CrossMark

SN REFSDAL: PHOTOMETRY AND TIME DELAY MEASUREMENTS OF THE FIRST EINSTEIN CROSS SUPERNOVA

S. A. Rodney[1], L.-G. Strolger[2], P. L. Kelly[3], M. Bradač[4], G. Brammer[2], A. V. Filippenko[3], R. J. Foley[5,6], O. Graur[7,8], J. Hjorth[9], S. W. Jha[10], C. McCully[11,12], A. Molino[13,14], A. G. Riess[15,2], K. B. Schmidt[12,16], J. Selsing[9], K. Sharon[17], T. Treu[18,21], B. J. Weiner[19], and A. Zitrin[20,22]

[1] Department of Physics and Astronomy, University of South Carolina, 712 Main St., Columbia, SC 29208, USA; srodney@sc.edu
[2] Space Telescope Science Institute, 3700 San Martin Dr., Baltimore, MD 21218, USA
[3] Department of Astronomy, University of California, Berkeley, CA 94720-3411, USA
[4] University of California Davis, 1 Shields Avenue, Davis, CA 95616
[5] Department of Physics, University of Illinois at Urbana-Champaign, 1110 W. Green Street, Urbana, IL 61801, USA
[6] Astronomy Department, University of Illinois at Urbana-Champaign, 1002 W. Green Street, Urbana, IL 61801, USA
[7] Center for Cosmology and Particle Physics, New York University, New York, NY 10003, USA
[8] Department of Astrophysics, American Museum of Natural History, Central Park West and 79th Street, New York, NY 10024, USA
[9] Dark Cosmology Centre, Niels Bohr Institute, University of Copenhagen, Juliane Maries Vej 30, DK-2100 Copenhagen, Denmark
[10] Department of Physics and Astronomy, Rutgers, The State University of New Jersey, Piscataway, NJ 08854, USA
[11] Las Cumbres Observatory Global Telescope Network, 6740 Cortona Dr., Suite 102, Goleta, California 93117, USA
[12] Department of Physics, University of California, Santa Barbara, CA 93106-9530, USA
[13] Instituto de Astronomia, Geofísica e Ciências Atmosféricas, Universidade de São Paulo, Cidade Universitária, 05508-090, São Paulo, Brazil
[14] Instituto de Astrofísica de Andalucía (CSIC), E-18080 Granada, Spain
[15] Department of Physics and Astronomy, The Johns Hopkins University, 3400N. Charles St., Baltimore, MD 21218, USA
[16] Leibniz-Institut für Astrophysik Potsdam (AIP), An der Sternwarte 16, D-14482 Potsdam, Germany
[17] Department of Astronomy, University of Michigan, 1085 S. University Avenue, Ann Arbor, MI 48109, USA
[18] Department of Physics and Astronomy, University of California, Los Angeles, CA 90095, USA
[19] Department of Astronomy, University of Arizona, Tucson, AZ 85721, USA
[20] California Institute of Technology, 1200 East California Boulevard, Pasadena, CA 91125, USA

ABSTRACT

We present the first year of *Hubble Space Telescope* imaging of the unique supernova (SN) "Refsdal," a gravitationally lensed SN at $z = 1.488 \pm 0.001$ with multiple images behind the galaxy cluster MACS J1149.6 +2223. The first four observed images of SN Refsdal (images S1–S4) exhibited a slow rise (over ∼150 days) to reach a broad peak brightness around 2015 April 20. Using a set of light curve templates constructed from SN 1987A-like peculiar Type II SNe, we measure time delays for the four images relative to S1 of 4 ± 4 (for S2), 2 ± 5 (S3), and 24 ± 7 days (S4). The measured magnification ratios relative to S1 are 1.15 ± 0.05 (S2), 1.01 ± 0.04 (S3), and 0.34 ± 0.02 (S4). None of the template light curves fully captures the photometric behavior of SN Refsdal, so we also derive complementary measurements for these parameters using polynomials to represent the intrinsic light curve shape. These more flexible fits deliver fully consistent time delays of 7 ± 2 (S2), 0.6 ± 3 (S3), and 27 ± 8 days (S4). The lensing magnification ratios are similarly consistent, measured as 1.17 ± 0.02 (S2), 1.00 ± 0.01 (S3), and 0.38 ± 0.02 (S4). We compare these measurements against published predictions from lens models, and find that the majority of model predictions are in very good agreement with our measurements. Finally, we discuss avenues for future improvement of time delay measurements—both for SN Refsdal and for other strongly lensed SNe yet to come.

Key words: galaxies: clusters: general – galaxies: clusters: individual (MACS J1149.6+2223) – gravitational lensing: strong – supernovae: general – supernovae: individual (SN Refsdal)

1. INTRODUCTION

The discovery of supernova (SN) Refsdal, the first strongly lensed SN resolved into multiple images, was described by Kelly et al. (2015c). SN Refsdal was located in the arm of a face-on spiral host galaxy at $z = 1.49$. This spiral arm is distorted into an Einstein ring by the gravitational potential of a foreground elliptical galaxy. That elliptical galaxy lens also resides within MACS J1149.6+2223, a strong-lensing galaxy cluster at $z = 0.54$ that is fast becoming one of the crown jewels of the Massive Cluster Survey (Ebeling et al. 2001, 2007). The galactic-scale lens, augmented by the cluster lens, causes SN Refsdal to appear to us as four images with separations of ∼2″, arranged in an "Einstein Cross"

configuration (see Figure 1) reminiscent of the quadruply-imaged quasar that originated this term (Huchra et al. 1985; Adam et al. 1989).

The host galaxy of SN Refsdal, is itself strongly lensed by the MACS J1149.6+2223 cluster, and was identified as a particularly spectacular example of a multiply-imaged galaxy in some of the earliest lens modeling efforts (Smith et al. 2009; Zitrin & Broadhurst 2009). Due to the spatial magnification afforded by the cluster lens, this galaxy has provided a rare opportunity to study the substructure of a $z = 1.5$ galaxy at scales down to ∼100 pc. This galaxy shows evidence for active star formation (Smith et al. 2009; Livermore et al. 2012, 2015) with a young stellar population at the SN Refsdal position (Adamo et al. 2013). Yuan et al. (2011) reported a steep metallicity gradient for the galaxy, and Yuan et al. (2015) measured a low metallicity from nine H II regions at similar

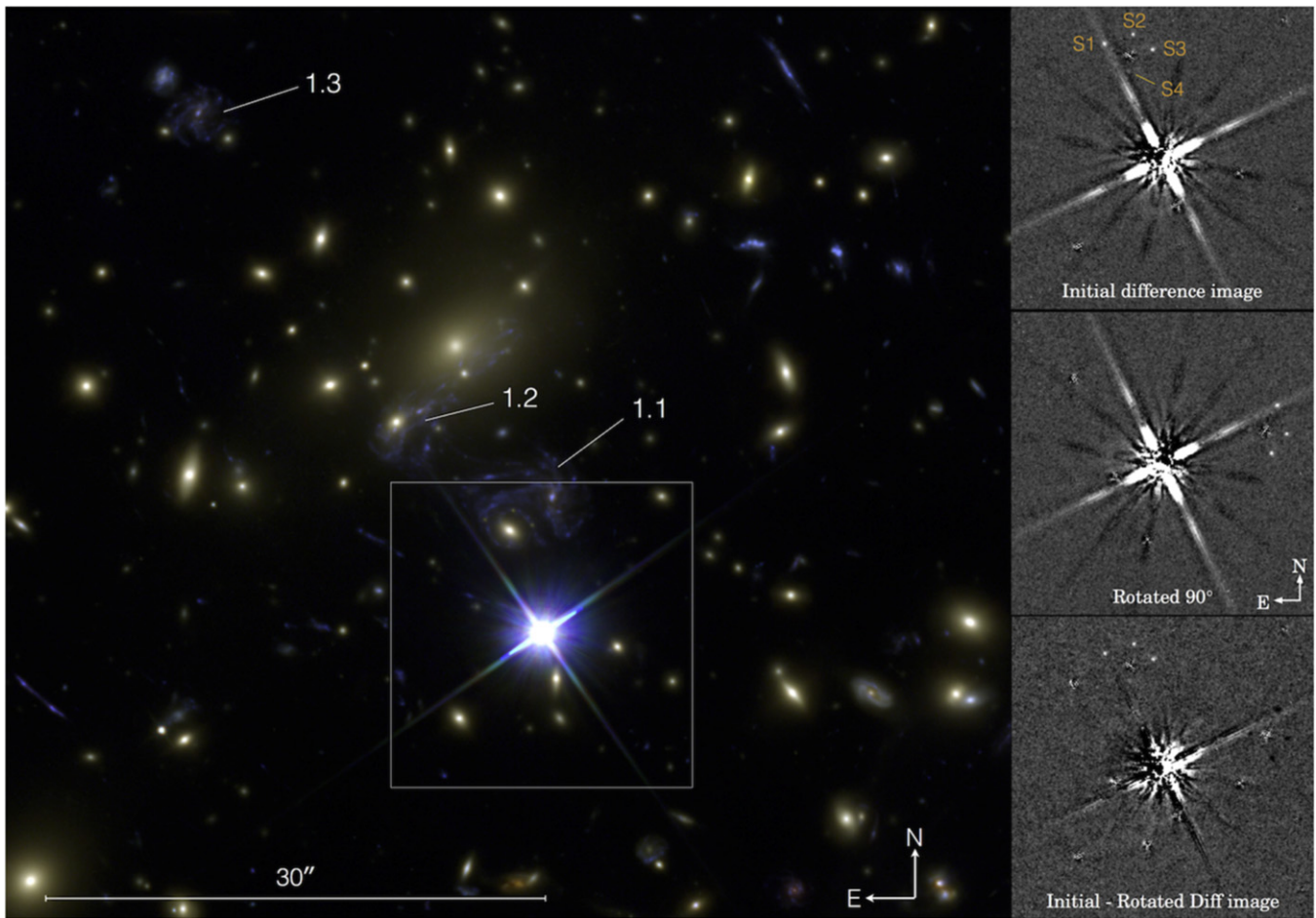---

[21] Packard Fellow.
[22] Hubble Fellow.

Figure 1. MACS J1149.6+2223 field, showing the positions of the three primary images of the SN Refsdal host galaxy (labeled 1.1, 1.2, and 1.3). SN Refsdal appears as four point sources in an Einstein Cross configuration in the southeast spiral arm of image 1.1. The highlighted box is shown at the same scale in panels on the right side, which illustrate the removal of contaminating diffraction spikes from a difference image. Each difference image is centered on the location of the contaminating star (top panel), then rotated clockwise by 90° (middle panel). The rotated difference image is then subtracted from the initial difference image, removing most of the flux from the contaminating diffraction spike at the location of the SN Refsdal point sources (bottom panel).

galactocentric radii. Using integral field spectroscopy with the Very Large Telescope (VLT) MUSE spectrograph, Karman et al. (2015) found Mg II emission at the SN Refsdal position, and inferred from the [O II] to Mg II ratio that the SN exploded in a low metallicity and high ionization environment.

This host galaxy presents at least three distinct images in the plane of the MACS J1149.6+2223 field, and the image in which SN Refsdal was discovered is typically labeled as image 1.1 (Smith et al. 2009). Lens models consistently indicate that the second image of this galaxy, 1.2, is a trailing image (e.g., Kelly et al. 2015c; Oguri 2015; Sharon & Johnson 2015; Treu et al. 2016), and indeed a new transient source appeared at the expected location in 2015 December. This new source is consistent with being the predicted reappearance image, SX (Kelly et al. 2015b). A third image of the host galaxy, image 1.3, is understood to be a leading image, and the first image of SN Refsdal most likely appeared there some 20+ years ago, although the available archival *Hubble Space Telescope* (*HST*) observations cannot confirm or refute this expectation.

Since the discovery of SN Refsdal, many lens modeling teams have produced updated lens models and generated predictions for the SN time delays and magnifications, in some cases taking advantage of the very deep imaging and spectro-sopic data from the *HST* and other observatories (Diego et al.

2016; Grillo et al. 2015; Jauzac et al. 2016; Kawamata et al. 2015; Oguri 2015; Sharon & Johnson 2015). Treu et al. (2016) describes the collaborative development of some of these updated lens models by five independent teams, and highlights the rare opportunity for a true blind test of these models. By generating these predictions in advance of the reappearance of SN Refsdal as image SX, the modelers have provided falsifiable predictions that can be directly confronted with a true measurement of the time delays and magnification ratios. An initial comparison based on the first detection of the reappearance image SX showed that several of the models are consistent with observations (Kelly et al. 2015b). However, a complete evaluation will need to await the full light curve of image SX, which will be collected over the coming year with an ongoing *HST* imaging campaign (GO-PID:14199, PI: P. Kelly).

The comparison of SN Refsdal observations against model predictions is similar in concept to previous tests of lens models using Type Ia SNe as standardizable candles (Nordin et al. 2014; Patel et al. 2014; Rodney et al. 2015a). For this small sample of lensed Type Ia SNe, it was possible to constrain the absolute magnification along a single sight line and compare to model predictions. SN Refsdal is not a Type Ia SN (Kelly et al. 2015c), but instead appears to be a peculiar Type II SN (Kelly

et al. 2015a) so the absolute magnification can not be determined to the same level of precision. However, with magnification factors as high as μ 20, SN Refsdal is much more strongly lensed than any cluster-lensed SNe previously seen. Furthermore, as the only known SN with resolved multiple images, SN Refsdal offers the first chance to test lens models using time delay measurements. This exercise will inform future prospects for using strongly lensed SNe as probes of both galaxy and cluster lenses, and may be valuable for understanding the prevalence of microlensing effects (Dobler & Keeton 2006 and see Section 5.2).

In this paper we present the first year of *HST* photometry of the first four observed images of SN Refsdal. A companion paper (Kelly et al. 2015a) describes the classification of SN Refsdal as a peculiar Type II SN similar to SN 1987A, based on the *HST* light curve as well as *HST* and VLT spectroscopy. An outline of the content of this paper is as follows: Section 2 describes the *HST* imaging observations, data processing, and photometry. To measure the gravitational lensing time delays and magnification ratios, in Section 3 we use light curve templates, and in Section 4 we use flexible polynomial light curve models. Finally, we offer a summary and discussion of results in Section 5.

2. OBSERVATIONS AND PHOTOMETRY

The imaging observations of SN Refsdal presented here were all obtained with *HST* using the Wide-Field Camera 3 (WFC3) with the infrared (IR) and UV-optical (UVIS) detectors, and the Advanced Camera for Surveys (ACS) Wide Field Camera (WFC). Here we present all *HST* observations from the discovery epoch on 2014 November 10 through the observations of 2015 November 15, one year later. The MACS J1149.6+2223 field was continuously observed by *HST* throughout this period with a span of no more than 3 weeks between each visit, except during the period from 2015 July 21 to 2015 October 30, when the field was too close to the Sun for safe observations with *HST*.

As detailed in Kelly et al. (2015c), SN Refsdal was discovered in images collected for the Grism Lens-Amplified Survey from Space (GLASS) program (Schmidt et al. 2014; Treu et al. 2015).[23] The MACS J1149.6+2223 cluster field was subsequently and extensively observed in the course of the *HST* Frontier Fields program (HFF, GO-13504; PI: Lotz)[24] providing a very rich set of optical and near-IR imaging. The HFF imaging cadence was supplemented by observations from the Frontier Fields Supernova program (FrontierSN, GO-13790; PI: Rodney), which extended the WFC3-IR imaging beyond the end of the HFF campaign to complete the near-IR light curves at later times. Additional imaging—as well as deep (34 orbits) grism observations—was provided by an *HST* follow-up program allocated through Director's Discretionary time (GO/DD-14041; PI: Kelly; Kelly et al. 2015a). The *HST* monitoring of this field continues under an ongoing imaging program (GO-14199; PI: Kelly).

To construct multi-color light curves of the four SN Refsdal sources, we first sorted the available observations into 45 imaging epochs, each of which contains observations that were collected within 2 observer-frame days of each other. We then processed the *HST* image data using tools from the

DrizzlePac software suite.[25] The same-filter observations for each epoch were registered to a common astrometric frame using TweakReg and combined with AstroDrizzle (Fruchter et al. 2010). The composite images were drizzled to a pixel scale of $0\rlap.{''}06\,\text{pixel}^{-1}$ for WFC3/IR and $0\rlap.{''}03\,\text{pixel}^{-1}$ for WFC3/UVIS and ACS/WFC data. Most of the resulting composite images in IR bands have total effective exposure times of ~1200 s (half of the exposure time typically available in one *HST* orbit). The ACS-WFC observations are primarily from the deeper HFF visits, and have composite exposure times of ~5000 s (two full *HST* orbits). Table 4 lists all the composite observation epochs, including exposure times.

As the final step in the data processing pipeline, we subtracted off a template image to remove contaminating light from the static foreground cluster galaxies and SN Refsdal's host galaxy. These templates were constructed from *HST* images collected prior to 2014 April 15, with contributions from the GLASS and HFF programs, but primarily from data collected as part of the Cluster Lensing And Supernova survey with *Hubble* (GO-12068; PI Postman, Postman et al. 2012). As can be seen in Figure 1, the location of SN refsdal is uncomfortably close to a 15th magnitude star USNO 1050–06589751 (R.A., decl. = 11:49:35.41, 22:23:38.0, Monet et al. 2003) In some *HST* imaging visits the telescope was oriented such that diffraction spikes from this star overlapped the position of one or more of the SN Refsdal source positions. In the template images, this impacted only image S4 in the F125W filter. To resolve this we generated a special set of templates that excluded those spike-contaminated template images. These slightly shallower templates were used only to gather photometry for image S4.

For epochs with SN Refsdal present, when the telescope orientation led to diffraction spike contamination of one or more of the four images, we cannot simply discard the contaminated observations. Instead, these images were processed through an additional "despiking" procedure to enable less biased photometric measurements. The diffraction spike pattern on *HST* in the WFC3-IR detector is close to symmetric about both axes, so we could generate a rough model for the contaminating spike by centering the image on the star, and then rotating the difference image by 90° in a clockwise direction. We then remove the spike by subtracting the rotated difference image from the original unrotated version, which effectively removes the majority of the contaminating flux at the Refsdal source locations, as shown in Figure 1. We examined modifications to this approach, such as using a 180° or 270° rotation, or a median of three rotated versions. We found that a single 90° clockwise rotation was most effective, and alternatives did not substantially affect the resulting photometry. By inserting and recovering artificial point sources in the spike-contaminated regions, we have confirmed that this despiking procedure does increase the statistical uncertainty of our photometric measurements, but results in a net improvement by reducing the potential for systematic biases.

2.1. Photometry

For our photometric measurements on the difference images, we used the PythonPhot[26] software package (Jones et al. 2015), developed in part for use on other high-z SNe

---

[23] http://glass.astro.ucla.edu and https://archive.stsci.edu/prepds/glass
[24] http://www.stsci.edu/hst/campaigns/frontier-fields

[25] http://drizzlepac.stsci.edu
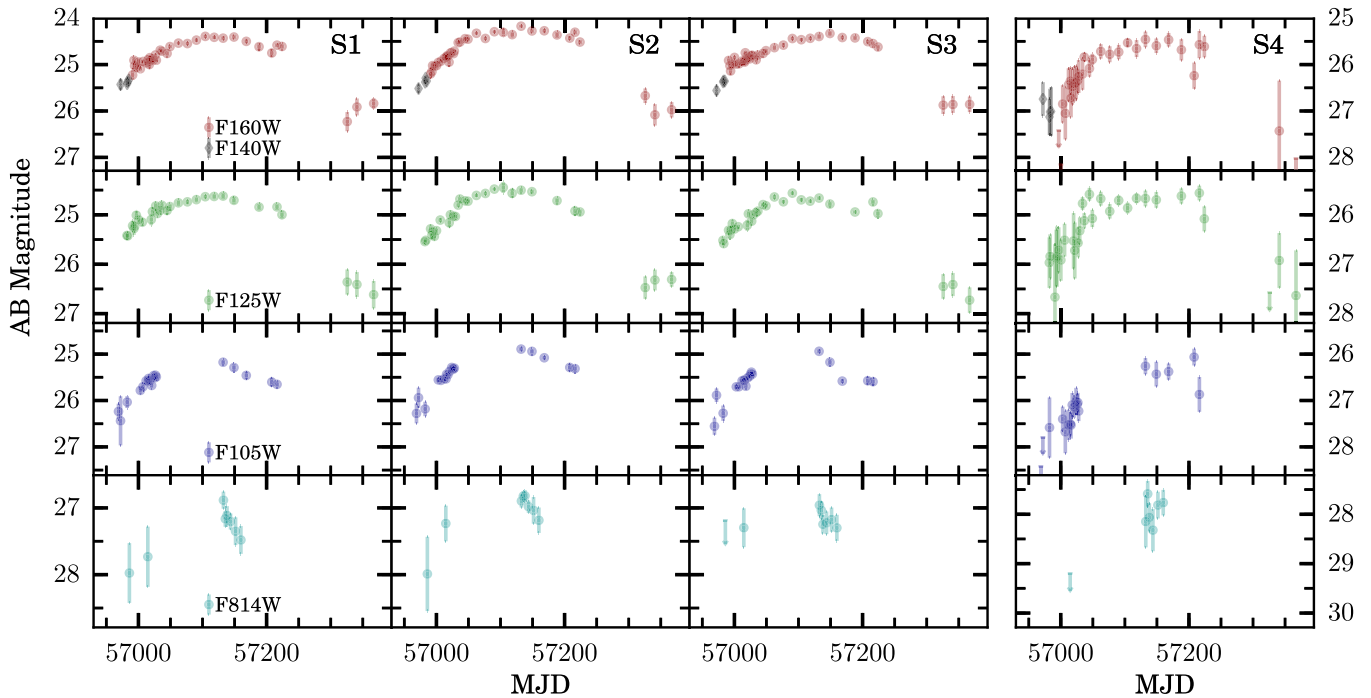[26] https://github.com/djones1040/PythonPhot

Figure 2. The observed light curves of SN Refsdal images S1–S4. Each panel shows AB magnitudes plotted against observer-frame days. Each column shows the light curve of one of the SN Refsdal images, S1–S4 from left to right, with F160W and F140W in the top row, then F125W, F105W, and F814W separately in rows two to four, respectively.

observed with *HST* (e.g., Rodney et al. 2015a, 2015b). We measured the flux using a point-spread function (PSF) fitting procedure similar to DAOPHOT (Stetson 1987). As in Rodney et al. (2015a, 2015b), we used an empirical PSF model generated from the *HST* imaging of the G2V standard star P330E, observed in a series of *HST* calibration programs.

To estimate photometric uncertainties in each image, we planted and extracted 500 fake stars (copies of the model PSF) at random locations in the region defined by the sky annulus. We measured the flux density of each fake star with PSF fitting and fit a normal distribution to the histogram of recovered fake star flux densities. We then define two components of the uncertainty from the best-fit normal distribution. First, δf_μ is the difference between the measured mean of the distribution and the value of the flux assigned to all the planted fake stars, which is typically very close to zero but can give an estimate of systematic biases in cases where the sky region around the SN is strongly contaminated by diffraction spikes or residuals from the lensing galaxy. Second, δf_σ is the standard deviation of the best-fit normal distribution, and gives an empirical measure of the uncertainty due to sky noise and detector read noise. A final uncertainty component is δf_ν, the Poisson noise error, computed from the total count of photons measured in the PSF fit or the aperture. These are added in quadrature to give the total uncertainty, where $\delta f^2 = \delta f_\mu^2 + \delta f_\sigma^2 + \delta f_\nu^2$.

These photometric measurements are reported in Table 4 and Figure 2 shows the resulting multi-band light curves for images S1–S4. In Table 4 we mark with an asterisk any photometric measurement that was collected from despiked images prepared as in Figure 1.

2.2. Photometric Classification

Details of the classification of SN Refsdal are presented in a companion paper (Kelly et al. 2015a), which draws on all available photometric and spectroscopic data. There we present spectroscopy of SN Refsdal from *HST* and the VLT, taken roughly 75 days apart in the rest-frame. In all spectra, we identify broad Hα emission consistent with a Type II SN at the redshift of the host galaxy. This classification is reinforced by the slow rise to peak brightness (over ~150 days) observed in all four SN Refsdal light curves (see Figure 2). This light curve shape is most consistent with the well-studied archetype, SN 1987A, a peculiar Type II SN that is understood to be the explosion of a blue supergiant star. Although no single line of evidence provides a definitive classification of the SN sub-type, the preponderance of evidence indicates that SN Refsdal is a Type II SN, and most likely a member of the rare SN 1987A-like sub-class.

3. LIGHT CURVE TEMPLATE FITTING

In recent years, high precision time delays have been measured for a growing sample of multiply-imaged quasars, using increasingly sophisticated observations and techniques (e.g., Fassnacht et al. 2002; Kochanek 2006; Courbin et al. 2011; Eulaers et al. 2013; Tewes et al. 2013b) Measuring time delays from lensed SNe like SN Refsdal should in principle be much simpler than is typically the case for lensed quasars. The time variation of quasars is stochastic, being driven by essentially random events on the accretion disk of the central supermassive black hole, so the intrinsic shape of a quasar light curve can not be known a priori. As such, quasar time delay methods must adopt a very flexible function to describe the light curve, and rely on purely empirical constraints (e.g., Liao et al. 2015; Tewes et al. 2013a). In contrast, for lensed SNe it should typically be possible to classify the SN based on both photometry and spectroscopy, and then identify a well-matched SN light curve template. In that case, a template-based approach for time delay

measurements will almost always be preferable to using a flexible function, as the template provides a strong informative prior for the intrinsic light curve shape.

In this section we derive our first measurements of relative time delays and magnification ratios from the SN Refsdal data using light curve template matching. This approach makes the assumption that the SN Refsdal light curve shape can be well approximated by a light curve model based on well-studied SNe from the nearby universe. A second set of time delay and magnification measurements using different assumptions will be presented in Section 4.

3.1. Template Fitting Method

As described in Section 2.2, SN Refsdal's slow rise to maximum light is clearly inconsistent with the rise times for the most common SN types (e.g., Ia, Ib/c, II-P, and II-L; see Li et al. 2011). For completeness, we also evaluated the quality of fit from these normal SN classes, using a library of 42 templates drawn from the Supernova Analysis software suite (SNANA Kessler et al. 2009). Unsurprisingly, the photometric peculiarity of SN Refsdal is born out quantitatively, as our light curve models for these normal SN sub-classes are highly incompatible with the data, returning a χ^2 per degree of freedom $\nu\chi_\nu^2 \gg 50$. These models are therefore formally rejected, and the remainder of our analysis focuses on the peculiar SN 1987A-like sub-class, which provides the best matches to the observed shape of the SN Refsdal light curve.

We constructed templates based on the prototype SN 1987A itself (Hamuy & Suntzeff 1990), and also using the 87A-like events SN 1998A (Woodings et al. 1998; Pastorello et al. 2005), SN 2000cb (Hamuy et al. 2001; Kleiser et al. 2011), SN 2006V, and SN 2006au (Taddia et al. 2012), and SN 2009E (Pastorello et al. 2012). All of these template SNe have well-sampled light curve coverage in the B, V, and R bands extending over at least 80 days in the rest-frame. As detailed below, each template was corrected to appear as it would at the redshift of SN Refsdal and through the observed HST passbands. We then implemented a Bayesian parameter estimation framework (similar to Rodney & Tonry 2009, 2010) to simultaneously find the color corrections needed to match each model to the SN Refsdal data, as well as the best-fit time delays and magnifications for all four SN Refsdal images.

The model light curves are defined using

$$m(\lambda', t') = M(\lambda, t) + K(\lambda, t; \lambda') + C_\lambda, \qquad (1)$$

where the time t is the rest-frame age relative to the date of peak brightness in the rest-frame R band, MJD_{pk} (a free parameter in the model). The rest-frame time is dilated to the observer frame using $t' = t\,(1+z)$. The model apparent magnitude in an observed passband at given observed age, $m(\lambda', t')$, is governed by a model absolute magnitude in a model passband at the model's rest-frame age, $M(\lambda, t)$, corrected to an observed passband with $K(\lambda, t; \lambda')$ (see Strolger et al. 2015, for an example of the applied K-correction). A magnitude shift C_λ is then added as a separate free parameter for each photometric passband, which accounts for both cosmological dimming and any color difference between the model and SN Refsdal (due to dust extinction or intrinsic color differences). Linear interpolation is used to infer model magnitudes between observed points in the template light curves.

To take into account the gravitational lensing effects, we include six more free parameters that are applied as corrections to the observed data: three time shifts Δt_i and three achromatic magnitude shifts Δm_i that give the time delays and magnifications of the three sources $i = $ (S2, S3, S4) relative to our reference source S1. The model light curves are then simultaneously compared to all four SN Refsdal sources in the F105W, F125W and F160W bands (rest-frame B, V, and R) to derive a likelihood distribution from each light curve template T_k, using

$$p(\boldsymbol{D}|T_k, \boldsymbol{\theta}) = \prod_i \frac{p(\boldsymbol{\theta})}{\sqrt{2\pi}\,\sigma_i} e^{-(m_{obs}(t_i) - m_k(\lambda_i, t_i))^2/(2\sigma_i^2)}. \qquad (2)$$

Here θ denotes the set of 10 free parameters: date of peak brightness, three relative time delays Δt, three lensing magnitude shifts Δm_i, and three "color" shifts C_λ for the three photometric bands used. We use flat priors $p(\theta)$ for all of the parameters, with time shifts allowed over the range $[-100, 100]$, and magnitude shifts in the range $[-3, 3]$. The product is over all observed epochs t_i, and the uncertainty for each epoch σ_i^2 is a quadratic sum of the photometric uncertainty and a "model uncertainty" of 0.15 mag. This term accounts for the fact that there is no perfect light curve analog available within our limited template library, due to the diversity and rarity of SN 1987A-like events (e.g., Pastorello et al. 2012; Taddia et al. 2012). The choice of 0.15 mag in all epochs follows (Rodney & Tonry 2009), where that value was found to approximately compensate for a similarly sparse library of core collapse light curve templates. Including this model uncertainty removes our ability to independently test for goodness of fit, so in this section we are making the strong assumption that the templates, blurred by this error term, are a good model for the observations.

Alternatively, we can set the model uncertainty to zero, which effectively assumes that all possible SN 1987A-like light curve shapes are represented within our set of six viable templates. In this case, all the best-fitting models return $\chi_\nu^2 \gtrsim 7$, indicating that the models are poor representations of the intrinsic SN light curve shape. Nevertheless, with no model uncertainty term we still find that the range of time delay and magnification estimates are consistent with the values derived using 0.15 mag for the model uncertainty.

To sample the likelihood distributions defined by Equation (2) over the ten-dimensional parameter space, we use the Markov Chain Monte Carlo ensemble sampling tools from the emcee software package (Foreman-Mackey et al. 2013).

3.2. Template Fitting Results

A summary of the template fitting results is given in Table 1. To derive a single set of measurements from these models, we use the approach of Bayesian Model Averaging (BMA; Leamer 1978; Draper 1995; Raftery 1995), which provides a weighted average for each parameter of interest, incorporating the posterior probabilities in the weighting. The BMA posterior mean and variance for each parameter ϕ in θ are given by

$$E[\phi|\boldsymbol{D}] = \sum_k \hat{\phi}_k p(T_k|\boldsymbol{D})$$

$$\mathrm{Var}[\phi|\boldsymbol{D}] = \sum_k (\mathrm{Var}[\phi|\boldsymbol{D}, T_k] + \hat{\phi}_k^2) p(T_k|\boldsymbol{D}) - E[\phi|\boldsymbol{D}]^2$$

$$\qquad (3)$$

Table 1
Time Delay and Magnification Ratio Measurements from SN Light Curve Template Fitting

Model	$p(T_k|\mathbf{D})$	$\mathrm{MJD_{pk}}$	$\Delta t_{\mathrm{S2:S1}}$	$\Delta t_{\mathrm{S3:S1}}$	$\Delta t_{\mathrm{S4:S1}}$	$\mu_{\mathrm{S2}}/\mu_{\mathrm{S1}}$	$\mu_{\mathrm{S3}}/\mu_{\mathrm{S1}}$	$\mu_{\mathrm{S4}}/\mu_{\mathrm{S1}}$
SN 1987A	0.21	$57148.0^{+3.2}_{-1.8}$	$7.58^{+1.59}_{-3.52}$	$5.45^{+1.33}_{-4.13}$	$20.80^{+1.99}_{-4.51}$	$1.127^{+0.031}_{-0.031}$	$1.019^{+0.028}_{-0.028}$	$0.331^{+0.012}_{-0.012}$
SN 1998A	1e-09	$57169.0^{+1.9}_{-2.1}$	$6.99^{+1.40}_{-4.16}$	$5.87^{+1.93}_{-4.98}$	$28.87^{+3.09}_{-2.09}$	$1.159^{+0.032}_{-0.032}$	$1.038^{+0.038}_{-0.038}$	$0.380^{+0.018}_{-0.018}$
SN 2000cb	0.34	$57125.0^{+1.1}_{-2.0}$	$1.98^{+3.31}_{-1.33}$	$-0.41^{+2.74}_{-0.98}$	$19.69^{+2.68}_{-1.17}$	$1.138^{+0.031}_{-0.031}$	$1.000^{+0.028}_{-0.028}$	$0.344^{+0.016}_{-0.016}$
SN 2006 V	0.44	$57143.8^{+0.8}_{-1.8}$	$3.01^{+1.24}_{-1.23}$	$1.03^{+1.70}_{-0.94}$	$28.09^{+2.21}_{-0.70}$	$1.159^{+0.032}_{-0.043}$	$1.009^{+0.028}_{-0.037}$	$0.347^{+0.016}_{-0.016}$
SN 2006au	3e-05	$57161.2^{+0.3}_{-0.1}$	$6.24^{+0.30}_{-0.28}$	$3.33^{+0.35}_{-3.06}$	$31.19^{+0.78}_{-0.87}$	$1.127^{+0.031}_{-0.031}$	$1.000^{+0.037}_{-0.028}$	$0.384^{+0.018}_{-0.018}$
SN 2009E	3e-05	$57160.2^{+1.6}_{-1.5}$	$4.35^{+1.93}_{-2.06}$	$2.97^{+1.75}_{-2.35}$	$21.85^{+3.16}_{-1.98}$	$1.107^{+0.031}_{-0.031}$	$1.000^{+0.028}_{-0.028}$	$0.353^{+0.013}_{-0.016}$
BMA Mean	$\cdots$	57138 ± 10	4 ± 4	2 ± 4	24 ± 5	1.15 ± 0.05	1.01 ± 0.04	0.34 ± 0.02

Note.
Mean value for each parameter, computed using the Bayesian Model Averaging method (see the text for details).
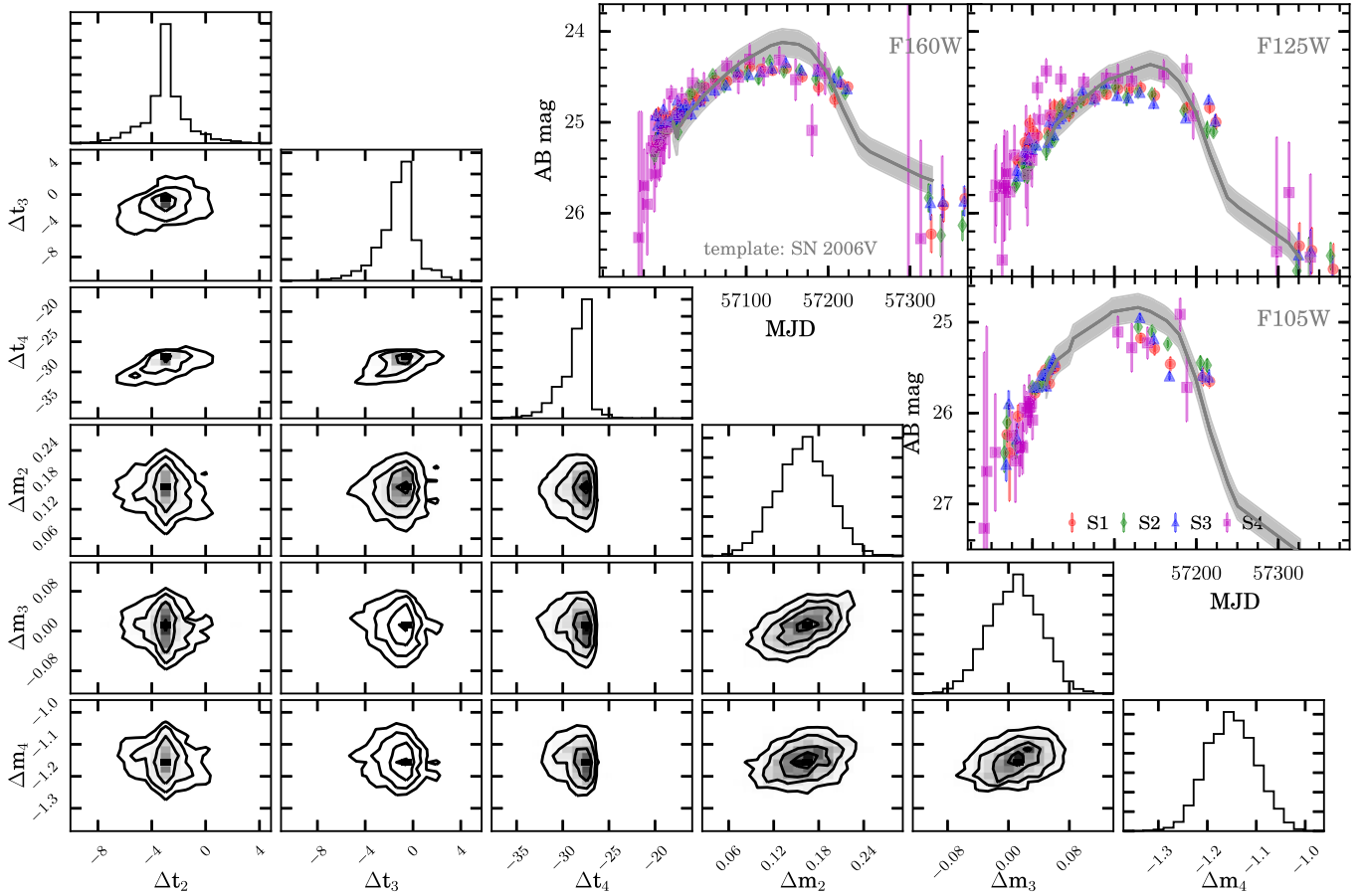

Figure 3. Results of the template fits to the SN Refsdal light curves using the best-matching template, which is the SN 1987A-like Type II SN 2006 V. Three panels in the upper right show the composite light curve from images S1–S4, after applying the time and magnitude shifts that bring S2–S4 into the frame of the reference light curve S1 and maximize the likelihood function (Equation (2)) for the SN 2006 V template. The SN 2006 V template light curve is overplotted as a solid gray line, with the shaded band indicating the assumed 0.15 mag model uncertainty. Panels in the lower left show 2D marginalized probability contours for each of the six fit parameters that define relative time delays and magnifications. Contours are shown at 0.5, 1.0, 1.5 and 2.0σ. Histograms at the top of each column show the 1D marginalized probability distribution for the parameter corresponding to that column.

where $\hat{\phi}_k = E[\phi|D, T_k]$ is the expectation value assuming template T_k is the correct model. The posterior probability values $p(T_k|\mathbf{D})$ (reported in column two of Table 1) are computed by applying Bayes' Theorem with a flat prior $p(T_k)$ for all templates (see, e.g., Hoeting et al. 1999, for further discussion of the BMA method).

Figure 3 shows the maximum likelihood light curve model, which is based on the SN 2006 V template. This template

effectively matches the general character of the SN Refsdal light curve, with a slow rise to maximum followed by a sharp drop. There are, however, notable systematic deviations, such as the sharpness of the peak and the steepness of the drop-off in the F160W band. Figure 3 also plots the 1D and 2D probability distributions for the six free parameters in the model that set the relative time delays and magnifications (i.e., these panels do not show four "nuisance" parameters that set the date of peak brightness and the SN color). The templates SN 1987A and SN
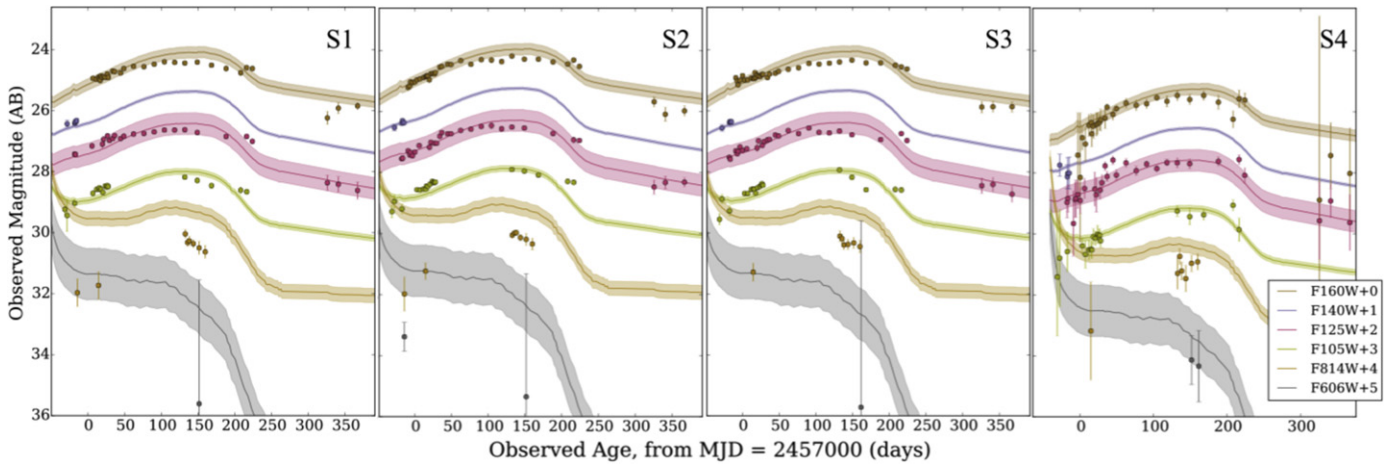
Figure 4. Observed light curves for the four SN Refsdal images, S1–S4, as labeled, with vertical offsets as indicated in the legend. Curves show the "maximally constrained" template fit, based on the SN 1987A light curves, color corrected to match a peak blackbody temperature of $T = 5300\,K$. Shaded bands indicate uncertainties from template photometric error and uncertainties in the K-correction.

2000cb provide a similar quality of fit to the data, as reflected in their posterior probabilities in Table 1. The time delay and magnification measurements from other models are broadly consistent, although they are substantially less effective at matching the observed photometry.

The composite mean and uncertainty for each parameter—derived from the BMA method (Equation (3))—are reported in the final row of Table 1. This locates the time of peak for image S1 in the F160W band—arbitrarily selected as our reference light curve—to be $MJD_{pk} = 57138$ (2015 April 26), with an uncertainty of ± 10 days. This parameter in particular should be taken with caution, as the best-fitting model shown in Figure 3 is clearly mis-representing the behavior of SN Refsdal near peak brightness.

3.3. Maximally Constrained Model Fit

The fitting procedure described above is limited insofar as it only employs the rest-frame B, V, and R bands. This is a necessary restriction, as the majority of known SN 1987A-like events do not have extensive observations in rest-frame ultraviolet bands that could be used to fit the F814W, F606W, and F435W observations of SN Refsdal. Furthermore, we have left out an important physical constraint, by allowing the color of each template to be completely free, with a separate parameter C_λ shifting each bandpass independently.

To derive a more physically constrained fit to the SN Refsdal light curves, we followed a prescription similar to that shown by Taddia et al. (2012). We adopt the SN 1987A template for this purpose, as it has the most complete coverage in both wavelength and time, and in Section 3.2 we have seen that it is one of three models that can provide an adequate fit to the rest-frame BVR light curve. We first correct the SN 1987A template for host extinction using $E(B - V) = 0.16$ mag (Fitzpatrick & Walborn 1990) and $R_V = 4.5$ (De Marchi & Panagia 2014) as appropriate for 30 Doradus, the star-forming region within the Large Magellanic Cloud where SN 1987A exploded. We then "zero out" the peak colors of the SN 1987A model by applying separate magnitude shifts in each band at the epoch of the rest-frame R band peak brightness (i.e., forcing $B - V \equiv 0$, $V - R \equiv 0$, etc.). We apply the same shift across the light curve, so that the color evolution of the model still matches the

observed color curve of SN 1987A. We then apply a temperature-based color-correction following

$$m_\lambda - V = -2.5 \log\left[\int B_\lambda(T) S_\lambda\, \lambda\, d\lambda\right] + C_{\lambda,AB} - C_{V,AB}, \quad (4)$$

where the magnitude correction for a given passband, m_λ, is defined by the color-correction relative to the V-band, which is the product of the filter throughput, S_λ, and the Planck function at a given temperature, $B_\lambda(T)$. $C_{\lambda,AB}$ and $C_{V,AB}$ define the system response through the given passband and the V-band.

Figure 4 shows the maximum likelihood fit of this revised SN 1987A model to the observed SN Refsdal data in all bands (including the optical bands from ACS/WFC). From this fit we find that the color–temperature of SN Refsdal around maximum light is $T \approx 5300\,K$. This is consistent with the range of temperatures (~ 4000–$9000\,K$) seen for other SNe 1987A-like events at the same epoch (Pastorello et al. 2005, 2012; Taddia et al. 2012). As discussed in Kelly et al. (2015a), the SN Refsdal rest-frame optical colors (with $B - V \approx 0.5$ mag) are slightly bluer than most SN 1987A-like objects. This blue color is most consistent with more luminous SN 1987A analogs, such as SNe 2006 V and 2006au (Taddia et al. 2012).

This modified SN 1987A model is not as good a fit to the data as the best-fit SN 2006 V model with completely free color terms from Section 3.2. This may be the result of poorer matching of SN 1987A-like SEDs to smooth blackbody spectra in bluer wavelengths, as was seen in Taddia et al. (2012), presumably due to atmospheric line blanketing in these events. There is a solid physical basis for the idea that color differences in the class of SN 1987A-like explosions stem primarily from differences in their photospheric temperature. These temperature differences may arise from different explosion energies driven by a diversity in progenitor masses. However, temperature differences alone can not explain the wide diversity of this SN sub-class, indicating that other physical parameters also strongly influence the color and color evolution.

Keeping these caveats in mind, we can nevertheless derive alternative constraints on gravitational lensing parameters from this color-temperature-corrected SN 1987A model. We find again a set of broadly consistent time delay and magnification estimates: $\Delta t_{S2:S1} = -1.0 \pm 1.2$ days, $\Delta t_{S3:S1} = 0.4 \pm 1.1$ days, and $\Delta t_{S4:S1} = 14.1 \pm 2.9$ days; $\mu_{S2}/\mu_{S1} =$

1.14 ± 0.07, $\mu_{S3}/\mu_{S1} = 1.05 \pm 0.07$,   and $\mu_{S4}/\mu_{S1} = 0.34 \pm 0.09$. The uncertainties here reflect only statistical error estimates, inferred from the photometric and model errors.

4. TIME DELAY MEASUREMENTS WITH FLEXIBLE LIGHT CURVE MODELS

As Figure 3 shows, even the best-fit template-based model shows systematic residuals and does not provide a good representation of the observed data. SN Refsdal is not quite a clone of other observed 87A-like Type II SNe. Thus as a second approach for measuring the time delays between the four Refsdal sources, we used a series of flexible light curve models (splines and polynomials) to represent the underlying light curve shape. By adopting these free-form curves in place of the rigid SN light curve templates, we can derive time delays that are agnostic about the classification of SN Refsdal. This allows for the possibility that SN Refsdal is unlike any of the available SN templates, and we may thereby avoid a systematic bias that could be introduced by assuming an incorrect light curve shape. The cost of this more flexible approach is that we lose the physical/empirical priors on the light curve shape and color that a well-matched template would afford. This second approach is therefore much closer to the methodology typically used for measuring lensed quasar time delays (e.g., Tewes et al. 2013a; Liao et al. 2015), where there is no way to apply an informative prior for the intrinsic light curve shape.

The first year of the SN Refsdal light curve is fundamentally very simple: a slow rise to a broad peak, and a gradual decline. To approximate this intrinsic light curve shape with the simplest possible functional form, we start by adopting a low-order Chebyshev polynomial of the first kind, which gives the magnitude of image i in band j at time t as

$$
\begin{aligned}
m_{i,j}(t) = c_0 &+ c_1 T_1(t + \Delta t_i) \\
&+ c_2 T_2(t + \Delta t_i) + [...] + \Delta m_i
\end{aligned} \tag{5}
$$

where the coefficients c_n are free parameters in the model and the polynomial components T_n are defined by the Chebyshev recurrence relation $T_{n+1}(x) = 2x T_n(x) - T_{n-1}(x)$, with $T_0(x) = 1$ and $T_1(x) = x$. The effects of gravitational lensing are represented in separate time shifts Δt_i and magnitude shifts Δm_i for each of the four images, though we fix image S1 as our reference point by setting $\Delta t_1 \equiv 0$ and $\Delta m_1 \equiv 0$ (i.e., we are only fitting for *relative* time delays and magnifications). Note that the intrinsic color of the SN is accounted for by having a separate polynomial fit to each band. When fitting this model to the SN Refsdal data, we use only the F160W, F125W and F105W bands, for which we have sufficient data to effectively constrain the peak of the light curve independently in each band. All together, this means a second-order polynomial model has 15 free parameters: three polynomial coefficients in each of three bands to define the light curve shape and color, and six parameters for the time delays and magnifications of S2-S4 relative to S1. Increasing the degree of the polynomial by one adds three additional free parameters (one new polynomial coefficient for each passband).

To allow for more complex intrinsic light curve shapes, we also evaluated cubic spline fits, using one, two and three internal spline knots at fixed positions along the time axis. The knots were arbitrarily set to MJD = [57150] for the single-knot

spline, [57000, 57200] for the two-knot spline, and [57000, 57100, 57200] for the three-knot spline.

To check whether this arbitrary knot placement could bias the measurement of lensing parameters, we also evaluated the use of more sophisticated algorithms for optimizing the number and location of internal cubic spline knots. We used tools from the PyCS (Tewes et al. 2013a)[27] and SNPy software packages (Burns et al. 2011, 2015).[28] The PyCS program was orginally developed by the Cosmological Monitoring of Gravitational Lenses (COSMOGRAIL) collaboration (Eigenbrod et al. 2005)[29] for the measurement of gravitational lensing time delays from single-filter quasar light curve sets. To collapse the multi-band SN Refsdal light curves into a form suitable for use with PyCS, we used a combination of the F125W and F160W observations, which were collected concurrently in almost every epoch, and together have the most complete and well-sampled coverage of the Refsdal light curve. The SNPy software suite was developed by the Carnegie Supernova Project (Hamuy et al. 2006) to provide general purpose SN light curve fitting tools, especially for Type Ia SNe. The SNPy spline fitting tools automatically enforce a restriction on the flexibility of the spline curve model by using the "hyperspline" algorithm (Thijsse et al. 1998). This method is designed to find a spline representation for noisy data without allowing the spline to follow every noise feature. This is achieved by starting with an interpolating spline (one knot at every observed data point) and iteratively removing knots to optimize the Durbin–Watson statistic (Durbin & Watson 1950, 1951), which tests for serial correlation in the least squares regression. When applying SNPy we defined a separate spline curve for the F160W, F125W, and F105W bands. Once again we found that the added flexibility and optimal placement of spline knots in these two packages did not lead to any significant changes in the inferred time delays or magnifications.

4.1. Flexible Curve Fitting Results

Table 2 reports the time delays and magnitude shifts of the sources S2–S4 relative to S1, derived from the polynomial and spline fits described above. The best model, as measured by the total posterior probability $p(M_k|\boldsymbol{D})$, is a cubic spline with a single internal knot. Figure 5 shows this best-fit spline model, along with marginalized posterior probability distributions for each of the lensing parameters. All other models except the second-order Chebyshev polynomial provide a similar quality of fit to the data. Furthermore, the relative time delays and magnifications inferred from all models are quite consistent. As in Section 3, we use the BMA method to combine the parameter estimates from all of these models, deriving the values given in the final row. These measurements are fully consistent within the uncertainties with the values inferred from SN light curve template fitting.

To explore whether the light curve can be effectively described with fewer parameters, we also evaluated a set of "minimalist" polynomial and spline models. In this case we assume that all bands (F160W, F125W, and F105W) have the same intrinsic light curve shape, meaning that they would reach peak brightness at the same epoch. This is not a good assumption for SN 1987A-like explosions, which tend to reach

---

[27] http://obswww.unige.ch/~tewes/cosmograil/public/pycs/index.html
[28] http://csp.obs.carnegiescience.edu/data/snpy
[29] http://cosmograil.epfl.ch

Table 2
Time Delay and Magnification Ratio Measurements from Polynomial and Spline Fits

Model	$p(M_k|D)$	$\mathrm{MJD_{pk}}$	$\Delta t_{S2:S1}$	$\Delta t_{S3:S1}$	$\Delta t_{S4:S1}$	μ_{S2}/μ_{S1}	μ_{S3}/μ_{S1}	μ_{S4}/μ_{S1}
Chebyshev, deg=2	0.02	57136.7	8.3 ± 1.5	-2.1 ± 1.6	32.5 ± 4.4	1.17 ± 0.01	0.99 ± 0.01	0.39 ± 0.01
Chebyshev, deg=3	0.13	57133.8	6.1 ± 1.5	0.2 ± 1.4	30.6 ± 3.6	1.17 ± 0.01	1.00 ± 0.01	0.39 ± 0.01
Chebyshev, deg=4	0.19	57133.5	6.7 ± 1.4	1.0 ± 1.3	24.7 ± 3.4	1.17 ± 0.01	1.00 ± 0.01	0.38 ± 0.01
Chebyshev, deg=5	0.19	57133.4	6.5 ± 1.4	1.0 ± 1.3	24.2 ± 3.4	1.17 ± 0.01	1.00 ± 0.01	0.38 ± 0.01
Spline, 1 knots	0.19	57132.7	7.4 ± 1.4	1.0 ± 1.3	29.0 ± 3.9	1.17 ± 0.01	1.01 ± 0.01	0.38 ± 0.01
Spline, 2 knots	0.15	57131.5	8.5 ± 1.3	0.4 ± 1.4	31.3 ± 3.9	1.18 ± 0.01	1.00 ± 0.01	0.39 ± 0.01
Spline, 3 knots	0.13	57127.0	7.4 ± 1.2	-0.1 ± 1.2	23.6 ± 3.4	1.17 ± 0.01	1.00 ± 0.01	0.38 ± 0.01
BMA mean	$\cdots$	57132 ± 3	7 ± 2	0.6 ± 2	27 ± 6	1.17 ± 0.02	1.00 ± 0.01	0.38 ± 0.02

Note.
The date of peak brightness inferred for the reference curve S1 in the F160W band. Note that this is not a singular parameter in these light curve models, but rather is accounted for in the coefficients of the polynomial or spline curve functions. We report here the value derived from locating the peak of the maximum likelihood model.
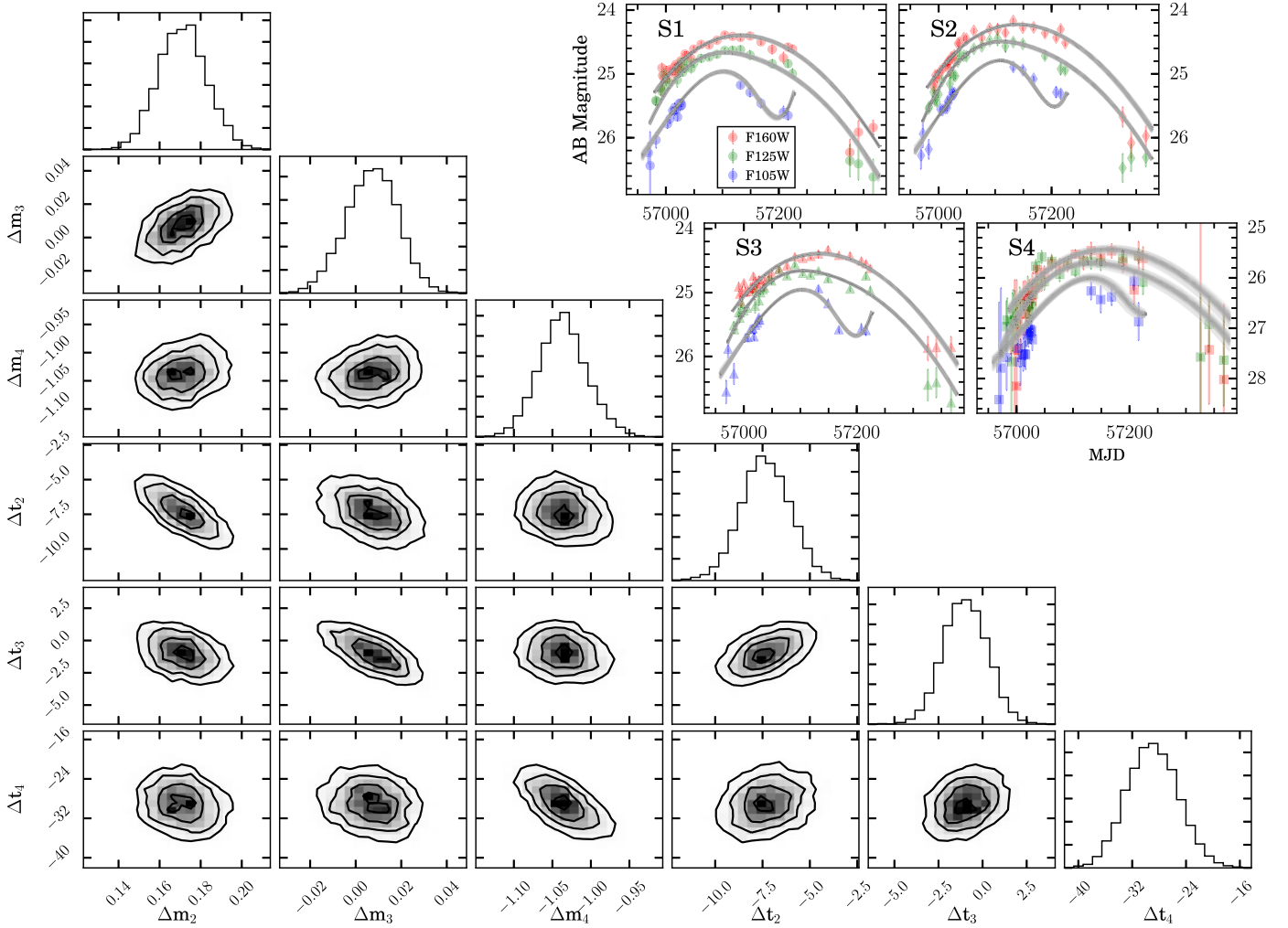

Figure 5. Results of fitting the SN Refsdal light curves using a cubic spline with a single internal knot. Four panels in the upper right show the spline fits to the observed data, with each panel showing a single Refsdal image (S1–S4, as labeled). The F160W, F125W, and F105W data are plotted together as red, green, and blue points, respectively. Overlaid gray curves show the optimized spline functions, which are fit simultaneously to all four images. Each gray band comprises a sample of 100 curves drawn randomly from the MCMC chain, to give an indication of the range of variation in the shapes of curves that have parameters close to their optimal values. Panels in the lower left show marginalized 2D posterior probability distributions for each of the 6 lensing parameters (magnitude shifts and time delays relative to the reference light curve, S1). As in Figure 3, contours mark the 0.5, 1.0, 1.5, and 2.0σ confidence regions, and histograms at the top of each column show 1D marginalized posterior probability distributions.

peak brightness much earlier in bluer bands (e.g., Pastorello et al. 2012; Taddia et al. 2012). Using both the net posterior probability and the Bayesian Information Criterion (Schwarz 1978) as metrics to evaluate the fitness of these simpler models, we found that we consistently get a better fit when using the more complex alternative. That is, using three separate
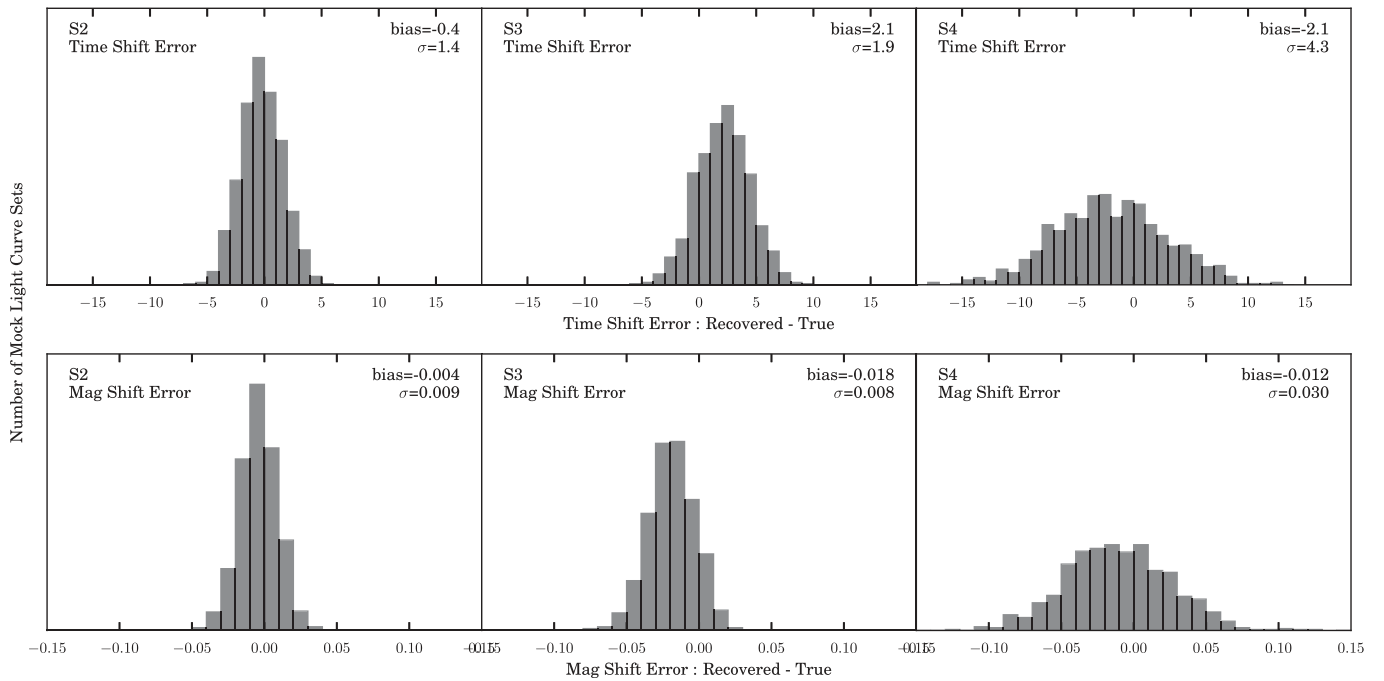
Figure 6. Histograms from the mock light curve analysis, using a single-knot cubic spline to represent the intrinsic light curve shape. Three panels in the top row illustrate the distributions of time delay errors (in days) as the recovered time delay minus the input (true) time delay for each mock light curve. The bottom row shows the same difference, but for the magnitude shift (in mags) accounting for the lensing magnification of each mock light curve.

polynomials or splines to describe each band independently gives a better representation of the light curve shape, regardless of the degree of the polynomial or the number of knots in the spline. Expanding the input data to include the F140W and F814W observations does not change these conclusions.

4.2. Uncertainty Estimates from Mock Light Curves

The uncertainty estimates given in Tables 1 and 2 reflect the statistical uncertainties due to photometric measurement error. The error in the BMA mean also accounts for some of the systematic errors that may be introduced by adopting an inappropriate functional form to describe the intrinsic light curve shape. As an alternative means to estimate systematic uncertainties, we follow the algorithm of Tewes et al. (2013a), generating 1000 mock light curves from the best-fit polynomial and spline curve fits, after introducing artificial time delays and magnifications drawn from uniform distributions about the best-fit values. We then fit the mock curves with the same procedures described above and measure the difference between the input and recovered values of the time delays and magnifications.

Each mock light curve is constructed with observations at the actual dates and in the same filters where SN Refsdal was observed. For each mock data point we start with the exact magnitude predicted by the best-fit model (polynomial or spline) for that epoch. Then we add a magnitude offset $\Delta m_{\mathrm{noise}}$ drawn from a normal distribution. If the observed data point was within 1.5σ of the model, then we set the standard deviation of that normal distribution equal to the photometric error of that observed data point. For data points where the difference between the observation and the best-fit model is $>1.5\sigma$, we set the standard deviation equal to the residual between the observation and the model, and we preserve the sign of the residual when applying the offset. This ensures that

our mock light curves have a similar mix of random offsets due to photometric error and correlated systematic offsets due to possible mismatches between the true light curve shape and the best-fit model. However, the runs of data points with similar residuals appear at different phases relative to the light curve peak for each mock light curve, because we have introduced random time delay shifts for the mock S1–S4 events.

Figure 6 shows histograms of the "recovered-minus-actual" time delays and magnitude shifts from this mock light curve analysis. Fitting a Gaussian to each histogram, we report the mean and standard deviation in each panel. We then adopt the standard deviation of the Gaussian fit as the statistical uncertainty for each lensing parameter. A non-zero value in the mean reflects a potential systematic bias in the fitting procedure, and we include this in our total uncertainty estimate, adding in quadrature to the statistical uncertainty.

5. SUMMARY AND DISCUSSION

The light curve of SN Refsdal after one full year of *HST* imaging observations reveals this to be an unusual object. The extremely broad shape of the light curve in rest-frame optical bands is incompatible with the relatively rapid rise of normal Type I and Type II SNe. The observed shape has distinct similarities with SN 1987A, and in a companion paper we have concluded that SN Refsdal is most likely a member of the peculiar Type II sub-class defined by the SN 1987A proto-type (Kelly et al. 2015a).

In this work we have explored two methods for measuring time delays from the SN Refsdal light curves. We first used a set of six SN light curve templates with shapes similar to SN 1987A, allowing the template colors to float as free parameters to accommodate the very blue rest-frame optical colors of SN Refsdal. We then adopted flexible polynomial functions as an alternative description of SN Refsdal's intrinsic light curve

Table 3
Summary of Time Delay and Magnification Ratio Measurements

Parameter	Template LC Fits[a]	Polynomial Curve Fits[a]
MJD_{pk}	57138 ± 10 days	57132 ± 4 days
$\Delta t_{S2:\ S1}$	4 ± 4 days	7 ± 2 days
$\Delta t_{S3:\ S1}$	2 ± 5 days	0.6 ± 3 days
$\Delta t_{S4:\ S1}$	24 ± 7 days	27 ± 8 days
μ_{S2}/μ_{S1}	1.15 ± 0.05	1.17 ± 0.02
μ_{S3}/μ_{S1}	1.01 ± 0.04	1.00 ± 0.01
μ_{S4}/μ_{S1}	0.34 ± 0.02	0.38 ± 0.02

Note.
[a] Mean values from the BMA combinations given in Tables 1 and 2, with uncertainties updated to incorporate the mock light curve analysis of Section 4.2.

shape. We find that the SN can be well represented by a very simple set of low-order Chebyshev polynomials or cubic splines, and from these fits we derive consistent results for the relative time delays and magnifications. The results from these complementary time delay measurement strategies are summarized in Table 3.

Each method independently provides measurements of the time delays for S2, S3, and S4 (relative to S1) with a precision of ± 2 to 8 days. This level of precision is promising, as it suggests that a similarly cadenced monitoring campaign could deliver a relative precision of $\sim 1\%$ on the time delay to the next image, SX, expected to reach peak brightness approximately one year after the observed S1 peak (Kelly et al. 2015b). Similarly, the magnification ratios relative to S1 are measured to better than 3% precision for S2 and S3, and better than 10% for S4.

5.1. Comparison to Model Predictions

As the first multiply-imaged SN ever seen, SN Refsdal has garnered great interest from the lens modeling community. The presentation of the SN Refsdal discovery by Kelly et al. (2015c) included a first analysis of the SN lensing, using the "light traces mass" (LTM) modeling approach (Broadhurst et al. 2005; Zitrin et al. 2009). Within a week of the initial announcement, two other teams produced revised strong lensing models for the MACS J1149.6+2223 cluster (Oguri 2015; Sharon & Johnson 2015). These were tuned to give more accurate predictions for the magnifications and time delays of the Einstein Cross images, as well as for the SX image. Subsequently, other lens modeling groups have produced updated models, taking advantage of improved imaging and spectroscopic data on this field to generate models using better catalogs of multiply-imaged galaxies, as well as better precision in reproducing those strong-lensing constraints (Diego et al. 2016; Jauzac et al. 2016). Most recently, Treu et al. (2016) presented a coordinated effort from five lens modeling groups to produce new MACS J1149.6+2223 lens models with a set of collectively vetted strong lensing constraints. The individual results from each group are being published separately, providing details on each group's modeling approach (Grillo et al. 2015; Kawamata et al. 2015; J. M. Diego et al. 2016, in preparation; K. Sharon et al. 2016, in preparation; A. Zitrin et al. 2016, in preparation). A primary goal of this effort was

to examine how different choices and assumptions in the modeling methodology can affect the predictions of magnifications and time delays for a strongly lensed source such as Refsdal.

Figure 7 presents a comparison of our measured magnification ratios and time delays for images S1–S4 against all published lens model predictions available as of 2015 December 15. The earliest models, first posted within a week of the SN Refsdal discovery, are shown as squares (Oguri 2015; Sharon & Johnson 2015).[30] The recently updated model of Jauzac et al. (2016) is plotted with circles, and the set of models from the Treu et al. (2016) model comparison program (Die-a, Gri-g, Ogu-a, Ogu-g, Sha-a, Sha-g, Zit-g, Zit-c) are plotted with trangles.

For all but two of these models, the time delay and magnification ratio measurements presented here were not available for use as input strong-lensing constraints, or as an intermediate check to guide the model development. Thus, the comparison of our measurements against these models is effectively a true blind test of the predictive power of each model. The two exceptions are the Zit-c model and the Jau15.2 model, which were updated after the initial release of these SN measurements. These "unblind" models are marked by asterisks and plotted with a black outline in Figure 7.

The Zit-c model is a corrected version of the Zit-g model, updated after the SN measurements presented in this work were known. The Zit-c model does not use any measured time delays or magnification ratios as input constraints, but does use the previously known positions of the S1–S4 images (as do all the models evaluated here). The key change in the Zit-c model is that it allows the total mass of the lens galaxy to be a free parameter, which ensures that its critical curves pass through the Einstein cross (S1–S4), as required by those positional constraints. For computational efficiency, both versions of this model were computed using a relatively low resolution grid ($0\farcs065$ pix^{-1}), and the predictions of the model with respect to the S1–S4 time delays are therefore limited by this grid scale. The main source of difference between the predictions of the Zit-g and Zit-c models (which both use the LTM approach) and the predictions by other (analytic) models is the different parametrization of the mass distribution. See Treu et al. (2016) for further details.

The Jau15.1 model is the version presented by Jauzac et al. (2016; first appearing in arXiv eprint v3), which adopts the model-predicted positions for the SN sources S1–S4 when computing the time delays. These model-predicted source locations do not match the observed locations, leading to biases in the predicted time delays. After the initial release of the SN measurements, a revision of Jauzac et al. (2016) (arXiv eprint v4) introduced the model labeled here as Jau15.2. This version instead uses the *observed* locations of images S1–S4 and adopts an analytic approach to compute the time delays that effectively forces S1–S4 to be spatially and temporally coincident at the source plane. These unblind time delays are presented as $\Delta t_{CATS\text{-}src}$ by Jauzac et al. (2016).

Although there are a few predictions that are discrepant at $>2\sigma$, the overall agreement between model predictions and our measurements is quite good. Unlike the situation for SN HFF14Tom, discussed in Rodney et al. (2015a), there is no

---

[30] Note that Oguri (2015) did not report uncertainties for the Ogu15 model.

Table 4
Photometry of the Four Einstein Cross Images of SN Refsdal

Filter	MJD	Exp. Time (s)	S1 (AB mag)	S2 (AB mag)	S3 (AB mag)	S4 (AB mag)
F160W	56990.9	1159	25.22 ± 0.11*	25.20 ± 0.10	24.91 ± 0.07	⋯
F160W	56993.0	1159	24.91 ± 0.07*	25.15 ± 0.11	25.02 ± 0.08	⋯
F160W	56994.0	1159	25.03 ± 0.07*	25.02 ± 0.07	25.14 ± 0.08	⋯
F160W	56996.8	1159	25.06 ± 0.10*	25.04 ± 0.09	24.99 ± 0.08	>25.96
F160W	57000.1	2318	24.95 ± 0.07*	25.02 ± 0.06	24.85 ± 0.04	>26.23
F160W	57003.0	5512	25.09 ± 0.04	24.95 ± 0.04	24.99 ± 0.07	26.85 ± 0.40
F160W	57007.1	5512	24.93 ± 0.03	24.93 ± 0.05	24.98 ± 0.06	27.05 ± 0.56
F160W	57012.0	5512	24.96 ± 0.05	24.87 ± 0.04	24.92 ± 0.06	26.44 ± 0.36
F160W	57015.0	5512	24.89 ± 0.05	24.84 ± 0.04	24.93 ± 0.07	26.35 ± 0.30
F160W	57015.9	5512	24.95 ± 0.04	24.83 ± 0.05	24.95 ± 0.06	26.72 ± 0.42
F160W	57017.0	1015	25.00 ± 0.09	24.81 ± 0.05	24.75 ± 0.05	⋯
F160W	57017.9	5512	24.85 ± 0.04	24.82 ± 0.04	24.90 ± 0.06	26.48 ± 0.39
F160W	57019.7	1421	24.90 ± 0.07	24.95 ± 0.09	24.82 ± 0.06	⋯
F160W	57021.9	6121	24.88 ± 0.05	24.75 ± 0.03	24.85 ± 0.06	26.41 ± 0.38
F160W	57023.9	5512	24.90 ± 0.03	24.74 ± 0.04	24.88 ± 0.06	26.38 ± 0.36
F160W	57026.6	11023	24.79 ± 0.03	24.76 ± 0.04	24.80 ± 0.06	26.34 ± 0.26
F160W	57027.9	5512	24.91 ± 0.04	24.71 ± 0.03	24.80 ± 0.05	26.26 ± 0.36
F160W	57034.0	1159	24.69 ± 0.04	24.53 ± 0.04	24.90 ± 0.06	26.20 ± 0.34
F160W	57036.6	1159	24.71 ± 0.06	24.51 ± 0.05	24.80 ± 0.07	25.84 ± 0.09
F160W	57044.7	1209	24.76 ± 0.05	24.44 ± 0.04	24.76 ± 0.05	26.08 ± 0.18
F160W	57049.2	1209	24.61 ± 0.04	24.45 ± 0.04	24.70 ± 0.05	25.89 ± 0.12
F160W	57062.4	1209	24.54 ± 0.03	24.33 ± 0.03	24.63 ± 0.06	25.72 ± 0.16
F160W	57076.4	1209	24.55 ± 0.04	24.44 ± 0.04	24.58 ± 0.05	25.78 ± 0.17
F160W	57090.4	1209	24.47 ± 0.04	24.29 ± 0.03	24.44 ± 0.04	25.73 ± 0.15
F160W	57104.3	1209	24.39 ± 0.04	24.31 ± 0.09	24.46 ± 0.04	25.53 ± 0.09
F160W	57118.2	1209	24.41 ± 0.04	24.35 ± 0.06	24.43 ± 0.04	25.65 ± 0.18
F160W	57132.1	759	24.43 ± 0.04	24.17 ± 0.03	24.39 ± 0.06	25.46 ± 0.17
F160W	57149.1	759	24.41 ± 0.05	24.27 ± 0.05	24.33 ± 0.05	25.59 ± 0.16
F160W	57168.3	1209	24.50 ± 0.05	24.27 ± 0.04	24.42 ± 0.05	25.46 ± 0.15
F160W	57188.2	1209	24.61 ± 0.08*	24.36 ± 0.07	24.43 ± 0.05	25.68 ± 0.23*
F160W	57208.1	1209	24.75 ± 0.09*	24.43 ± 0.04	24.51 ± 0.05	26.24 ± 0.28*
F160W	57216.2	1209	24.58 ± 0.04	24.30 ± 0.05	24.55 ± 0.05	25.57 ± 0.29*
F160W	57224.0	1209	24.61 ± 0.05	24.51 ± 0.05	24.62 ± 0.05	25.61 ± 0.23*
F160W	57325.8	1259	26.23 ± 0.21	25.67 ± 0.16	25.87 ± 0.19	>25.84
F160W	57341.0	1259	25.91 ± 0.18	26.08 ± 0.23	25.86 ± 0.18	27.43 ± 1.08
F160W	57367.1	1259	25.84 ± 0.11	25.97 ± 0.17	25.86 ± 0.17	28.02 ± 1.50
F140W	56972.1	1168	25.43 ± 0.11	25.52 ± 0.09	25.56 ± 0.10	26.74 ± 0.36
F140W	56982.4	15935	25.41 ± 0.07	25.38 ± 0.05	25.37 ± 0.05	27.13 ± 0.39
F140W	56983.2	5212	25.37 ± 0.07	25.29 ± 0.05	25.35 ± 0.06	27.02 ± 0.50
F140W	56984.9	5212	25.33 ± 0.06	25.35 ± 0.05	25.36 ± 0.05	27.01 ± 0.52
F125W	56982.2	10423	25.42 ± 0.08	25.54 ± 0.05	25.52 ± 0.06	26.97 ± 0.51
F125W	56983.4	10423	25.43 ± 0.07	25.52 ± 0.06	25.58 ± 0.06	26.84 ± 0.45
F125W	56990.9	1159	25.22 ± 0.10	25.28 ± 0.07	25.31 ± 0.06	27.67 ± 1.07*
F125W	56993.0	1159	25.28 ± 0.13	25.41 ± 0.08	25.40 ± 0.07	26.88 ± 0.58*
F125W	56994.0	1159	25.23 ± 0.10	25.34 ± 0.08	25.32 ± 0.08	26.85 ± 0.61*
F125W	56996.7	1159	25.01 ± 0.09	25.44 ± 0.07	25.18 ± 0.07	26.71 ± 0.46*
F125W	57000.1	2318	25.10 ± 0.08	25.32 ± 0.05	25.29 ± 0.06	26.92 ± 0.42*
F125W	57005.9	5212	25.15 ± 0.08	25.11 ± 0.04	25.24 ± 0.05	26.51 ± 0.33
F125W	57020.0	406	25.10 ± 0.12	25.16 ± 0.10	25.21 ± 0.11	26.53 ± 0.56*
F125W	57021.8	812	24.94 ± 0.09	25.00 ± 0.06	24.98 ± 0.07	26.72 ± 0.57*
F125W	57026.9	1623	24.82 ± 0.07	25.03 ± 0.07	25.13 ± 0.07	26.57 ± 0.29
F125W	57029.6	5212	24.97 ± 0.05	25.03 ± 0.04	25.00 ± 0.03	26.32 ± 0.27
F125W	57033.9	1159	24.91 ± 0.06	24.80 ± 0.04	24.95 ± 0.07	25.77 ± 0.13
F125W	57036.6	1159	24.81 ± 0.07	24.68 ± 0.05	24.93 ± 0.05	26.12 ± 0.15
F125W	57044.7	1209	24.91 ± 0.08	24.72 ± 0.05	24.79 ± 0.04	25.59 ± 0.13
F125W	57049.2	1209	24.83 ± 0.06	24.73 ± 0.04	24.82 ± 0.05	26.07 ± 0.17
F125W	57062.4	1209	24.76 ± 0.06	24.61 ± 0.03	24.64 ± 0.04	25.67 ± 0.14
F125W	57076.4	1209	24.74 ± 0.04	24.57 ± 0.03	24.74 ± 0.04	25.93 ± 0.15
F125W	57090.4	1209	24.69 ± 0.05	24.48 ± 0.03	24.56 ± 0.04	25.71 ± 0.10
F125W	57104.3	1209	24.63 ± 0.04	24.45 ± 0.11	24.70 ± 0.04	25.86 ± 0.11
F125W	57118.2	1209	24.63 ± 0.06	24.56 ± 0.09	24.72 ± 0.04	25.67 ± 0.10
F125W	57132.1	759	24.62 ± 0.07	24.50 ± 0.05	24.66 ± 0.04	25.67 ± 0.16
F125W	57149.1	759	24.71 ± 0.07	24.53 ± 0.05	24.78 ± 0.05	25.70 ± 0.16

Table 4
(Continued)

Filter	MJD	Exp. Time (s)	S1 (AB mag)	S2 (AB mag)	S3 (AB mag)	S4 (AB mag)
F125W	57188.2	1209	24.84 ± 0.07	24.71 ± 0.06	24.94 ± 0.04	25.62 ± 0.15*
F125W	57216.2	1209	24.84 ± 0.06	24.92 ± 0.09*	24.74 ± 0.05*	25.56 ± 0.15*
F125W	57224.0	1209	25.00 ± 0.08	24.94 ± 0.05*	24.97 ± 0.07*	26.08 ± 0.25
F125W	57325.8	1159	26.36 ± 0.25	26.47 ± 0.23	26.45 ± 0.25	>26.22
F125W	57340.9	1159	26.41 ± 0.25	26.32 ± 0.21	26.41 ± 0.23	26.92 ± 0.55
F125W	57367.1	1159	26.62 ± 0.28	26.31 ± 0.15	26.73 ± 0.26	27.63 ± 0.91
F105W	56968.9	1218	26.24 ± 0.17	26.28 ± 0.21	26.55 ± 0.18	>26.60
F105W	56972.1	356	26.44 ± 0.53	25.94 ± 0.22	25.89 ± 0.15	>26.18
F105W	56982.4	5612	26.03 ± 0.13	26.18 ± 0.16	26.27 ± 0.17	27.58 ± 0.65
F105W	57003.0	5612	25.78 ± 0.07	25.56 ± 0.04	25.70 ± 0.05	27.40 ± 0.27
F105W	57007.1	5612	25.70 ± 0.08	25.57 ± 0.05	25.70 ± 0.05	27.68 ± 0.46
F105W	57012.0	5612	25.58 ± 0.06	25.53 ± 0.03	25.58 ± 0.04	27.52 ± 0.33
F105W	57015.0	5612	25.58 ± 0.06	25.53 ± 0.04	25.56 ± 0.05	27.51 ± 0.25
F105W	57015.9	5612	25.52 ± 0.06	25.43 ± 0.04	25.57 ± 0.04	27.53 ± 0.29
F105W	57017.9	5612	25.58 ± 0.05	25.45 ± 0.04	25.69 ± 0.05	27.10 ± 0.26
F105W	57020.7	5612	25.68 ± 0.06	25.37 ± 0.04	25.51 ± 0.04	27.14 ± 0.23
F105W	57023.9	5612	25.46 ± 0.04	25.29 ± 0.04	25.47 ± 0.04	27.02 ± 0.29
F105W	57025.7	5612	25.51 ± 0.06	25.31 ± 0.03	25.43 ± 0.03	27.08 ± 0.23
F105W	57026.6	5612	25.46 ± 0.06	25.32 ± 0.03	25.39 ± 0.04	27.04 ± 0.21
F105W	57027.9	5612	25.49 ± 0.06	25.30 ± 0.03	25.43 ± 0.04	27.23 ± 0.22
F105W	57132.1	759	25.18 ± 0.05	24.89 ± 0.05	24.94 ± 0.05	26.26 ± 0.17
F105W	57149.1	759	25.29 ± 0.08	24.94 ± 0.05	25.17 ± 0.07	26.43 ± 0.27
F105W	57168.3	1209	25.46 ± 0.08	25.08 ± 0.04	25.58 ± 0.04	26.38 ± 0.17
F105W	57208.1	1209	25.60 ± 0.10	25.29 ± 0.08	25.57 ± 0.09	26.06 ± 0.18*
F105W	57216.3	2412	25.65 ± 0.07	25.31 ± 0.07	25.59 ± 0.08*	26.87 ± 0.38*
F814W	56985.9	4768	27.98 ± 0.45	27.99 ± 0.56	>27.43	⋯
F814W	57014.7	10616	27.73 ± 0.45	27.23 ± 0.27	27.30 ± 0.29	>27.58
F814W	57132.6	9826	26.88 ± 0.13	26.90 ± 0.10	26.96 ± 0.17	28.15 ± 0.52
F814W	57135.6	14538	27.16 ± 0.12	26.82 ± 0.07	27.04 ± 0.14	27.59 ± 0.26
F814W	57138.2	19652	27.11 ± 0.14	26.82 ± 0.08	27.25 ± 0.14	28.06 ± 0.29
F814W	57143.8	29880	27.20 ± 0.12	26.98 ± 0.09	27.22 ± 0.17	28.32 ± 0.43
F814W	57151.6	14943	27.35 ± 0.21	27.04 ± 0.20	27.18 ± 0.18	27.82 ± 0.27
F814W	57159.9	10092	27.48 ± 0.21	27.18 ± 0.19	27.30 ± 0.20	27.77 ± 0.26
F606W	56985.8	4542	>27.89	28.37 ± 0.47	>27.75	⋯
F606W	57151.6	14724	>27.99	>27.74	>28.00	29.14 ± 0.80
F606W	57161.4	10092	>27.89	>27.90	>27.65	>28.07
F435W	57131.7	4744	27.67 ± 0.52	>27.53	>27.71	>27.61
F435W	57138.2	9488	27.88 ± 0.48	>27.51	>27.64	>27.63

Note. Magnitudes are reported as 3σ lower limits when the measured flux from PSF fitting was less than the flux uncertainty. Empty entries indicate instances where the PSF fitting did not converge. Asterisks mark values measured on images processed with the "despiking" algorithm illustrated in Figure 1.

indication of a consistent systematic bias among all models for any of the time delays or magnifications. In the case of SN Refsdal, the Einstein Cross configuration is dominated by a single (galaxy) lens, which is a very different lensing regime than for a SN like HFF14Tom on the outskirts of the strong-lensing region of a cluster-scale lens.

The overall agreement between model predictions and observed time delays is an encouraging indication of the accuracy of this current generation of well-vetted models. Furthermore, all of the most up-to-date models (Grillo et al. 2015; Jauzac et al. 2015; Kawamata et al. 2015; Treu et al. 2016) agree that the date of peak brightness for the reappearance of SN Refsdal in image SX should occur within 1–1.5 years. That prediction is also broadly in agreement with earlier models based on a more limited set of input data (Diego et al. 2015; Kelly et al. 2015c; Oguri 2015; Sharon & Johnson 2015). A transient source at the expected position of image SX has now been detected, with magnitudes that are consistent with the magnification ratios and time delays predicted by several of these models (Kelly et al. 2015b). As the full light curve of this new image is measured over the coming year, we will soon be able to complete this direct test of those falsifiable model predictions.

The natural experiment afforded by SN Refsdal gives us an opportunity to examine whether a particular modeling strategy or set of input constraints can deliver better estimates of the time delays and magnifications. The SN observations should be particularly useful for identifying subtle systematic biases common to many models, as in Rodney et al. (2015a). To simplify this assessment, we define the "total tension" between
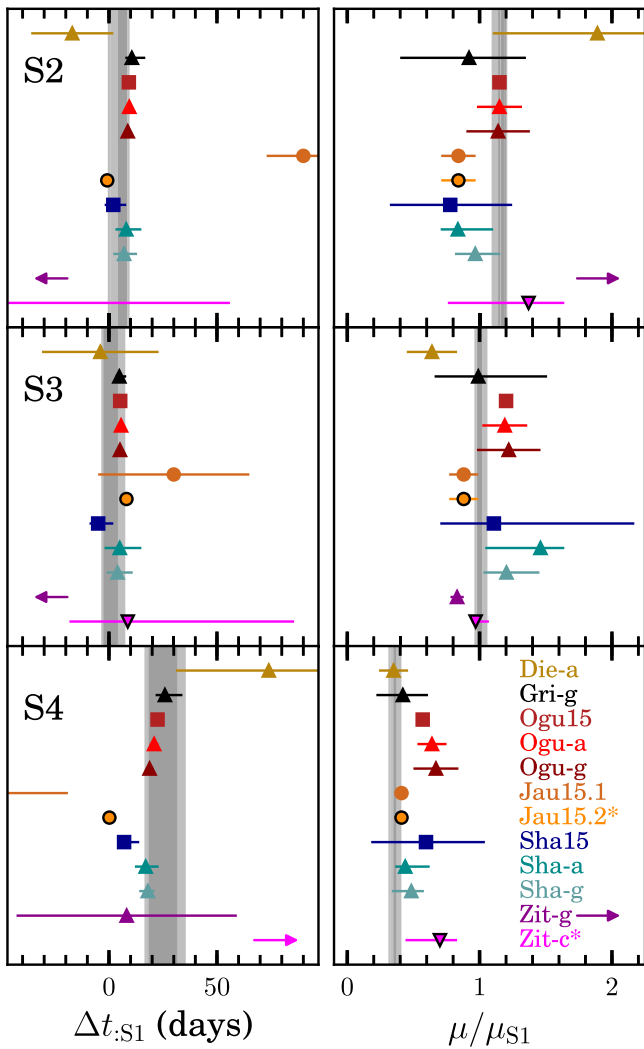
Figure 7. Comparison of the measured time delays against values predicted by lensing models. The three rows of panels present results for images S2, S3, and S4 from top to bottom. Panels in the left column plot time delays and in the right column they show magnification ratios (relative to S1 in both cases). Vertical gray bars indicate the measurements from Table 3. The darker shaded regions indicate overlapping measurements from the two methods presented in Sections 3 and 4. Predicted time delays from published lens models are plotted as colored points, using the key given in the lower right panel (see the text for details). The two "unblind" models are marked with asterisks and plotted with black outlines. Arrows indicate points that fall outside the plotted range.

Figure 8. Tension between models and measurements, plotted against metrics assessing the quantity and quality of input strong lensing constraints. The ordinate in all three panels marks the number of standard deviations separating the model predictions from the observations of time delays and magnification ratios, collapsed to a single value using Equation (6). The abscissa in the left panel is the number of multiply imaged systems used as input constraints for each model, and in the middle panel it is the total number of images (counting all instances of each lensed galaxy), and includes all the distinct knots from the Refsdal host galaxy that are used as input positional constraints. The right panel plots the tension against the fraction of multiply-imaged systems that have a spectroscopic redshift (a crude metric for "quality" of input constraints, following Rodney et al. 2015b). Symbols and colors are as in Figure 7.

a given model and the SN Refsdal measurements as

$$\tau = \sum_{i=2-4} \frac{(\Delta t_i(\mathrm{obs}) - \Delta t_i(\mathrm{mod}))^2}{(\sigma_{t_i}^2(\mathrm{obs}) + \sigma_{t_i}^2(\mathrm{mod}))}$$
$$+ \frac{\left(\frac{\mu_i}{\mu_1}(\mathrm{obs}) - \frac{\mu_i}{\mu_1}(\mathrm{mod})\right)^2}{(\sigma_{\mu_i}^2(\mathrm{obs}) + \sigma_{\mu_i}^2(\mathrm{mod}))} \quad (6)$$

Figure 8 plots this total tension against three metrics that quantify the input strong-lensing constraints used by each model: the number of multiple-image systems, the total number of images (including knots within the SN Refsdal host), and the fraction of multiply-imaged galaxies that have a spectroscopic redshift. Rodney et al. (2015a) performed a similar comparison, using the absolute magnification measurement from a lensed Type Ia SN to test the accuracy of 17 lens models for the

cluster Abell 2744. That analysis suggested that simply increasing the *quantity* of strong lensing constraints did not in and of itself lead to a more accurate magnification prediction. This SN Refsdal test reinforces that suggestion, as shown in the left and middle panels of Figure 8: the models using the greatest number of multiply-imaged systems tend to have a greater total tension with observations, and the same is true for those models with the greatest total number of images and knots from the SN Refsdal host galaxy.

Rodney et al. (2015a) found evidence that the *quality* of input lensing constraints is mildly correlated with successful model predictions. The rightmost panel of Figure 8 again provides some support for this idea: the models that are most accurate in predicting the SN Refsdal time delays and magnifications all have a large fraction of strong-lensing constraints with spectroscopic redshifts (the same crude "quality" metric that was used by Rodney et al. 2015a). Two of the model families from the Treu et al. (2016) comparison are particularly informative in this analysis. The Sharon et al. model series (Sha15, Sha-a, Sha-g) and the Oguri et al. model series (Ogu15, Ogu-a, Ogu-g) were each generated by the same modeling team, using the same modeling toolkit, with the same basic model assumptions. In Figure 8 these sequences are plotted with connecting lines, and both follow the trend of increasing accuracy (lower tension) as the spec-z fraction increases. As such, these sequences provide an especially clean indication that the quality of strong-lensing constraints is a key ingredient for model accuracy.

When a lensed SN is not available for empirical tests of lens models, it would be tempting to determine the best possible magnification or time delay predictions from a set of independent models using a "wisdom of the crowd" approach. For example, one might use a median or an average from a set

of models that includes one contribution from every modeling team. In the case of SN Refsdal, using such a method to combine model predictions would in fact deliver an accurate and precise estimate of the lensing time delays and magnifications. However, this is not ideal. As discussed by, e.g., Treu et al. (2016), the dispersion and bias of the predictions could be overestimated by incorrect assumptions in some of the models, or they could be underestimated if all models suffer from the same incorrect assumption. A more fruitful approach is to actually try to understand why some models perform better than others, and what assumptions are justified and what are not.

Additionally, one should note that the trends displayed in Figure 8 are fairly weak, and the number of models being tested is quite small. Thus, although these comparisons can certainly be informative, we should be very cautious about making inferences regarding the global fitness of these lens models based on their precision or accuracy in predicting the measurable properties of a single SN. The value and limitations of such comparisons are discussed in more detail by Rodney et al. (2015a) and Treu et al. (2016), and the interested reader is referred there.

5.2. Microlensing

Throughout this work, we have ignored the possible effects of *microlensing*: small-scale gravitational lensing perturbations due to massive objects along the light path of any one image in the quartet. Instead, we have assumed that each of the S1–S4 light curves is only affected by a single magnification factor that is static over the duration of the light curve. It is, however, quite possible that SN Refsdal is affected by either of two types of microlensing. The first is the traditional form of microlensing that has been observed in lensed quasars (e.g., Kochanek 2004). In this case, the effective transverse motion of stars in the lensing galaxy changes the intervening lensing potential and causes fluctuations in the light curve on a timescale of months or years (e.g., Wyithe & Turner 2001; Schechter & Wambsganss 2002; Schechter et al. 2004). Dobler & Keeton (2006) describe the second form of microlensing that is unique to lensed SNe. The SN light passing through the lensing plane is distorted by a web of lensing potentials formed by all the intervening stars in the lensing galaxy. As the photosphere of the SN expands, it intersects a larger section of this complex web, which can result in microlensing fluctuations that affect the light curve on timescales of weeks to months. Analysis of such microlensing features in a lensed SN light curve could potentially be used to make inferences about the mass fraction and projected spatial density of the stellar population in the lensing galaxy (Kolatt & Bartelmann 1998; Dobler & Keeton 2006).

Dobler & Keeton (2006) find that most microlensed SN should be expected to exhibit fluctuations of ~0.2 mag on short timescales (days to weeks) and shifts of >0.5 mag on long timescales (months). These distortions will significantly limit the precision that can be achieved in measurement of their time delays. However, the microlensing environments modeled by Dobler & Keeton (2006) are for a SN lensed by a single isolated galaxy, and the situation for SN Refsdal may be less dire, as the added shear and convergence from the MACS J1149.6+2223 cluster potential may result in light paths through the lensing galaxy that are farther from the galactic nucleus and therefore intersect relatively sparse stellar environments, even accounting for the intracluster light.

The SN Refsdal light curve is not as finely sampled or covering as long a baseline as is now typical for multiply-imaged quasars, which in some cases are monitored over 5–10 years (e.g., Courbin et al. 2011; Eulaers et al. 2013; Tewes et al. 2013b). However, a microlensing analysis of SN Refsdal would have several distinct advantages relative to the quasar observations. First, the intrinsic SN light curve with a single broad peak is much simpler than the stochastically varying quasar light curves. This means that deviations from a smoothly varying light curve can in principle be immediately attributed to microlensing variations—assuming that systematic uncertainties in the photometry are well controlled. The data set for SN Refsdal also includes valuable color information from simultaneous observations in multiple pass-bands, which has not been commonly collected for many long-baseline quasar light curves. Gravitational lensing effects are achromatic in general, so a microlensing event should always be independently detected in multiple bands. A limited chromatic dependence can be generated for microlensing if the effective source size depends on wavelength (Kochanek 2004), but this will generally be negligible for the small wavelength difference between optical and IR bands typically used for SN observations. Finally, a microlensing analysis of SN Refsdal will soon be able to take advantage of the light curve of the fifth image (SX), which is lensed only by the MACS J1149.6+2223 cluster potential and not also directly affected by any cluster member galaxies. The light curve of SX, though expected to be substantially fainter than S1–S4, should be relatively free of microlensing, since its light path does not pass through any individual cluster galaxy, and it should be less affected by intracluster stars or the outskirts of the MACS J1149.6+2223 Brightest Cluster Galaxy.

A full analysis of possible microlensing in the SN Refsdal light curves is beyond the scope of this work. However, we can make a preliminary assessment of whether there are any indications of especially strong microlensing events that could bias our time delay and magnification measurements. Figure 9 shows the maximum likelihood model derived from fitting three second-order Chebyshev polynomials to the SN Refsdal light curves in the F160W, F125W, and F105W bands. This model does not have the largest posterior probability among the flexible function models we have investigated. However, it is useful for this analysis because it has very limited flexibility, so it cannot be distorted to accommodate microlensing events at the edges of the observed data as part of the intrinsic light curve shape. Even for this very simple model, the residuals shown in Figure 9 are consistently within ± 0.2 mag for all epochs between MJD ~ 56,950 and 57,250. The final epochs deviate more strongly, but this is likely due to the inability of this simple model to accommodate a change in slope after the peak. Although this is far from a complete analysis, the absence of significant deviations (>0.2 mag) relative to this minimally flexible model gives a preliminary indication that major microlensing events did not affect the SN Refsdal light curves.

5.3. Future Measurements of SN Time Delays

In the near future, we may expect that additional detections of strongly lensed SNe with measurable time delays will be few and far between. For most massive galaxy clusters, the rate of such SNe visible above a magnitude limit of $m_{AB} \sim 27.0$ is
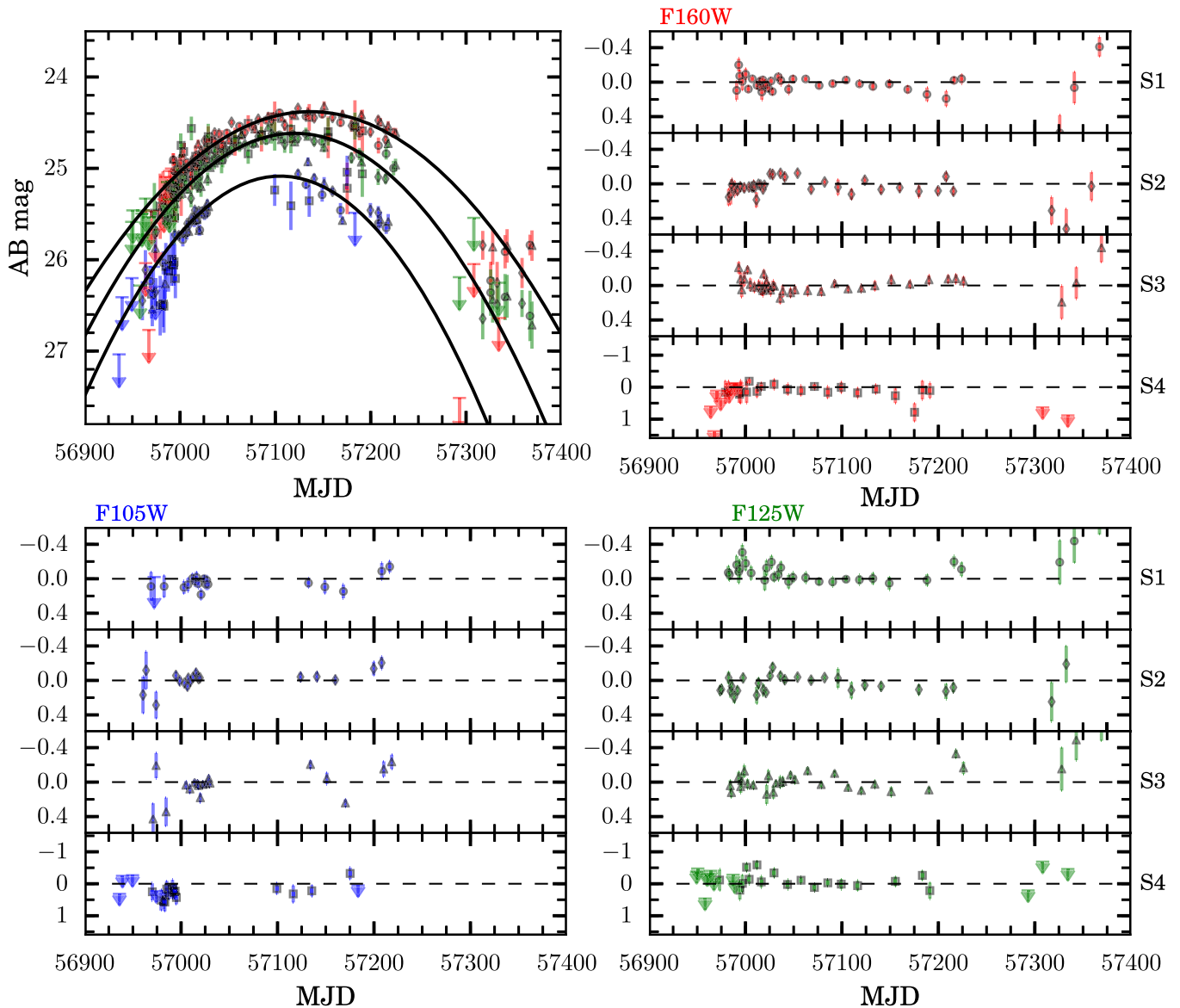
Figure 9. Fit of the SN Refsdal light curves with the simplest viable functional form, a second-order Chebyshev polynomial. Points in the top left panel show the composite S1–S4 light curve data, after applying magnitude and time shifts so that S2, S3 and S4 match the reference light curve, S1. The F160W, F125W, and F105W bands are plotted in red, green and blue, respectively. Overplotted curves show the maximum likelihood polynomial fit, where separate polynomial coefficients are fit for each band. The adjoining panel sets show residuals after subtracting off this best-fit model, separated by image and band.

expected to be on the order of a few SNe per century (Gal-Yam et al. 2002; Bolton & Burles 2003; Sharon et al. 2010; Li et al. 2012). This means that roughly 100 clusters must be regularly monitored with deep imaging in order to have a realistic chance of detecting another event like SN Refsdal in any given year. The Reionization Lensing Cluster Survey (RELICS, *HST*-GO-14096; PI:D. Coe) is an ongoing *HST* program that is a step in that direction, with cadenced IR imaging of 46 strong-lensing galaxy cluster fields. However, each RELICS cluster target is only monitored over a period of 1–2 months, so RELICS and any similar cluster surveys will still have only a small chance of discovering another multiply imaged SN in the near future. Such programs are much more likely to find lensed SNe with significant magnification but no multiple images (Sullivan et al. 2000; Goobar et al. 2009), which can still be useful as a means for discovering distant SNe (Gunnarsson & Goobar 2003; Amanullah et al. 2011) or for

testing cluster lens models (Riehm et al. 2011; Nordin et al. 2014; Patel et al. 2014; Rodney et al. 2015a).

Although we have seen that cluster-lensed SNe like SN Refsdal are valuable as tools for testing cluster dark matter models, another strong motivation for measuring SN time delays is for cosmological constraints. After accounting for the difference in the gravitational potential traversed by the light path for each multiple image (Shapiro 1964), the time delay can be used to directly constrain the Hubble constant (Refsdal 1964) and other cosmological parameters (Linder 2004, 2011; Coe & Moustakas 2009). Distant quasars multiply-imaged by foreground galaxies have been used for such time delay cosmography measurements, providing valuable cosmological constraints that can complement or validate other methods such as Type Ia SNe, baryon acoustic oscillations and the cosmic microwave background (e.g., Saha et al. 2006; Oguri 2007; Coles 2008; Suyu et al. 2010, 2013, 2014; and see Treu & Ellis 2014 for a recent review).

Although rare, cluster-lensed SNe such as SN Refsdal could in principle contribute to future time delay cosmography efforts. However, cluster lenses are generally much more complex than isolated galaxies, which limits the possible precision of time delay cosmography. A more promising avenue for building up a cosmologically useful sample of strongly lensed SNe is through wide-field imaging surveys that can find lensed SNe behind *galaxy-scale* lenses (Oguri & Kawano 2003; Mörtsell et al. 2005). The first example of this came from the Pan-STARRS survey (Kaiser et al. 2010), as SN PS1–10afx (Chornock et al. 2013) was shown to be strongly lensed (though not multiply-imaged) by an intervening galaxy (Quimby et al. 2013, 2014). The Large Synoptic Survey Telescope (Tyson 2002) will dramatically extend this sample, as it is expected to deliver ~130 SNe strongly lensed by foreground galaxies (Oguri & Marshall 2010). The Wide Field Infrared Survey Telescope could substantially increase that sample, particularly at the high redshift end (Holz 2001; Mörtsell et al. 2005; Oguri & Marshall 2010). Time delays for these galaxy-scale lenses are on the order of days or months, not years or decades as is typically the case for cluster-scale lenses, which reduces the timescale over which an observational monitoring campaign needs to operate. Furthermore, the lensing potential of a solitary galaxy—especially an elliptical— is much simpler than a typical galaxy cluster. This means that with a sufficiently rapid observational cadence and concerted lens modeling efforts it will be feasible to use measurements of these lensed SN time delays as cosmographic tools, which will finally realize the original vision of SN Refsdal's namesake (Refsdal 1964).

REFERENCES

Adam, G., Bacon, R., Courtes, G., et al. 1989, A&A, 208, L15
Adamo, A., Östlin, G., Bastian, N., et al. 2013, ApJ, 766, 105
Amanullah, R., Goobar, A., Clément, B., et al. 2011, ApJL, 742, L7
Bolton, A. S., & Burles, S. 2003, ApJ, 592, 17
Broadhurst, T., Benítez, N., Coe, D., et al. 2005, ApJ, 621, 53
Burns, C. R., Stritzinger, M., Phillips, M. M., et al. 2011, AJ, 141, 19
Burns, C. R., Stritzinger, M., Phillips, M. M., et al. 2015, ascl:1505.023
Chornock, R., Berger, E., Rest, A., et al. 2013, ApJ, 767, 162
Coe, D., & Moustakas, L. A. 2009, ApJ, 706, 45
Coles, J. 2008, ApJ, 679, 17
Courbin, F., Chantry, V., Revaz, Y., et al. 2011, A&A, 536, A53
De Marchi, G., & Panagia, N. 2014, MNRAS, 445, 93
Diego, J. M., Broadhurst, T., Chen, C., et al. 2016, MNRAS, 456, 356
Diego, J. M., Broadhurst, T., Molnar, S. M., Lam, D., & Lim, J. 2015, MNRAS, 447, 3130
Dobler, G., & Keeton, C. R. 2006, ApJ, 653, 1391
Draper, D. 1995, J. R. Stat. Soc. B, 57, 45 http://www.jstor.org/stable/2346087
Durbin, J., & Watson, G. S. 1950, Biometrika, 37, 409
Durbin, J., & Watson, G. S. 1951, Biometrika, 38, 159
Ebeling, H., Barrett, E., Donovan, D., et al. 2007, ApJL, 661, L33
Ebeling, H., Edge, A. C., & Henry, J. P. 2001, ApJ, 553, 668
Eigenbrod, A., Courbin, F., Vuissoz, C., et al. 2005, A&A, 436, 25
Eulaers, E., Tewes, M., Magain, P., et al. 2013, A&A, 553, A121
Fassnacht, C. D., Xanthopoulos, E., Koopmans, L. V. E., & Rusin, D. 2002, ApJ, 581, 823
Fitzpatrick, E. L., & Walborn, N. R. 1990, AJ, 99, 1483
Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, PASP, 125, 306
Fruchter, A. S., Hack, W., Dencheva, N., Droettboom, M., & Greenfield, P. 2010, in STSCI Calibration Workshop Proc., BetaDrizzle: A Redesign of the MultiDrizzle Package, ed. S. D. C. Oliveira (Baltimore, MD: Space Telescope Science Institute), 382
Gal-Yam, A., Maoz, D., & Sharon, K. 2002, MNRAS, 332, 37
Goobar, A., Paech, K., Stanishev, V., et al. 2009, A&A, 507, 71
Grillo, C., Karman, W., Suyu, S. H., et al. 2015, ApJ, in press (arXiv:1511.04093)
Gunnarsson, C., & Goobar, A. 2003, A&A, 405, 859
Hamuy, M., Folatelli, G., Morrell, N. I., et al. 2006, PASP, 118, 2
Hamuy, M., Pinto, P. A., Maza, J., et al. 2001, ApJ, 558, 615
Hamuy, M., & Suntzeff, N. B. 1990, AJ, 99, 1146
Hoeting, J. A., Madigan, D., Raftery, A. E., & Volinsky, C. T. 1999, StaSc, 14, 382
Holz, D. E. 2001, ApJL, 556, L71
Huchra, J., Gorenstein, M., Kent, S., et al. 1985, AJ, 90, 691
Jauzac, M., Richard, J., Limousin, M., et al. 2016, MNRAS, 457, 2029
Jones, D. O., Scolnic, D. M., & Rodney, S. A. 2015, ascl:1501.010
Kaiser, N., Burgett, W., Chambers, K., et al. 2010, Proc. SPIE, 7733, 77330E
Karman, W., Grillo, C., Balestra, I., et al. 2015, arXiv:1509.07515
Kawamata, R., Oguri, M., Ishigaki, M., Shimasaku, K., & Ouchi, M. 2015, ApJ, in press (arXiv:1510.06400)
Kelly, P. L., Brammer, G., Selsing, J., et al. 2015a, ApJ, submitted (arXiv:1512.09093)
Kelly, P. L., Rodney, S. A., Treu, T., et al. 2015b, ApJL, submitted (arXiv:1512.04654)
Kelly, P. L., Rodney, S. A., Treu, T., et al. 2015c, Sci, 347, 1123
Kessler, R., Bernstein, J. P., Cinabro, D., et al. 2009, PASP, 121, 1028
Kleiser, I. K. W., Poznanski, D., Kasen, D., et al. 2011, MNRAS, 415, 372
Kochanek, C. S. 2004, ApJ, 605, 58
Kochanek, C. S. 2006, in Saas-Fee Advanced Course 33: Gravitational Lensing: Strong, Weak and Micro, ed. G. Meylan et al. (Berlin: Springer), 91
Kolatt, T. S., & Bartelmann, M. 1998, MNRAS, 296, 763
Leamer, E. E. 1978, Specification Searches: Ad Hoc Inference with Nonexperimental Data (New York: Wiley)
Li, W., Leaman, J., Chornock, R., et al. 2011, MNRAS, 412, 1441
Li, X., Hjorth, J., & Richard, J. 2012, JCAP, 11, 15
Liao, K., Treu, T., Marshall, P., et al. 2015, ApJ, 800, 11
Linder, E. V. 2004, PhRvD, 70, 043534
Linder, E. V. 2011, PhRvD, 84, 123529
Livermore, R. C., Jones, T., Richard, J., et al. 2012, MNRAS, 427, 688
Livermore, R. C., Jones, T. A., Richard, J., et al. 2015, MNRAS, 450, 1812
Monet, D. G., Levine, S. E., Canzian, B., et al. 2003, AJ, 125, 984
Mörtsell, E., Dahle, H., & Hannestad, S. 2005, ApJ, 619, 733
Nordin, J., Rubin, D., Richard, J., et al. 2014, MNRAS, 440, 2742
Oguri, M. 2007, ApJ, 660, 1
Oguri, M. 2015, MNRAS, 449, L86
Oguri, M., & Kawano, Y. 2003, MNRAS, 338, L25
Oguri, M., & Marshall, P. J. 2010, MNRAS, 405, 2579
Pastorello, A., Baron, E., Branch, D., et al. 2005, MNRAS, 360, 950
Pastorello, A., Pumo, M. L., Navasardyan, H., et al. 2012, A&A, 537, A141
Patel, B., McCully, C., Jha, S. W., et al. 2014, ApJ, 786, 9
Postman, M., Coe, D., Benítez, N., et al. 2012, ApJS, 199, 25
Quimby, R. M., Oguri, M., More, A., et al. 2014, Sci, 344, 396
Quimby, R. M., Werner, M. C., Oguri, M., et al. 2013, ApJL, 768, L20
Raftery, A. E. 1995, Sociological Methodology, 25, 111
Refsdal, S. 1964, MNRAS, 128, 307
Riehm, T., Mörtsell, E., Goobar, A., et al. 2011, A&A, 536, A94
Rodney, S. A., Patel, B., Scolnic, D., et al. 2015a, ApJ, 811, 70
Rodney, S. A., Riess, A. G., Scolnic, D. M., et al. 2015b, AJ, 150, 156
Rodney, S. A., & Tonry, J. L. 2009, ApJ, 707, 1064
Rodney, S. A., & Tonry, J. L. 2010, ApJ, 715, 323

Saha, P., Coles, J., Macciò, A. V., & Williams, L. L. R. 2006, ApJL, 650, L17
Schechter, P. L., & Wambsganss, J. 2002, ApJ, 580, 685
Schechter, P. L., Wambsganss, J., & Lewis, G. F. 2004, ApJ, 613, 77
Schmidt, K. B., Treu, T., Brammer, G. B., et al. 2014, ApJL, 782, L36
Schwarz, G. 1978, AnSta, 6, 461
Shapiro, I. I. 1964, PhRvL, 13, 789
Sharon, K., Gal-Yam, A., Maoz, D., et al. 2010, ApJ, 718, 876
Sharon, K., & Johnson, T. L. 2015, ApJ, 800, 26
Smith, G. P., Ebeling, H., Limousin, M., et al. 2009, ApJL, 707, L163
Stetson, P. B. 1987, PASP, 99, 191
Strolger, L.-G., Dahlen, T., Rodney, S. A., et al. 2015, ApJ, 813, 93
Sullivan, M., Ellis, R., Nugent, P., Smail, I., & Madau, P. 2000, MNRAS, 319, 549
Suyu, S. H., Auger, M. W., Hilbert, S., et al. 2013, ApJ, 766, 70
Suyu, S. H., Marshall, P. J., Auger, M. W., et al. 2010, ApJ, 711, 201
Suyu, S. H., Treu, T., Hilbert, S., et al. 2014, ApJL, 788, L35

Taddia, F., Stritzinger, M. D., Sollerman, J., et al. 2012, A&A, 537, A140
Tewes, M., Courbin, F., & Meylan, G. 2013a, A&A, 553, A120
Tewes, M., Courbin, F., Meylan, G., et al. 2013b, A&A, 556, A22
Thijsse, B. J., Hollanders, M. A., & Hendrikse, J. 1998, ComPh, 12, 393
Treu, T., Brammer, G., Diego, J. M., et al. 2016, ApJ, 817, 60
Treu, T., & Ellis, R. S. 2014, arXiv:1412.6916
Treu, T., Schmidt, K. B., Brammer, G. B., et al. 2015, ApJ, 812, 114
Tyson, J. A. 2002, Proc. SPIE, 4836, 10
Woodings, S. J., Williams, A. J., Martin, R., Burman, R. R., & Blair, D. G. 1998, MNRAS, 301, L5
Wyithe, J. S. B., & Turner, E. L. 2001, MNRAS, 320, 21
Yuan, T., Kobayashi, C., & Kewley, L. J. 2015, ApJL, 804, L14
Yuan, T.-T., Kewley, L. J., Swinbank, A. M., Richard, J., & Livermore, R. C. 2011, ApJL, 732, L14
Zitrin, A., & Broadhurst, T. 2009, ApJL, 703, L132
Zitrin, A., Broadhurst, T., Umetsu, K., et al. 2009, MNRAS, 396, 1985