systematics,

---

[(a)] John Woodruff Simpson Fellow.

can introduce inconsistencies that reduce the power of Type Ia SNe for model comparisons. Various techniques have been employed to address this problem, including the introduction of an intrinsic dispersion for each subsample, constrained by the requirement that the $\chi^2_{\rm dof}$ of the fit be equal to one in each case [5, 19] (See also the general discussion in ref. [1].) It is questionable whether this approach provides a statistically fair selection between different models. In our previous analysis using a single large compilation, rather than a merger of unrelated subsamples, the results were quite clear: In a pairwise comparison, the Bayes Information Criterion (BIC) favors $R_{\rm h} = ct$ over ΛCDM with a likelihood of $\approx 90\%$ versus only $\approx 10\%$ for the standard model. The ratio of likelihoods is even greater when comparing $R_{\rm h} = ct$ with wCDM.

Recent progress was made using merged samples [1] by introducing a new cross-calibration of the Sloan Digital Sky Survey (SDSS-II; $0.05 < z < 0.4$) [20] and SNLS ($0.2 < z < 1$) [18] samples. In this *Letter*, we repeat our one-on-one comparison of wCDM and ΛCDM with $R_{\rm h} = ct$, though this time using the joint analysis of these SDSS-II and SNLS sub-samples [1] with 613 SNe Ia. We examine whether the outcome of such a study using merged samples is consistent with that based solely on a single (though relatively large) compilation.

The Combined Supernova Sample. – The Union2.1 catalog [5], which currently includes 580 SN detections [21, 22], is a merger of subsamples, each with its own set of systematic and intrinsic uncertainties, commonly subsumed into an unknown *intrinsic* dispersion $\sigma_{\rm int}$. Note, however, that some systematic uncertainties may relate to the Hubble diagram as a whole, in which case they do not contribute to $\sigma_{\rm int}$. For this *Letter*, we do not have sufficient information to separate the two. For this reason, and the fact that the cross-calibration in the JLA allowed ref. [1] to find a a single sample-wide dispersion in place of individual $\sigma_{\rm int}$'s, we shall also adopt their approach here. In other circumstances, however, where a cross-calibration is not available [17, 23, 24], it is questionable whether model parameters can be estimated by minimizing an overall χ^2 (while constraining the $\chi^2_{\rm dof}$ of each subsample to equal unity), since the unknown $\sigma_{\rm int}$'s should be estimated simultaneously with all other parameters [24]. As an alternative, the method of maximum likelihood estimation (MLE) has been shown to yield superior results [17, 23, 24], though when multiple $\sigma_{\rm int}$'s are used, the analysis becomes computationally challenging. We shall see that, even though we shalll adopt the single sample-wide dispersion of ref. [1], MLE is still required because the probabilities depend on nuisance parameters that need to be optimized with the fits (see discussion following Eq. (6) below).

Some catalogs available for this work are large and well suited to the analysis we carry out in this *Letter*. For example, about half of the Type Ia SNe in Union2.1 came from the SNLS [18]. Since the same instruments and reduction techniques were employed for all 252 of these ($0.15 < z < 1.1$) events, a single $\sigma_{\rm int}$ is sufficient to characterize the unknown intrinsic scatter.

In this *Letter*, we optimize the cosmological fits using the joint lightcurve analysis (JLA) [1], a true recalibration of all the data in SNLS and SDSS-II, based on the use of tertiary standard stars observed by both experiments as reference, a common point-spread function, and the same procedural steps, yielding a photometric accuracy approaching ≈ 5 mmag (but see the discussion below of a possible relic offset between them). The merged sample has a consistent calibration and systematics.

Two features of the JLA catalog need to be addressed, however. First, the complete catalog contains not only the SNLS and SDSS-II events used in the cross-calibration, but also an additional low-z ($z < 0.1$) sample, and the HST SNe at high redshift ($z \approx 1$). Neither of these two groupings was involved in the cross-calibration, so the problem of disparate systematics and intrinsic differences remains for them. The low-z events were calibrated against secondary photometric standards [25], but one must allow for measurement uncertainties in the reference F-subdwarf BD +17 4708 system used there. The calibration of the HST SNe was based on the Riess et al. [26] interpretation. Consistent with our goal of avoiding unknown systematics and calibration uncertainties as much as possible, we shall not include these two (small) subsamples in our analysis because their calibration was handled differently from that of SNLS and SDSS-II.

Second, ref. [1] managed to avoid some of the problems associated with the introduction of intrinsic dispersions, $\sigma_{\rm int}$, but could not completely eliminate them. However, instead of treating them as 'nuisance' parameters to be estimated during the analysis itself (as was done previously), they attempted to find a model-independent dispersion as a function of redshift, by partitioning the SNe into different redshift intervals and using the distribution of observed magnitudes to guess an overall dispersion in each bin. In their assessment, the various dispersions 'measured' in this way are consistent with a single, constant value $\sigma_{\rm coh} = 0.106 \pm 0.006$ (labeled in this fashion to distinguish it from the older $\sigma_{\rm int}$'s). This is the sample-wide 'intrinsic' dispersion we shall also use for the JLA in this *Letter*, consistent with the approach of ref. [1].

The distance modulus of a given supernova is inferred from a fit of its spectral evolution using one of several light-curve models. For this step, two methods are commonly used, SiFTO and SALT2 [27]. Ref. [1] computed the (observed) distance modulus of each Type Ia SN using SALT2, and since we are closely following their approach, we shall also use this lightcurve fitter in our analysis. An argument in favour of this choice for the JLA is that SALT2 is data-driven, and does not introduce any significant bias between low and high-redshift distances [28], which is essential when dealing with a large sample that spans the redshift range we have here. SALT2 includes the following components: (i) the apparent mag-

nitude, m_B, of the SN at maximum light; (ii) the 'shape,' X_1, of its lightcurve; and (iii) its deviation, C, from the mean Type Ia SN $B - V$ color. The formula for the distance modulus is $\mu_B \equiv m_B + \alpha \cdot X_1 - \beta \cdot C - M_B$, where M_B is the absolute B-band magnitude of a Type Ia SN with $X_1 = 0$ and $C = 0$. An allowance may also be made for the assumed host galaxy mass [5,29], introduced as an adjustment ΔM_{host} to M_B for SNe in host galaxies with a mass $> 10^{10} M_\odot$ [1].

'Nuisance' parameters used in computing each distance modulus, such as α, β, M_B (and ΔM_{host} if present), must be fitted simultaneously with parameters characterizing the cosmology itself. Many previously reported model comparisons have not uniformly recalibrated the data for each cosmology being tested. For example, in ref. [6] we adopted the lightcurve parameters optimized for ΛCDM and used them for all the models being tested. But this simplifying assumption can vitiate the outcome of model selection. One goal of this *Letter* is to relax this restriction and re-optimize the nuisance parameters separately for each model. When this procedure is followed, past experience has shown that the cosmology preferred by the Type Ia SN data is not always wCDM or ΛCDM [17].

Model Comparisons. – For each SN, the theoretical distance modulus μ_{th} is calculated from the measured redshift z by the definition $\mu_{\mathrm{th}}(z) \equiv 5 \log[D_{\mathrm{L}}(z)/10 \text{ pc}]$, where $D_{\mathrm{L}}(z)$ is the model-dependent luminosity distance. wCDM and ΛCDM assume specific constituents in the density, expressed as $\rho = \rho_{\mathrm{r}} + \rho_{\mathrm{m}} + \rho_{\mathrm{de}}$, where ρ_{r}, ρ_{m} and ρ_{de} are, respectively, the energy densities of radiation, (luminous and dark) matter, and dark energy. These densities are often represented in terms of today's critical density, $\rho_{\mathrm{c}} \equiv 3c^2 H_0^2 / 8\pi G$, as $\Omega_{\mathrm{m}} \equiv \rho_{\mathrm{m}}/\rho_{\mathrm{c}}$, $\Omega_{\mathrm{r}} \equiv \rho_{\mathrm{r}}/\rho_{\mathrm{c}}$, and $\Omega_{\mathrm{de}} \equiv \rho_{\mathrm{de}}/\rho_{\mathrm{c}}$. H_0 is the Hubble constant, and the other symbols have their usual meanings. In $R_{\mathrm{h}} = ct$, on the other hand, whatever constituents are present in ρ beyond matter and radiation, the principal constraint is the zero active mass condition [13–16], which corresponds to a *total* equation-of-state $p = -\rho/3$.

wCDM has a dark energy with an equation-of-state $p_{\mathrm{de}} = w_{\mathrm{de}} \rho_{\mathrm{de}}$ and $w_{\mathrm{de}} \neq -1$. Its luminosity distance is given by the expression

$$D_{\mathrm{L}}^{w\mathrm{CDM}}(z) = \frac{c}{H_0} \frac{(1 + z_{\mathrm{hel}})}{\sqrt{|\,\Omega_k\,|}} \, \mathrm{sinn}$$

$$\left\{ \int_0^z \frac{\mathrm{d}z \,|\,\Omega_k\,|^{1/2}}{\sqrt{\Omega_{\mathrm{m}}(1+z)^3 + \Omega_k(1+z)^2 + \Omega_{\mathrm{de}}(1+z)^{3(1+w_{\mathrm{de}})}}} \right\} \quad (1)$$

where z and z_{hel} are the CMB rest frame and heliocentric redshifts of the SN, and $\Omega_k = 1 - \Omega_{\mathrm{m}} - \Omega_{\mathrm{de}}$ represents the spatial curvature of the Universe—appearing as a term proportional to k in the Friedmann equation. In addition, sinn is sinh when $\Omega_k > 0$ and sin when $\Omega_k < 0$. For a flat Universe with $\Omega_k = 0$, the right-hand side of this expression simplifies to the form $(1 + z_{\mathrm{hel}})c/H_0$ times the integral (though without the Ω_k term).

Depending on the application, the standard model may contain as many as ten parameters, though only three of these are critical for supernova work. One may adjust Ω_{m}, k (or equivalently Ω_{de}), and w_{de}. It is well known that H_0 is degenerate with M_B when constructing a SN Hubble diagram, so it is not free if M_B is one of the optimized variables [5].

If dark energy is assumed to be a cosmological constant, with $w_{\mathrm{de}} = -1$, the standard model becomes ΛCDM. The principal parameters in this model are Ω_{m} and the spatial curvature constant k. One often sees flatness (i.e., $k = 0$) assumed on the basis of other kinds of observation, but there are good reasons to avoid this if possible. Unlike the distinction between $w_{\mathrm{de}} \neq -1$, which represents dynamical dark energy, and $w_{\mathrm{de}} = -1$, which represents a cosmological constant, there is no theoretical basis to distinguish $k = 0$ and $k \neq 0$, though some models of inflation would have driven $k \to 0$ via an expansion-enforced dilution of the Universe's net energy (kinetic plus gravitational). But since we are here considering only a basic ΛCDM model, the value of k is an initial condition, not a fundamental constraint, and must be measured from the observations, as is routinely done, e.g., with anisotropies in the cosmic microwave background. When one uses parameters optimized in previous studies, however, one is obliged to use other parameters optimized in correspondence with these values, which would actually yield less favorable fits to the SN data. In other words, when comparing models, it is not fair statistically to adopt a previously optimized value of k, while ignoring other parameters that are then re-optimized with SN data. The two models we have at our disposal for unbiased SN work are therefore wCDM, with three free parameters: Ω_{m}, Ω_{de}, and w_{de}, and ΛCDM with Ω_{m} and $\Omega_{\mathrm{de}} \equiv \rho_\Lambda/\rho_{\mathrm{c}}$. But to demonstrate how critical this issue of pre-optimized parameters can be, we shall also show the result of model comparisons using flat ΛCDM. As we shall see, the JLA sample is so large now that even one change in the handling of the parameters in the standard model can greatly alter the outcome of the analysis.

The luminosity distance in $R_{\mathrm{h}} = ct$ is given by the simpler expression

$$D_{\mathrm{L}}^{R_{\mathrm{h}} = ct}(z) = \frac{c}{H_0}(1 + z_{\mathrm{hel}}) \ln(1 + z) \,. \quad (2)$$

Since H_0 is degenerate with M_B, the $R_{\mathrm{h}} = ct$ Universe has no parameters to adjust when we construct its SN Hubble diagram. Further discussion on observational differences between wCDM and $R_{\mathrm{h}} = ct$ appears in refs. [6–8, 13, 14, 38]. For a pedagogical treatment, see also ref. [30].

The cosmological parameters, along with the model-specific nuisance parameters, are estimated using an approach first described in refs. [23, 24], and more fully developed by us in ref. [17]. It is based on the joint likelihood function to be maximized for all these parameters, or as a multiplicative factor that modifies an assumed flat

Bayesian prior. This function is

$$L = \frac{\exp\left[-\frac{1}{2}(\hat{\mu}_{\mathbf{B}} - \hat{\mu}_{\mathrm{th}})^{\mathbf{T}} \, \mathbf{C}^{-1} \, (\hat{\mu}_{\mathbf{B}} - \hat{\mu}_{\mathrm{th}})\right]}{\sqrt{(2\pi)^n \det \mathbf{C}}} \, , \quad (3)$$

where $\hat{\mu}_{\mathbf{B}}$ ($\hat{\mu}_{\mathrm{th}}$) is the observed (theoretical) distance modulus vector with n components, n being the number of SNe, and $\mathbf{C}$ is the full $n \times n$ covariance matrix (including both statistical and systematic errors), defined by

$$\mathbf{C} = \mathbf{D}_{\mathrm{stat}} + \mathbf{C}_{\mathrm{stat}} + \mathbf{C}_{\mathrm{sys}} \, . \quad (4)$$

In this, $\mathbf{D}_{\mathrm{stat}}$ is the diagonal part of the statistical uncertainty, given by

$$(\mathbf{D}_{\mathrm{stat}})_{ii} = \sigma^2_{m_B,i} + \alpha^2 \sigma^2_{X_1,i} + \beta^2 \sigma^2_{\mathcal{C},i}$$
$$+ C_{m_B \, X_1 \, \mathcal{C},i} + \sigma^2_{\mathrm{pec},i} + \sigma^2_{\mathrm{lens},i} + \sigma^2_{\mathrm{coh}} \, , \quad (5)$$

where $\sigma_{m_B,i}$, $\sigma_{X_1,i}$, and $\sigma_{\mathcal{C},i}$ are the standard errors of the peak magnitude and light-curve parameters of the i'th SN. The term $C_{m_B \, X_1 \, \mathcal{C},i}$ derives from the covariances among $m_B, X_1, \mathcal{C}$, and itself depends quadratically on the nuisance parameters α, β. The dispersion $\sigma_{\mathrm{pec},i} = 5\sigma_z/(z_i \log 10)$ accounts for the uncertainty in cosmological redshift due to peculiar velocities, and $\sigma_{\mathrm{lens},i}$ accounts for the variation of magnitudes caused by gravitational lensing. We follow ref. [1] in using $c\sigma_z = 150$ km s^{-1}, as well as $\sigma_{\mathrm{lens},i} = 0.055 \times z_i$, as suggested in ref. [31]. The statistical and systematic covariance matrices, $\mathbf{C}_{\mathrm{stat}}$ and $\mathbf{C}_{\mathrm{sys}}$, are generally not diagonal [32], and for the JLA are given by

$$\mathbf{C}_{\mathrm{stat}} + \mathbf{C}_{\mathrm{sys}} = \mathbf{V}_0 + \alpha^2 \mathbf{V}_a +$$
$$\beta^2 \mathbf{V}_b + 2\alpha \mathbf{V}_{0a} - 2\beta \mathbf{V}_{0b} - 2\alpha\beta \mathbf{V}_{ab} \, , \quad (6)$$

where $\mathbf{V}_0$, $\mathbf{V}_a$, $\mathbf{V}_b$, $\mathbf{V}_{0a}$, $\mathbf{V}_{0b}$, and $\mathbf{V}_{ab}$ are available from ref. [1] at http : //supernovae.in2p3.fr. From the resulting model-specific likelihood function L, which can also be viewed as a Bayesian posterior, we determine the best-fit values for the parameters by maximizing over the joint parameter space. In this just-described estimation, we take into account the extensive analysis carried out in refs. [1,32], concerning systematic errors.

Each distance modulus $\mu_{B,i}$ depends on α, β, M_B, and ΔM_{host}. As stated, we maximize the likelihood L over all nuisance parameters. Kim [24] notes that such a 'full MLE' is better founded statistically, since it treats on the same level all cosmological and all nuisance parameters, the uncertainties in which can affect each other, including those of the intrinsic dispersion(s), if these are themselves not known a priori. (This will not be the case here, since σ_{coh} is fixed at 0.106. However, we do not expect the lack of individual optimization of σ_{coh} for each model to significantly affect our results [33].) We emphasize, however, that even though σ_{coh} is fixed in this approximation, the Gaussian normalization in our likelihood analysis is still not a constant. It depends on the value of α, β, M_B, and ΔM_{host} (see Eqn. 5). Thus, maximizing the

likelihood function L is not exactly equivalent to minimizing the χ^2 statistic, i.e., $\chi^2 = (\hat{\mu}_{\mathbf{B}} - \hat{\mu}_{\mathrm{th}})^{\mathbf{T}} \, \mathbf{C}^{-1} \, (\hat{\mu}_{\mathbf{B}} - \hat{\mu}_{\mathrm{th}})$.

In addition to computing the best-fit parameter values from the likelihood function L by MLE, we treat L in a Bayesian fashion as an unnormalized probability density function (PDF) on the joint parameter space, and employ Markov-chain Monte Carlo (MCMC) techniques to generate a large random sample of points from this space, distributed according to the PDF. The standard error for each estimated parameter is then obtained as an empirical standard deviation of this sample.

Because wCDM, ΛCDM and $R_{\mathrm{h}} = ct$ have different numbers of free parameters, comparing their likelihoods of being the 'correct' model requires the use of a model selection criterion. Since the samples we are dealing with here are very large, the most appropriate tool to use [7] is the Bayes Information Criterion, which approximates the computation of the (logarithm of the) 'Bayes factor' for deciding between models [34,35]. The BIC is defined for each model being fit by

$$\exp(-\mathrm{BIC}/2) \equiv n^{-p/2} L^* \, , \quad (7)$$

where L^* is the maximized likelihood, n ($= 613$ here) the data set size, and p the count of free parameters in the model. If BIC_α comes from model α, the unnormalized likelihood of model α being correct is the 'Bayes weight' $\exp(-\mathrm{BIC}_\alpha/2)$. Thus model α ($\alpha = 1, 2$) has likelihood

$$P(\alpha) = \frac{\exp(-\mathrm{BIC}_\alpha/2)}{\exp(-\mathrm{BIC}_1/2) + \exp(-\mathrm{BIC}_2/2)} \quad (8)$$

of being the correct choice. This has a Bayesian interpretation: $\exp(-\mathrm{BIC}_\alpha/2)$ is a large-sample ($n \to \infty$) approximation to an integral over the parameter space of model α, of its likelihood function L. In the limit, the standard error of each parameter shrinks like $n^{-1/2}$, and the integral of L equals up to a constant factor the $n^{-p/2} L^*$ of Eq. (7).

By convention, the magnitude of the difference $\Delta \equiv \mathrm{BIC}_2 - \mathrm{BIC}_1$ provides a numerical assessment of the evidence that model 1 is favoured over model 2. The rule of thumb is that if $\Delta \lesssim 2$, the evidence is weak; if $\Delta \sim 3$ or 4, it is mildly strong, and if $\Delta \gtrsim 5$, it is quite strong.

Hubble Diagram. – Several of the SNLS and SDSS-II supernovae fall outside the range of validity established for the lightcurve fitter, SALT2, and must therefore be removed [1]. The pruned catalogs include 239 events from SNLS and 374 from SDSS-II, for a total of 613 events.

The best-fit parameters for wCDM, ΛCDM and $R_{\mathrm{h}} = ct$, obtained by MLE, are provided in Tables 1 and 2, along with a standard error for each (estimated by MCMC).[1] The first comparison we make is an extension to our previous work based solely on the SNLS [17]. The principal

---

[1]A chain of 10^5 points in the parameter space distributed according to the likelihood function was generated from the Metropolis–Hastings algorithm with a uniform prior. In each case, the distributions of the estimated parameters, with a confidence interval for each, followed from a statistical analysis.
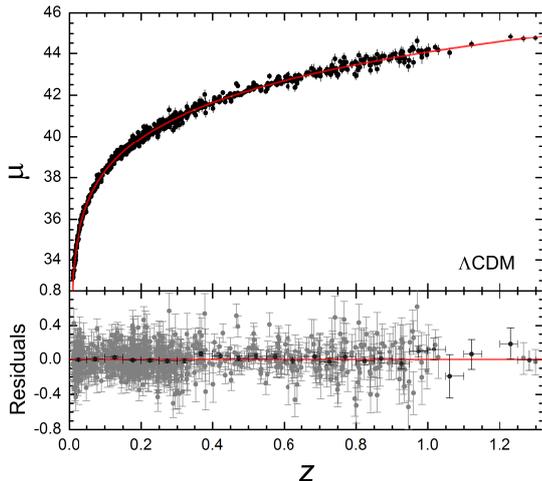
Fig. 1: (Top) Hubble diagram for the combined SNLS and SDSS-II sample of 613 Type Ia SNe, together with the best-fit wCDM model (red curve) (line 2 in Tables 1 and 2). (Bottom) Residuals for the best-fit model, shown in grey for individual sources, and in black for averages over redshift bins of 0.05.
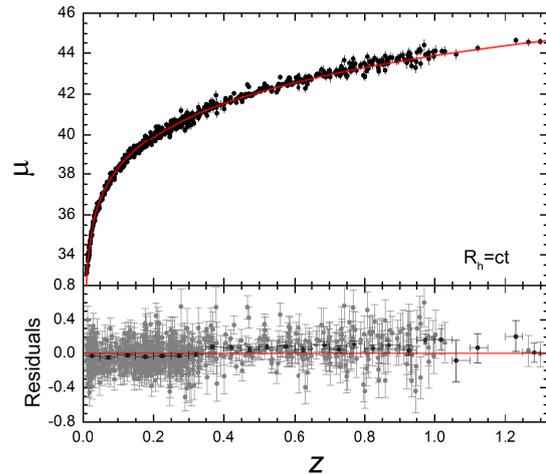

Fig. 2: Similar to Fig. 1, except that these data are calibrated using the $R_{\rm h} = ct$ Universe, and the red curve shows the best fit in this model (see line 1, Tables 1 and 2).

motivation in this *Letter* has been to examine whether the outcome of that analysis is supported by a similar comparative study involving a much bigger, merged sample. The corresponding results for the combined JLA sample are summarized in lines 1, 2 and 3 of Tables 1 and 2 for $R_{\rm h} = ct$, wCDM and ΛCDM. As indicated earlier, we also compare these models with flat ΛCDM (line 4) to demonstrate the impact of adopting a pre-optimized parameter value (for k). The optimization for $R_{\rm h} = ct$, wCDM and ΛCDM is based purely on the SN observations.

On the basis of the JLA, the BIC favors $R_{\rm h} = ct$ over wCDM, with a likelihood of 82.8% versus 17.2%. With $\Delta = 3.14$, the evidence in favour of $R_{\rm h} = ct$ is mildly strong. However, both SNLS and the combined JLA sample are so big that the BIC (see Eq. (7)) is sensitive to the number of free parameters. For example, with one fewer parameter than wCDM, ΛCDM is somewhat favoured over $R_{\rm h} = ct$ with $\Delta = 3.29$, and strongly favoured over wCDM with $\Delta = 6.43$. The sensitivity of this outcome to the various assumptions is further demonstrated by the sample selection. Notice, for example, that the SNLS on its own yields very different likelihoods. In this case, even a comparison between $R_{\rm h} = ct$ and flat ΛCDM shows that the likelihoods are about even, i.e., 43.2% versus 56.8%. The statistically fairer comparison between $R_{\rm h} = ct$ and wCDM and ΛCDM shows that the evidence in favour of the former is mildly—or even very—strong in both cases.

Yet in every case, the $\chi^2_{\rm dof}$ values for the best fit models are hardly distinguishable. It is clear that model selection using Type Ia SNe is therefore heavily influenced by the number of free parameters in the models. And given this sensitivity, it is necessary to avoid biasing the results by assigning pre-optimized values to the variables that are not theoretically constrained (such as k in this case).

The corresponding Hubble diagrams for the best-fit

wCDM (line 2) and $R_{\rm h} = ct$ (line 1) models are shown in Figs. 1 and 2, respectively, together with their residuals. A close inspection of the data in these plots shows that, though very similar, they are not identical, highlighting the importance of estimating the nuisance parameters individually for each different model. It is also quite evident from a comparison of the best-fit curves in these plots that both models fit the data extremely well; the $\chi^2_{\rm dof}$ values attest to this, and demonstrate a comparably high quality fit in each case. We also show in Figs. 3 and 4 the corresponding one- and two-dimensional projections of the posterior probability distributions for the free parameters in wCDM and $R_{\rm h} = ct$, generated by MCMC.

Discussion and Conclusions. – In this *Letter*, we have used the MLE method (with Bayesian extensions). In our previous study of the SNLS sample [17], we also employed MLE, but in addition, contrasted the outcome with that of the more conventional procedure of optimizing $\sigma_{\rm int}$ [5, 19] by requiring that $\chi^2_{\rm dof}$ equal unity. When using a single, homogeneous sample, these two approaches give essentially the same result, because (as we have seen) the $\chi^2_{\rm dof}$ of the optimized fit is almost always close to 1. There is less justification [1] for using the latter approach when several subsamples are merged into a single compilation. Thus, even though the use of MLE with a full covariance matrix is computationally difficult, it should be the method of choice for any model selection involving non-nested models and a blend of diverse subsamples.

The outcome of our analysis using the combined SNLS and SDSS-II sample is in agreement with that of our earlier study based solely on the SNLS for some cases, but not others. The $R_{\rm h} = ct$ universe is consistently favoured over wCDM, but the likelihoods are reversed when comparing $R_{\rm h} = ct$ with ΛCDM. If we introduce a previously optimized value for k, the outcome is strongly reversed.

But this is problematic for several reasons. First, one

Model	α	β	M_B	ΔM_{host}	σ_{coh}
			SNLS + SDSS-II		
1. $R_{\text{h}} = ct$	0.119 ± 0.007	2.600 ± 0.084	-18.932 ± 0.021	-0.052 ± 0.026	0.106 (fixed)
2. wCDM	0.121 ± 0.008	2.631 ± 0.087	-19.020 ± 0.037	-0.051 ± 0.028	0.106 (fixed)
3. ΛCDM	0.121 ± 0.008	2.631 ± 0.086	-19.026 ± 0.035	-0.051 ± 0.028	0.106 (fixed)
4. ΛCDM ($k = 0$)	0.121 ± 0.008	2.630 ± 0.085	-19.040 ± 0.030	-0.051 ± 0.028	0.106 (fixed)
			SNLS		
5. $R_{\text{h}} = ct$	0.108 ± 0.014	2.290 ± 0.153	-18.889 ± 0.016	-0.043 ± 0.023	0.069 ± 0.018
6. wCDM	0.114 ± 0.014	2.351 ± 0.079	-19.055 ± 0.079	-0.043 ± 0.025	0.056 ± 0.022
7. ΛCDM	0.114 ± 0.014	2.351 ± 0.162	-19.061 ± 0.078	-0.043 ± 0.025	0.056 ± 0.022
8. ΛCDM ($k = 0$)	0.113 ± 0.014	2.367 ± 0.156	-19.022 ± 0.039	-0.042 ± 0.024	0.056 ± 0.022

Table 1: Optimized parameters for different cosmological models.

Model	Ω_{m}	Ω_{de}	w_{de}	χ^2_{dof}	BIC
			SNLS + SDSS-II		
1. $R_{\text{h}} = ct$	$\cdots$	$\cdots$	$\cdots$	1.04 (609 dof)	-514.57
2. wCDM	$0.203^{+0.137}_{-0.196}$	$0.445^{+0.265}_{-0.205}$	$-0.956^{+0.276}_{-0.384}$	1.01 (606 dof)	-511.43
3. ΛCDM	0.250 ± 0.152	0.480 ± 0.202	-1 (fixed)	1.01 (607 dof)	-517.86
4. ΛCDM ($k = 0$)	0.353 ± 0.045	$1.0 - \Omega_{\text{m}}$	-1 (fixed)	1.01 (608 dof)	-523.70
			SNLS		
5. $R_{\text{h}} = ct$	$\cdots$	$\cdots$	$\cdots$	0.94 (234 dof)	-177.65
6. wCDM	0.368 ± 0.120	0.806 ± 0.310	$-0.912^{+0.302}_{-0.446}$	0.98 (231 dof)	-169.99
7. ΛCDM	0.453 ± 0.130	0.869 ± 0.336	-1 (fixed)	0.97 (232 dof)	-173.61
8. ΛCDM ($k = 0$)	0.360 ± 0.051	$1.0 - \Omega_{\text{m}}$	-1 (fixed)	0.97 (233 dof)	-178.20

Table 2: Optimized parameters for different cosmological models (cont.)
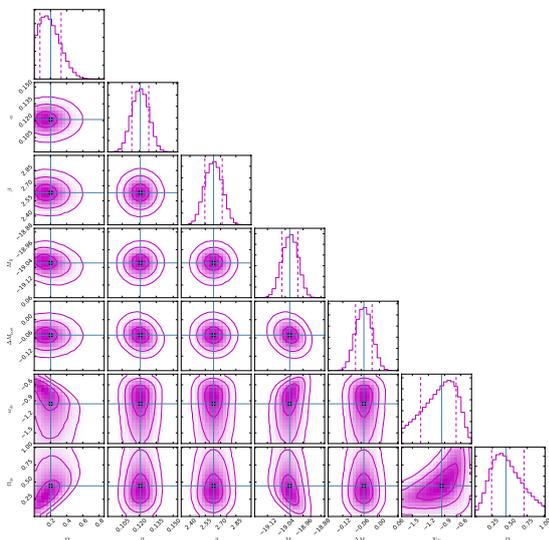
Fig. 3: Posterior probability distributions of the free parameters in wCDM. Contours are 1, 2, and 3 σ. The vertical lines are the best-fit results (solid), and the enclosed 68% credible region (dashed). Made with `triangle.py` from ref. [36].

Fig. 4: Same as Fig. 3, but now for the $R_{\text{h}} = ct$ Universe.

might expect that if a model is correct, it should fit either the SNLS, with ≈ 250 events distributed in redshift $0 < z < 1$, or the combined JLA sample with three times as many SNe spread over a similar redshift range, comparably well, at least qualitatively. The quality of the fit improves as the sample size increases, but one would not expect the model selection to change from one sample to the other because, in both cases, the SNe are distributed across the region (near $z \approx 0.6$) where the transition from deceleration to acceleration is thought to have occurred.

Perhaps an indication of why there may be differences between the analysis of SNLS on its own, versus the merged SNLS and SDSS-II sample, is provided by the tendency of binned residuals in both Figs. 1 and 2 to be slightly lower at $z \lesssim 0.35$ than those at $z \gtrsim 0.35$. The

average difference is about 0.04 magnitudes for wCDM and about 0.08 magnitudes for $R_{\text{h}} = ct$. On the other hand, for the 4 hightest redshift bins, we find an average residual magnitude of 0.139 for $R_{\text{h}} = ct$ and a slightly worse 0.143 for wCDM. Since these residuals do not exhibit any monotonic trend, the implication seems to be that there exists a systematic offset across $z \sim 0.35$. One possible origin for this behavior could be that the calibration of the SNLS and SDSS-II sources is not completely self-consistent after all, and since the SDSS-II events occurred at $0.05 < z < 0.4$, while the SNLS events were recorded at $0.2 < z < 1$, we may simply be seeing the impact of an unresolved measurement offset in their magnitude. Notice, for example, that the implied magnitude offset is comparable to the measured sample-wide 'intrinsic' dispersion $\sigma_{\text{coh}} = 0.106 \pm 0.006$ (see above). The offset may be slightly smaller for wCDM due to the additional free parameter that allows greater flexibility in shaping the best-fit curve. In both cases, however, the key point is that the offset appears to be independent of redshift above

and below the crossover at $z \approx 0.35$.

Some other studies, e.g., ref. [37], have reached different conclusions from those presented here. As we have discussed extensively in this paper, although the statistical analysis of Type Ia SNe may be improved with the merger of disparate subsamples, each subsample comes with its own set of systematic and intrinsic uncertainties. Ref. [37] carried out the model selection using both the Union2.1 and JLA samples, following the conventional approach of minimizing an overall χ^2, though with the constraint that $\chi^2_{\mathrm{dof}} = 1$ for each subsample. One should not be surprised, therefore, that the χ^2_{dof} for the whole compilation is also close to 1. A correct statistical approach, however, would estimate the unknown σ_{int}'s simultaneously with the model-specific and nuisance parameters [17, 24]. One should therefore use MLE.

More importantly, Union2.1 contains over 17 subsamples. The total number of 'nuisance' parameters is therefore 20, since all of the σ_{int}'s should be re-estimated for each model for a truly unbiased test. Moreover, the expressions used by ref. [37] to compute the BIC are incorrect because the σ_{int}'s themselves are not known a priori. So the information criteria must be calculated in terms of the likelihood function [17], not χ^2. Finally, since ref. [37] used several previously optimized parameter values, those results are not unbiased like the outcomes shown in lines 1–3 of Tables 1 and 2.

Comparing our results using the SNLS on its own and the merged SNLS and SDSS-II sample shows that the model selection using these Type Ia SNe is still somewhat ambiguous. Our analysis has shown that the cross-calibration in the JLA may be imperfect, with a residual offset between the two samples of ~ 0.08 mag, comparable to the sample-wide dispersion $\sigma_{\mathrm{coh}} \sim 0.106$ mag inferred in ref. [1]. Such an offset tends to favour models with a larger number of free parameters, with a greater flexibility in adjusting the shape of their luminosity distance to fit the data. This may explain why $R_{\mathrm{h}} = ct$ is favoured over wCDM, but not always over ΛCDM.

An important goal of future work with Type Ia SNe should be to improve the consistent cross-calibration of independent datasets, along the lines initiated by ref. [1], though with even greater precision. If $R_{\mathrm{h}} = ct$ were to eventually become the cosmology favoured by the Type Ia SN data, such an outcome would be relevant to the growing tension between the predictions of an inflationary cosmology and the *Planck* measurements. The Universe did not require an early period of inflated expansion to avoid the horizon problem in $R_{\mathrm{h}} = ct$ [38], so the $R_{\mathrm{h}} = ct$ explanation for the uniformity of the physical conditions across the Universe may be more realistic than the currently held belief of an inflated expansion.

$$* * *$$

REFERENCES

[1] BETOULE, M. ET AL., *Astron. Astrophys.*, **568** (2014) 22.
[2] PERLMUTTER, S. ET AL., *Nature*, **391** (1998) 51.
[3] RIESS, A. G. ET AL., *Astron. J.*, **116** (1998) 1009.
[4] SCHMIDT, B. P. ET AL., *Astrophys. J.*, **507** (1998) 46.
[5] SUZUKI, N. ET AL., *Astrophys. J.*, **746** (2012) 85.
[6] MELIA, F., *Astron. J.*, **144** (2012) 110.
[7] MELIA, F. AND MAIER, R. S., *Mon. Not. R. Astron Soc.*, **432** (2013) 2669.
[8] WEI, J.-J., WU, X. AND MELIA, F., *Astrophys. J.*, **772** (2013) 43.
[9] MELIA, F., *Astrophys. J.*, **764** (2013) 72.
[10] MELIA, F., *Astron. Astrophys.*, **561** (2014) A80.
[11] DELUBAC, T. ET AL., *Astron. Astrophys.*, **574** (2014) A59.
[12] MELIA, F. ET AL., *Int. J. Mod. Phys. D*, **27** (2017) 1750055.
[13] MELIA, F., *Mon. Not. R. Astron Soc.*, **382** (2007) 1917.
[14] MELIA, F. ET AL., *Mon. Not. R. Astron Soc.*, **419** (2012) 2579.
[15] MELIA, F., *Front. Phys.*, **11** (2016) 119801.
[16] MELIA, F., *Front. Phys.*, **12** (2017) 129802.
[17] WEI, J. ET AL., *Astron. J.*, **149** (2015) 102.
[18] GUY, J. ET AL., *Astron. Astrophys.*, **523** (2010) A7.
[19] AMANULLAH, R. ET AL., *Astrophys. J.*, **716** (2010) 712.
[20] SAKO, M. ET AL., *Astrophys. J. Suppl. Ser.*, (2015) arXiv:1401.3317S.
[21] KUZNETSOVA, N. ET AL., *Astrophys. J.*, **673** (2008) 981.
[22] RIESS, A. G. ET AL., *Astrophys. J.*, **730** (2011) 119.
[23] D'AGOSTINI, G., , **2005** (arXiv:physics/0511182) .
[24] KIM, A. G., *Publ. Astron. Soc. Pac.*, **123** (2011) 230.
[25] LANDOLT, A. U. AND UOMOTO, A. K., *Astron. J.*, **133** (2007) 768.
[26] RIESS, A. G. ET AL., *Astrophys. J.*, **659** (2007) 98.
[27] CONLEY, A. ET AL., *Astrophys. J.*, **681** (2008) 482.
[28] MOSHER, J. ET AL., *Astrophys. J.*, **793** (2014) 16.
[29] LAMPEITL, H. ET AL., *Astrophys. J.*, **722** (2010) 566.
[30] MELIA, F., *Australian Phys*, **49** (2012) 83.
[31] JÖNSSON, J. ET AL., *Mon. Not. R. Astron Soc.*, **405** (2010) 535.
[32] CONLEY, A. ET AL., *Astrophys. J. Suppl. Ser.*, **192** (2011) 1.
[33] RUBIN, D. ET AL., *Astrophys. J.*, **813** (2015) 137.
[34] SCHWARZ, G., *Ann. Statist.*, **6** (1978) 461.
[35] KASS, R. E. ET AL., *J. Amer. Stat. Assoc.*, **90** (1995) 773.
[36] FOREMAN-MACKEY, D. ET AL., *Publ. Astron. Soc. Pac.*, **125** (2013) 306.
[37] SHAFER, D. L., *Phys. Rev. D*, **91** (2015) 103516.
[38] MELIA, F., *Astron. Astrophys.*, **553** (2013) A76.