Technische Universität München
School of Natural Sciences

Strongly Lensed Type Ia Supernovae: Time Delays from Machine Learning

Simon Huber

Vollständiger Abdruck der von der TUM School of Natural Sciences zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzender:
Prof. Dr. Andreas Weiler
Prüfende der Dissertation:
1. Prof. Dr. Sherry Suyu
2. Hon.-Prof. Dr. Wolfgang Hillebrandt

Die Dissertation wurde am 26.10.2022 bei der Technischen Universität München eingereicht und durch die TUM School of Natural Sciences am 01.12.2022 angenommen.

Abstract

The Hubble constant H_0 sets the present-day expansion rate of our Universe. However, there is a significant tension between early and late-time probes, pointing towards new physics or unaccounted systematic measurement errors. Therefore, there is a great desire for further independent H_0 investigations, and time-delay cosmography with strongly lensed Type Ia supernovae (LSNe Ia) can provide a direct H_0 measurement. This method is based on three pillars; a time-delay measurement between the multiple images of a variable source, a lens mass model, and a reconstruction of mass perturbations along the line of sight to the source. Time-delay cosmography has been applied successfully to strongly lensed quasars. In contrast to quasars, SNe fade away and reveal the host emission and the lens kinematics, which offers an excellent opportunity to break degeneracies in the lens mass model. Independently, the degeneracies can be broken via the magnification factor. Therefore, especially Type Ia supernovae (SNe Ia) as standardizable candles are very promising if microlensing is weak. Thus, LSNe Ia might outperform lensed quasars.

This thesis focuses mainly on the time-delay measurement of LSNe Ia. We first investigate microlensing uncertainties in color curves of SNe Ia, and find that the distortions are almost negligible in the first three rest-frame weeks after the explosion, in agreement with a similar and independent study in the literature. Still, the practical usage is limited since broadband follow-up from optical to infrared would be required to capture a few color curves with robust features, like peaks. Therefore, we build our methods to measure time delays on *light curves* of SNe Ia, where in principle, one band (ideally the *i* band) can be enough, to achieve a well-measured time delay, but typically, not more than three bands are required. For our novel methods presented in this thesis, we use machine learning aimed at improving the time-delay accuracy.

First, we use two simple machine learning models, a fully connected neural network (FCNN) and a random forest (RF). A test of the two approaches shows that unbiased results are only expected from the RF. However, the tests highlight that the RF is precise enough for practical use and therefore promising to measure time delays in LSNe Ia. Our study shows further that observational noise is the dominant source of uncertainty instead of microlensing. Thus, the depth of the follow-up observation is critical in maximizing the number of LSNe Ia with well-measured time delays. This study provides further guidelines like the choice of bands for future follow-up observations. The disadvantage of the investigated machine learning models is that they must be trained for each LSN Ia separately.

Therefore, we build in a second approach a more general method to measure time delays in LSNe Ia, based on a Siamese long short-term memory network combined with a FCNN (LSTM-FCNN). One can apply this network to future LSNe Ia detected in the Legacy Survey of Space and Time, for which additional follow-up observations are triggered in the *i* band every other day with a mean single-epoch 5σ depth of 24.5 mag. We show that we can keep the bias within 1% for time delays longer than 20 days, similar to the RF and free-knot spline estimator from `PyCS` used in one of our previous works. However, we significantly outperform the RF and `PyCS` in terms of precision. For an example where `PyCS` provides an uncertainty of the time delay measurement around 2.0 days, we achieve 1.6 days with the RF and 0.7 days with the LSTM-FCNN. Further, the LSTM-FCNN also provides a mean precision of around 0.7 days on a comprehensive sample of LSNe Ia . Therefore, the time-delay measurement methods developed in this thesis pave the way for a competitive determination of H_0 with time-delay cosmography using LSNe Ia.

Zusammenfassung

Die Hubble-Konstante H_0 legt die aktuelle Expansionsrate in unserem Universum fest. Messungen aus dem frühen und späten Universum zeigen jedoch signifikante Unstimmigkeiten, was auf neue Physik oder nicht beachtete systematische Fehler hindeutet. Deshalb besteht ein großer Bedarf an weiteren unabhängigen Messungen von H_0. Kosmografie mit Zeitverzögerungsmessungen von Type Ia Supernovae in "starken" Gravitationslinsensystemen (LSNe Ia) ist eine solche Methode um H_0 direkt zu messen. Diese Methode basiert auf drei Säulen: Einem Messen der Zeitverzögerung zwischen den verschiedenen Bildern einer variablen Quelle, einem Modell für die Gravitationslinse und einer Rekonstruktion der Masse entlang der Sichtlinie zur Lichtquelle. Kosmografie mit Zeitverzögerungsmessungen wurde bereits erfolgreich an Quasaren in Gravitationslinsensystemen angewendet. Im Gegensatz zu Quasaren verblassen die SNe mit der Zeit und enthüllen die Ursprungs-Galaxy der SN und die Kinematik der Linse, was einem die Möglichkeit gibt, Entartungen im Modell der Linse zu beseitigen. Unabhängig davon können die Entartungen auch durch den Vergrößerungsfaktor aufgehoben werden. Falls der Mikrolinseneffekt klein ist, sind vor allem SNe Ia deshalb sehr vielversprechend, da sie Standardkerzen sein können. Wegen dieser Vorteile erscheint es realistisch, dass LSNe Ia sogar bessere Resultate als Quasare liefern könnten.

In dieser Dissertation liegt der Fokus hauptsächlich auf den Zeitverzögerungsmessungen von LSNe Ia. Als Erstes untersuchen wir Unsicherheiten die vom Mikrolinseneffekt verursacht werden, wobei wir vor allem Farbkurven von SNe Ia betrachten. Die Untersuchung zeigt, dass der Mikrolinseneffekt in den Farbkurven bis 3 Wochen nach der Explosion, gemessen im Ruhesystem der SN, sehr klein ist. Dies stimmt gut mit einer ähnlichen, aber unabhängigen Studie in der Literatur überein. Das Problem ist aber, dass man mit einer großen Bandbreite vom sichtbaren Licht bis ins Infrarote beobachten müsste, um Farbkurven mit markanten Stellen wie Maxima zu erhalten. Deshalb setzen wir in den folgenden Methoden auf *Lichtkurven* von SNe Ia, wo im Prinzip bereits ein Filter (idealerweise der i-Filter) genug sein kann, um eine Zeitverzögerung präzise zu messen. Typischerweise werden aber nicht mehr als drei Filter benötigt. Für unsere neuartigen Methoden in dieser Dissertation benutzen wir maschinelles Lernen mit dem Ziel, die Genauigkeit von Zeitverzögerungsmessungen zu erhöhen.

Als erstes benutzen wir dafür zwei einfache Modelle des maschinellen Lernens und zwar ein "vollständig verbundenes neuronales Netzwerk" und einen "Random Forest" (RF). Ein Test der beiden Methoden zeigt, dass man nur vom RF Resultate ohne systematischen Fehler erwarten kann. Des Weiteren zeigen die Tests, dass der RF präzise genug ist für die Praxis und somit vielversprechend, um Zeitverzögerungen in LSNe Ia zu messen. Zusätzlich stellen sich die Unsicherheiten vom Beobachten selbst und nicht der Mikrogravitationslinseneffekt als die dominante Fehlerquelle heraus. Deshalb ist die Tiefe der Beobachtung entscheidend, um die Anzahl der LSNe Ia mit präziser Zeitverzögerungsmessung zu erhöhen. Die Studie legt weitere Beobachtungsstrategien nahe, wie z. B. die Wahl bestimmter Filter. Der Nachteil von den untersuchten Modellen ist, dass man ein Modell für jede LSNe

Ia separat trainieren muss.

In einem zweiten Ansatz entwickeln wir deshalb eine allgemeinere Methode, um Zeitverzögerungen für LSNe Ia zu messen. Diese basiert auf einem “Siamese long short-term memory” Netzwerk in Kombination mit einem “vollständig verbundenen neuronalen Netzwerk”. Wir nennen dieses System im Folgenden LSTM-FCNN. Man kann dieses Netzwerk auf zukünftige LSNe Ia anwenden, die im “Legacy Survey of Space and Time” erwartet werden. Dies erfolgt unter der Annahme, dass man die entdeckten Systeme ungefähr jeden zweiten Tag im i-Band mit einer 5σ Tiefe von 24.5 mag beobachtet. Wir zeigen, dass der systematische Fehler kleiner als 1% ist, sobald die Zeitverzögerung größer als 20 Tage ist, ähnlich wie beim RF und dem “free-knot spline estimator” von `PyCS`, den wir in einer unserer früheren Arbeiten verwendet haben. Wenn es aber um den statistischen Fehler geht, ist das LSTM-FCNN dem RF und `PyCS` deutlich überlegen. Für ein Beispiel wo `PyCS` bei einer Zeitverzögerungsmessung einen statistischen Fehler um die 2.0 Tage hat, erreichen wir 1.6 Tage mit dem RF und 0.7 Tage mit dem LSTM-FCNN. Des Weiteren finden wir auch für einen umfangreichen Satz von LSNe Ia eine durchschnittliche Präzision von 0.7 Tage mit dem LSTM-FCNN. Somit ebenen die in dieser Dissertation entwickelten Methoden den Weg für eine kompetitive Bestimmung von H_0 mit LSNe Ia.

Contents

List of Figures

List of Tables

1 Introduction

According to our standard cosmology model, the Big Bang is an event that occurred 13.8 billion years ago when our Universe was in a very compact, dense, and hot initial state from where it expanded to our observable Universe today. In an expanding Universe, distances between gravitationally unbound galaxies increase with time, not because of their proper motion but because of the expanding space between them. This effect can be observed in faraway galaxies' redder appearance. Their light is shifted towards longer wavelengths because the wavelengths get stretched as light travels through the expanding cosmos. Therefore, the present-day expansion rate is one of the fundamental parameters of our Universe and is quantified by the Hubble constant, H_0, which sets the scale and age of our Universe. In Sect. 1.1, we will give a historical introduction to the Hubble constant and summarize current measurements showing a significant tension in H_0 and the need for further independent probes. One such method, which is of particular interest in this thesis, is time-delay cosmography using Type Ia supernovae[1] (SNe Ia) strongly lensed into multiple images by an intervening galaxy or galaxy cluster. We will give an introduction to lensed supernovae in Sect. 1.2. Suppose such new independent techniques confirm the H_0 tension. In that case, it will point towards new physics beyond the successful flat Lambda cold dark matter (ΛCDM) cosmological model coming, e.g., from a different form of dark energy or a modified theory of gravity.

1.1 Expanding universe and the Hubble constant

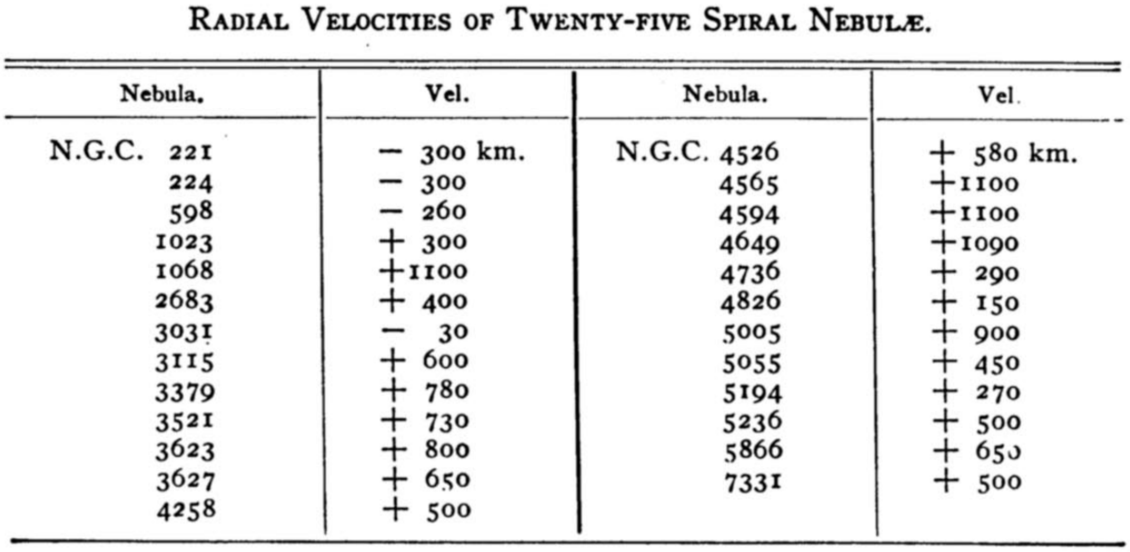
1.1.1 History

The first hint of an expanding Universe was discovered by Slipher (1917), although the observation was not understood at that time. Slipher (1917) measured the redshifts from spectral lines of galaxies, referred to as nebulae at that time, given that the existence of galaxies other than the Milky Way was unknown. In Fig. 1.1, we see the original table from Slipher (1917), where the radial velocities of 25 spiral nebulae are listed. The radial velocities come from the redshift of the observed spectral lines under the assumption of a doppler shift related to the peculiar motion of the nebulae. Interestingly, almost every nebula was redshifted and therefore moving away from us. Furthermore, the mean velocities were about 30 times larger than the average velocity of known stars and much larger than the velocities of any other known class of celestial bodies at that time. This discovery led Slipher (1917) to the conclusion that these nebulae have to be categorized in their own class, which we know today as galaxies. Further, the interpretation of the redshift, just as a doppler shift, is inaccurate, and the observation is indeed a combination of the doppler redshift, from the peculiar motion of the nebulae, and the cosmic redshift due to the expanding space itself, where the cosmic redshift causes that most nebulae are redshifted. The theoretical definition of doppler shift, cosmic redshift, and the combination of both will follow in Sect. 2.1 in

---

[1] A supernova is an explosion of a star, where one distinguishes between different types depending on the progenitor.

Eqs. (2.15), (2.28), and (2.32). However, in 1917 this interpretation was not possible, given that the distances to the nebulae were unknown, and therefore it was not even clear if they were part of the Milky Way.



RADIAL VELOCITIES OF TWENTY-FIVE SPIRAL NEBULÆ.

Nebula.	Vel.	Nebula.	Vel.
N.G.C. 221	− 300 km.	N.G.C. 4526	+ 580 km.
224	− 300	4565	+1100
598	− 260	4594	+1100
1023	+ 300	4649	+1090
1068	+1100	4736	+ 290
2683	+ 400	4826	+ 150
3031	− 30	5005	+ 900
3115	+ 600	5055	+ 450
3379	+ 780	5194	+ 270
3521	+ 730	5236	+ 500
3623	+ 800	5866	+ 650
3627	+ 650	7331	+ 500
4258	+ 500		

Fig. 1.1. Table taken from Slipher (1917), showing the radial velocity of 25 spiral nebulae (galaxies). Positive velocities stand for receding nebulae and negative velocities for approaching nebulae. All nebulae except four have a positive radial velocity and therefore move away from us.

This changed with Edwin Hubble, who used Cepheid variable stars as distance indicators. For these stars, Leavitt (1908); Leavitt & Pickering (1912) found a relation between the periodicity and the absolute luminosity. By measuring the period of the variability of a Cepheid, one gets its absolute luminosity, which in comparison to the observed flux, gives a measure of the distance to the star. This method assumes that the source is isotropic and therefore spreads light equally over a surface area of $4\pi D_{\mathrm{lum}}^2$, where D_{lum} is the distance from us to the source. Edwin Hubble used this approach for Cepheid variable stars he identified in the Andromeda Nebula and other close nebulae. He showed that these nebulae are outside our Milky Way, proving that there are other galaxies than the one we are located in (Hubble, 1925). Cepheid variables are intrinsically not bright enough to measure distances to faraway galaxies. Still, Hubble used them to calibrate other distance indicators like the brightest resolved stars in nearby galaxies and therefore was able to build a catalog containing distances to galaxies not reachable with Cepheid variables (Hubble, 1926). Lemaître (1927) then related redshifts and distances from galaxies and found that the Universe is expanding with roughly $600\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, meaning that the further away a galaxy is, the faster it moves away from us[2]. Although this was an astonishing discovery, Lemaître's work stayed mostly unnoticed at this time. In 1929 it was Edwin Hubble who changed the understanding of our Universe with his famous diagram plotting the radial velocity of galaxies (as listed in Fig. 1.1) versus their distance (Hubble, 1929). The original Hubble

---

[2] $1\,\mathrm{pc} \approx 3.0857 \cdot 10^{16}\,\mathrm{m} \approx 3.26156$ light years, where one light year is the distance light travels in one Julian year (365.25 days).

diagram is shown in Fig. 1.2, which suggests an expanding Universe with a Hubble constant $H_0 \approx 500\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ on the same order as Lemaître (1927), but almost ten times higher than current values of $\sim$70 $\mathrm{km\,s^{-1}\,Mpc^{-1}}$. The significant deviation is related to several unaccounted issues by Hubble. First, there is a scatter in the period-luminosity relation dependent on the color of the Cepheid variables. A second issue comes from the brightest resolved stars as distance indicators because bright, compact HII (singly ionized hydrogen) regions can be confused with stars. The third problem is that the absolute luminosity of the brightest star might depend on the galaxy type, and further, there might be dust extinction. These issues have been mentioned and corrected by Sandage (1958), providing an estimate of H_0 between 50 and $100\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, still rather uncertain but consistent with current measurements.

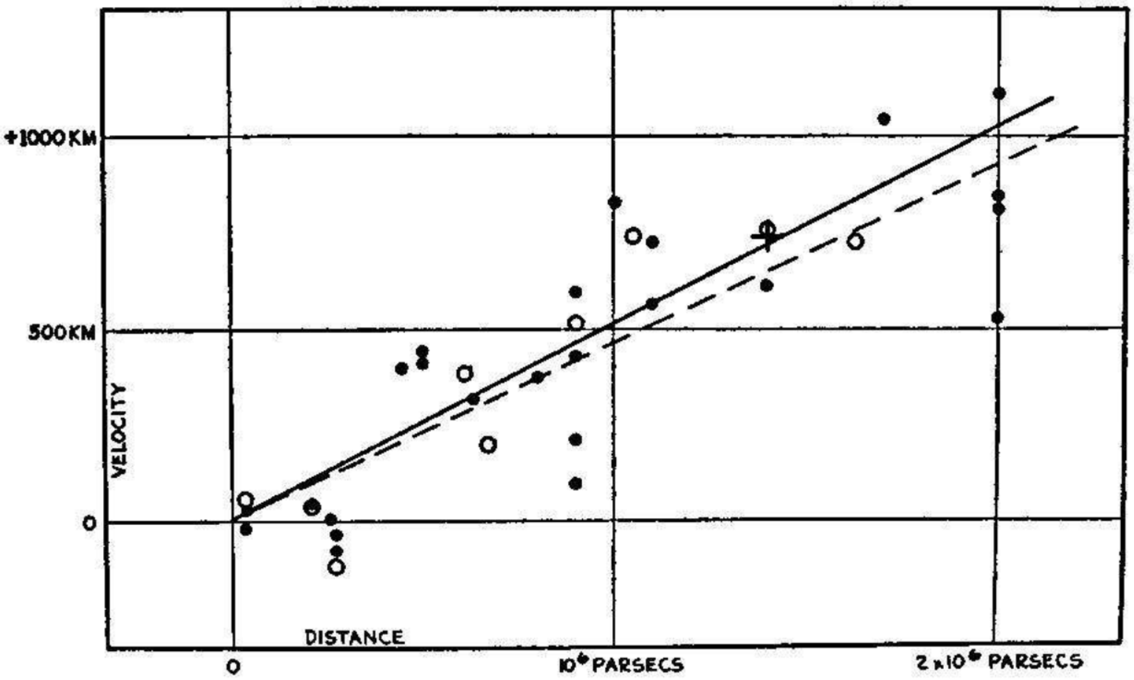


Fig. 1.2. Original Hubble diagram taken from Hubble (1929), showing the radial velocity (in km/s and not km, as displayed in the Figure) of galaxies as a function of their distance. The diagram suggests an expanding Universe with the Hubble constant $H_0 \approx 500\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$.

Before we come to recent measurements of the Hubble constant, it is interesting to note that Albert Einstein's General Theory of Relativity (Einstein, 1915, 1916), as we will introduce in Sect. 2.1.3, predicts a dynamic Universe. However, Einstein believed at that time that the Universe is static; therefore, he added the cosmological constant to his field equations to counterbalance the gravitational attraction of matter (Einstein, 1917). With the detection of the expanding Universe (Lemaître, 1927; Hubble, 1929), the idea of the cosmological constant was mostly dropped. However, with the discovery that the expansion of the Universe is not decelerated, as expected for a Universe dominated by matter, but instead accelerated (Riess et al., 1998; Perlmutter et al., 1999), the cosmological constant became relevant again. This constant can be interpreted as the simplest form of an unknown source of energy driving our Universe apart (dark energy; Turner, 1999). From there on, increasing amounts of observational evidence point toward our current standard model of cosmology, the so-called flat ΛCDM model, where our present Universe is composed of roughly 30% matter and 70% dark energy, consistent with the cosmological constant. The amount of radiation today is totally negligible in comparison to matter and dark energy. Further, matter accounts for ~80% dark matter and ~20% baryonic matter (Planck Collaboration, 2014, 2016, 2020). Dark matter is an invisible mass component, which was first introduced to explain the missing matter in the rotation curves of spiral galaxies (e.g., Faber & Gallagher, 1979; Rubin et al., 1980; Ashman, 1992; Persic et al., 1996).

1.1.2 Hubble tension

Since the first measurement of the Hubble constant by Lemaître (1927), it has been of great interest in observational cosmology to improve measurement methods to get more precise results. However, as precision improved, a tension between early and late-time H_0 measurements arose. The most precise results from an early-time probe come today from the Planck Collaboration (2020), achieving

$$H_0 = 67.4 \pm 0.5\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$$

from investigations of the cosmic microwave background (CMB). The CMB was emitted when the Universe was cold enough such that protons and electrons recombined and formed neutral hydrogen, allowing photons (that were previously scattering with ions) to escape. These photons, first accidentally discovered by Penzias & Wilson (1965), are measured today at a cosmological redshift of $z \approx 1100$, corresponding roughly to a time of 380000 years after the Big Bang[3]. Therefore, results on H_0 inferred from the CMB were extrapolated to our present time, corresponding to a cosmic redshift of zero, using the successful flat ΛCDM model. Hence, if the tension in comparison to late-time probes cannot be resolved due to unaccounted systematic errors in at least one of the measurements, this would point toward new physics beyond the flat ΛCDM model.

A well-established late-time probe for H_0 measures the relation between distance and redshift by calibrating a distance ladder. This method requires very bright standardizable candles reaching far into the Hubble flow, where the cosmological redshift dominates the doppler shift from peculiar motion. One such probe are SNe Ia, where the absolute luminosity can be related empirically to the width of the light curve (Phillips, 1993). Typically, multiple steps are required to calibrate SNe Ia, and there are many ways to realize that. The most precise results are achieved by the Supernova H_0 for the Equation of State (SH0ES) project (e.g., Riess et al., 2018, 2019, 2021). They use Gaia parallaxes in our Milky Way (Gaia Collaboration et al., 2021), detached eclipsing binaries in the Large Magellanic Cloud (Pietrzyński et al., 2019), and water masers in NGC 4258 (Reid et al., 2019) to calibrate Cepheid variables. These calibrated variables are then used to calibrate 42 SNe Ia (Riess et al., 2022), located in 37 hosts where Cepheid variables have been observed, reaching distances up to 50 Mpc ($z \lesssim 0.01$). Further, these calibrated SNe Ia enable the distance measurement to 277 SNe Ia from the Pantheon Plus sample (Scolnic et al., 2021; Brout et al.,

---

[3] Connecting redshift to cosmic time requires a cosmological model, and can be calculated via Eq. (2.30).

2021), located in the Hubble flow between $0.023 < z < 0.15$. The Hubble parameter is then determined from the relation between redshift and distance. For very low redshifts, this is the well-known Hubble law or Hubble–Lemaître law, as found by Lemaître (1927) and Hubble (1929), describing the linear relation between velocity and distance:

$$cz = H_0 D. \tag{1.1}$$

The term on the left-hand side is the velocity as plotted in Fig. 1.2 and comes from the approximation of the relativistic doppler effect for low redshift $v = cz$, as we will see in Sect. 2.1.2. In this context, the cosmic redshift is interpreted as a doppler redshift coming from a recession velocity due to the expanding space itself. However, from a full theoretical point of view, the relation between distance and redshift is not linear, and there are different definitions of cosmological distances, as we will see in Eqs. (2.34) and (2.35). Nonetheless, from such a distance-redshift relation, as we will learn in Sect. 2.1.4, and the SN Ia sample Riess et al. (2022) determine the Hubble constant

$$H_0 = 73.04 \pm 1.04\,\mathrm{km\,s^{-1}\,Mpc^{-1}},$$

which is in 4.9σ tension compared to the H_0 measurement from the Planck Collaboration (2020).

There are also other ways to calibrate a SN Ia sample in the Hubble flow, for example, using the Tip of the Red Giant Branch (TRGB). Red giants are former low-to-intermediate mass main sequence stars that evolved away from the main sequence. This evolution occurs when the hydrogen fuel in the core is exhausted and burned to helium, leading to a core contraction because the radiation pressure can not counterbalance the gravitational attraction anymore. Consequently, the core temperature increases, as well as the temperature and pressure in a hydrogen shell around the core, igniting hydrogen fusion there. Further, the photosphere expands to a certain size and cools during that process. As the shell burning proceeds, the star becomes more luminous, resulting in a luminous red star on the red giant branch in the Hertzsprung–Russell diagram. At the end phase of the red giant branch, low-mass red giants (like our sun will become one day) will reach a degenerate helium core. This means that the core becomes dense enough that the electron degeneracy pressure will prevent a further collapse. As soon as the core is hot enough, the helium burning starts almost simultaneously because of the degenerate core, resulting in a helium flash, corresponding to a maximum luminosity in the evolution of red giants, also known as the TRGB. This method serves as a distance indicator (e.g., Lee et al., 1993; Rizzi et al., 2007) to galaxies containing a large sample of red giants which cover the diversity of evolutionary tracks of these giants. Similar to the calibration of Cepheid variables, one can use, for example, eclipsing binaries to calibrate the TRGB.

Freedman (2021) used the TRGB to calibrate 15 hosts of 18 SNe Ia, which were further used to measure distances to 99 more distant SNe Ia up to a redshift of 0.08 from the

Carnegie Supernova Project, leading to

$$H_0 = 69.8 \pm 0.6\,(\mathrm{stat}) \pm 1.6\,(\mathrm{sys})\,\mathrm{km\,s^{-1}\,Mpc^{-1}}.$$

This late-time H_0 measurement agrees within 1.3σ with results from the Planck Collaboration (2020) and is also consistent on the 1.6σ level with Riess et al. (2022) and therefore, significantly lowers the tension between CMB measurements and the SH0ES project, suggesting unaccounted systematics in at least one of the measurements. However, other methods using the TRGB suggest higher H_0 values. One such approach was conducted by Soltis et al. (2021), where the trigonometric parallax of Omega Centauri[4] is measured to calibrate the TRGB for hosts of a SN Ia sample, yielding:

$$H_0 = 72.1 \pm 2.0\,\mathrm{km\,s^{-1}\,Mpc^{-1}}.$$

This result agrees with the SH0ES project, and Freedman (2021), but is inconsistent at the 2.3σ level with the CMB measurement (Planck Collaboration, 2020). A discussion of the presented results, also in the context of a few other H_0 measurement methods, is summarized by Valentino et al. (2021). However, the Hubble tension's origin is still unknown; therefore, new independent measurements are required to confirm or refute the tension. One such method is radiative transfer modeling of type II supernovae (SN II). This one-step method is based on luminosity estimates of SN II from the tailored-expanding-photosphere method (Dessart & Hillier, 2006; Dessart et al., 2008) by fitting synthetic spectra to observed ones. An extended version of TARDIS (Kerzendorf & Sim, 2014; Vogl et al., 2019) provides synthetic spectra for this application. Then, using an automated spectral fitting approach (Vogl et al., 2020), distances can be determined to a large sample of observed SNe II in the Hubble flow. With a proof-of-principle measurement, Vogl (2020) estimated

$$H_0 = 72.3^{+2.9}_{-2.8}\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$$

based on six SNe II, showing that this method will provide highly competitive constraints on H_0 with additional systems in the future. Other new approaches are based on gravitational wave sources (e.g., Abbott et al., 2017; Mukherjee et al., 2021; Gayathri et al., 2021), Megamasers (Pesce et al., 2020), and time-delay cosmography with strongly lensed supernovae (LSNe). This thesis focuses on LSNe, which will be introduced in the next section.

1.2 Gravitational lensing as a probe for the Hubble constant

According to general relativity, gravity is not a force, but instead, it is a manifestation of curved spacetime caused by massive objects. Light rays traveling through spacetime will follow the curvature, which corresponds effectively to a deflection. Therefore, if a faraway light source and a massive intervening object are close in projected separation (but separated by a large distance), one expects lens-like action, as first proposed by Einstein (1936).

---

[4] Omega Centauri is the biggest and most massive globular cluster in the Milky Way.

Einstein assumed, as a massive intervening object, referred to as gravitational lens or lens hereafter, a star, leading to lensing features on tiny scales far beyond what was observable at that time. However, Zwicky (1937) assumed galaxies as lenses leading to interesting applications in extragalactic astronomy, such as multiple images or arcs on the order of several arcseconds and, therefore, potentially observable. This phenomenon is known as strong gravitational lensing, but it took more than 40 years until the first time two images of a quasar[5] were observed (Walsh et al., 1979). The image separation of the system is $5.7''$, and it is located at a redshift of 1.405. The same redshift and almost identical spectra from both images confirmed that it is a gravitational lens with two images of the same quasar instead of two different quasars close along our line of sight.

Interesting to note is that the light from the different images travels along different path lengths and through different gravitational potentials, resulting in a time delay between the multiple images. This delay can be measured for variable sources by matching key features in the light curves from the two images in time. This time delay, in combination with a lens mass model and a reconstruction of mass perturbations along the line of sight, can be used to measure H_0. To obtain a lens model, one uses the observation of the lens with multiple images of the source. A typical approach is to predict the image positions from a lens model by optimizing the source position and parameters of the lens model until a good match between observed and predicted image positions is achieved. A well-established software for lens modeling is the Gravitational Lens Efficient Explorer (`GLEE`), developed by Suyu & Halkola (2010); Suyu et al. (2012). The line-of-sight effects can be accounted for via full ray tracing using multi-plane gravitational lensing, where every deflector is modeled as a single lens plane. However, this is a very complex process, and in practice, one typically uses approximations of the full ray-tracing (e.g., Greene et al., 2013; McCully et al., 2017; Birrer et al., 2017; Rusu et al., 2017) or weak gravitational lensing (e.g., Tihhonova et al., 2018).

1.2.1 Time-delay cosmography with lensed quasars

Time-delay cosmography for H_0 measurements was first proposed by Refsdal (1964) for supernovae (SNe) lensed into multiple images, but until now, it is mostly used for strongly lensed quasars (e.g., Suyu et al., 2010, 2013, 2014; Bonvin et al., 2018; Birrer et al., 2019; Sluse et al., 2019; Rusu et al., 2019; Chen et al., 2019; Wong et al., 2020; Birrer et al., 2020). The H_0 Lenses in COSMOGRAIL's Wellspring (H0LiCOW) program was initiated (Suyu et al., 2017) to get accurate and precise measurements of the Hubble constant from lensed quasars using time-delay measurements from The COSmological MOnitoring of GRAvItational Lenses (COSMOGRAIL; Courbin et al., 2018) collaboration. The H0LiCOW and COSMOGRAIL collaboration, together with The Strong lensing at High Angular Resolution Program (SHARP; Chen et al., 2019), determine in a blind analysis

---

[5] A quasar is a very luminous active galactic nucleus, powered by an accretion disc around a supermassive black hole.

$$H_0 = 73.3^{+1.7}_{-1.8}\,\mathrm{km\,s^{-1}\,Mpc^{-1}},$$

for six lensed quasar systems, assuming a flat ΛCDM cosmology and using physically motivated mass models (Wong et al., 2020). This measurement is in good agreement with the SH0ES project but in 3.2σ tension with the CMB measurement. The STRong-lensing Insights into the Dark Energy Survey (STRIDES; Shajib et al., 2020) collaboration has analyzed a seventh lensed quasar system, in good agreement with results from Wong et al. (2020). The main degeneracy in the mass models of the lens is the mass-sheet transformation (e.g., Falco et al., 1985; Schneider & Sluse, 2013; Kochanek, 2020), which leaves observables like the image position unchanged, but influences the H_0 measurement. The Time-Delay COSMOgraph (TDCOSMO; Millon et al., 2020) organization, consisting of H0LiCOW, COSMOGRAIL, SHARP, and STRIDES did an analysis without physically motivated mass models to allow for the mass-sheet transformation where the mass models are only constrained by stellar kinematics, leading to

$$H_0 = 74.5^{+5.6}_{-6.1}\,\mathrm{km\,s^{-1}\,Mpc^{-1}},$$

for the analysis of seven lensed quasars in a flat ΛCDM cosmology (Birrer et al., 2020). This result is in good agreement with Wong et al. (2020) and also consistent with Planck Collaboration (2020). Therefore, if no assumptions are made on the lens model, time-delay cosmography with the current sample of lensed quasars is, at the moment, not precise enough to break the Hubble tension, which is also confirmed by Denzel et al. (2021). Although it might be exaggerated to give up physically motivated mass models, LSNe will have several advantages over lensed quasar, helping to constrain the mass-sheet transformation.

1.2.2 Strongly lensed supernovae

As pointed out in Sect. 1.2.1, one of the main issues in the H_0 determination, using time-delay cosmography, are the degeneracies in the lens mass model. SNe, in comparison to quasars, have the advantage that they fade away, which helps to determine a more precise lens model by using the lens image with the multiple images of the SN host galaxy without contamination from the bright point source (Ding et al., 2021). Further, after the SNe has faded, one can trigger follow-up observations of the lensing galaxy to measure stellar kinematics (e.g., Barnabè et al., 2011a; Yıldırım et al., 2017; Shajib et al., 2018; Yıldırım et al., 2020) and determine the mass within a certain effective radius around the lens which helps to break model degeneracies. Additionally, suppose the source is a standardizable candle, like a SN Ia. In that case, we can also break the model degeneracies by determining the lensing magnification (Oguri & Kawano, 2003; Foxley-Marrable et al., 2018), that is, the additional brightness of the source due to the primary lens (galaxy or galaxy cluster). This approach works only in cases with small statistical magnifications from stars in the lens galaxy (microlensing). Besides, LSNe generally have a very characteristic light curve shape compared to the stochastically varying quasars, making the delay measurement easier and typically possible on shorter time scales. For these reasons LSNe and especially type Ia's are promising. We will investigate LSNe Ia in this thesis, where we focus on time-delay measurements as one of the three pillars of time-delay cosmography. A more detailed review article on LSNe was recently published (Liao et al., 2022), and another one will follow soon (Suyu et al., in prep.). Furthermore, in Oguri (2019), one can find a review article on lensed transients, including lensed SNe.

1.2.2.1 Detected lensed supernovae

So far, primarily lensed quasars have been used for time-delay cosmography in contrast to LSNe due to their rarer appearance. For example, Oguri & Marshall (2010) estimates that there will be roughly 80 times more strongly lensed quasars detected in future surveys in comparison to LSNe Ia. However, a few systems have been found in the last decade, which we will briefly summarize in the following.

PS1-10afx The PS1-10afx was reported by Chornock et al. (2013) as a new type of superluminous supernova. However, Quimby et al. (2013) showed that the spectrum and light curves in different bands are consistent with a normal SNe Ia located at $z = 1.39$, although the flux is 30 times too bright, which could be easily explained if the system is lensed. This is because the lensing magnification factor μ increases the observed flux linearly in comparison to the case of no lensing[6]. After the SNe faded, a spectrum showed the presence of a foreground lens at $z = 1.12$ (Quimby et al., 2014). Even though the system was likely a LSNe Ia, the multiple images have never been resolved, and the opportunity for further investigation is long gone.

SN Refsdal Multiple images of a LSNe were observed for the first time in a Hubble Space Telescope (HST) observation on November 10th, 2014. The LSNe found by Kelly et al. (2015, 2016a,b) was named SN "Refsdal", in honor of Sjur Refsdal and corresponded to a core-collapse SN at $z = 1.489$ behind the MACS J1149.6+2223 galaxy cluster at $z = 0.542$. This system is shown in Fig. 1.3. At detection, the four images S1 to S4 were visible, as highlighted in the inserted panel (b). These four images come from a double lensing event, where the SN source was first lensed by the MACS J1149 cluster into three images, and then one of these three images was further lensed into images S1 to S4 by a cluster member. Given that the time delay for a symmetric galaxy lens is much shorter (order of days) than the appearance of the SN (order of hundred days in this case), all four images are visible. The time delays between the three images produced only by the MACS J1149 cluster are much longer, and as predicted from the lens model, one of the images faded already as it occurred before S1 to S4. However, the third image SX, where the location is marked in the inserted panel (a), appeared afterward. When SN Refsdal was detected, the prediction of the reappearance was quite uncertain and ranged from 2015 to 2025 (Kelly et al.,

---

[6] Most of the time the magnification factor μ is greater than one, but there are also configurations where μ is smaller than one, and therefore the system is demagnified.

2015). More accurate cluster models were possible after Grillo et al. (2016) investigated 117 potential cluster members, where 68 members were confirmed spectroscopically with redshift values ranging from 0.5272 to 0.5660. This data was shared with several modeling teams before the SN reappearance. Grillo et al. (2016); Kawamata et al. (2016) predicted the time of appearance and magnification of image SX in excellent agreement with the later observation of SX. The image SX of SN Refsdal reappeared in an HST observation taken on December 11th in 2015 (Kelly et al., 2016a), ending an astonishing and successful blind test for gravitational lensing! As soon as time delays for image SX are available, Grillo et al. (2018, 2020) forecasted and showed that with SN Refsdal it is possible to determine H_0 with 6% uncertainty, including statistical and systematic errors.

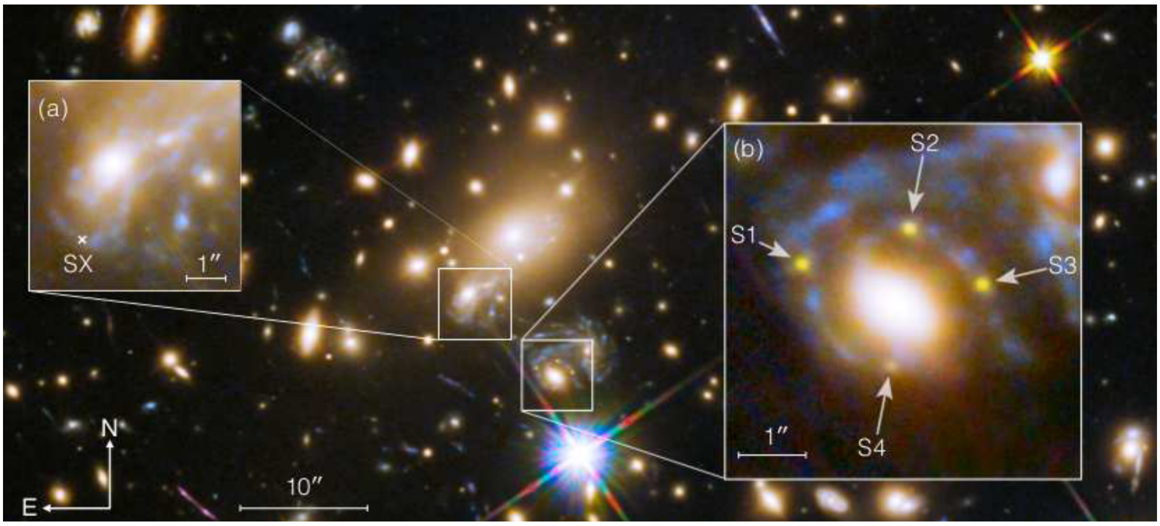


Fig. 1.3. SN Refsdal in the MACS J1149.6+2223 galaxy cluster. At detection, only the four images S1 to S4 in the inserted panel (b) were visible. These four images have been lensed by the MACS J1149 cluster and by a cluster member (bright galaxy surrounded by S1 to S4). Two other images were expected from the lensing of the MACS J1149 cluster only, where one of them occurred before S1 to S4, and one (labeled as SX) was observed afterward at the position marked in the inserted panel (a). Image taken from Grillo et al. (2018). Original image credit: NASA, ESA/Hubble.

iPTF16geu The first LSNe Ia, where multiple images have been resolved, was the iPTF16geu (Goobar et al., 2017) detected by the intermediate Palomar Transient Factory (iPTF). This system corresponds to a SNe Ia at $z = 0.409$ strongly lensed by an intervening galaxy at $z = 0.216$. The iPTF could not resolve the multiple images, but spectroscopy confirmed that the system is a normal SN Ia at a redshift ~0.4, although the flux was more than 50 times brighter than expected for that redshift. The history is very similar to PS1-10afx, but this time high resolution follow-up observations were triggered, and four multiple images in a symmetric configuration around the lensing galaxy were resolved, as shown in Fig. 1.4. For such a symmetric system, time delays are very short, and one would expect similar magnifications. A lens model by More et al. (2017) showed that only one image matched the expected magnification, two of the images were significantly brighter, and another one was substantially fainter. These deviations might be explained by dust extinction, substructure lensing, or microlensing (More et al., 2017; Mörtsell et al., 2020). Especially microlensing will be of interest in this thesis. This additional lensing effect from stars in the lensing galaxy can highly distort and magnify the light curves of the LSN Ia differently. More about microlensing will follow in Secs. 2.2.5 and 2.3. For time-delay cosmography, iPTF16geu is not suitable since all delays are lower than 1.5 days (Dhawan et al., 2020) and therefore far too uncertain.

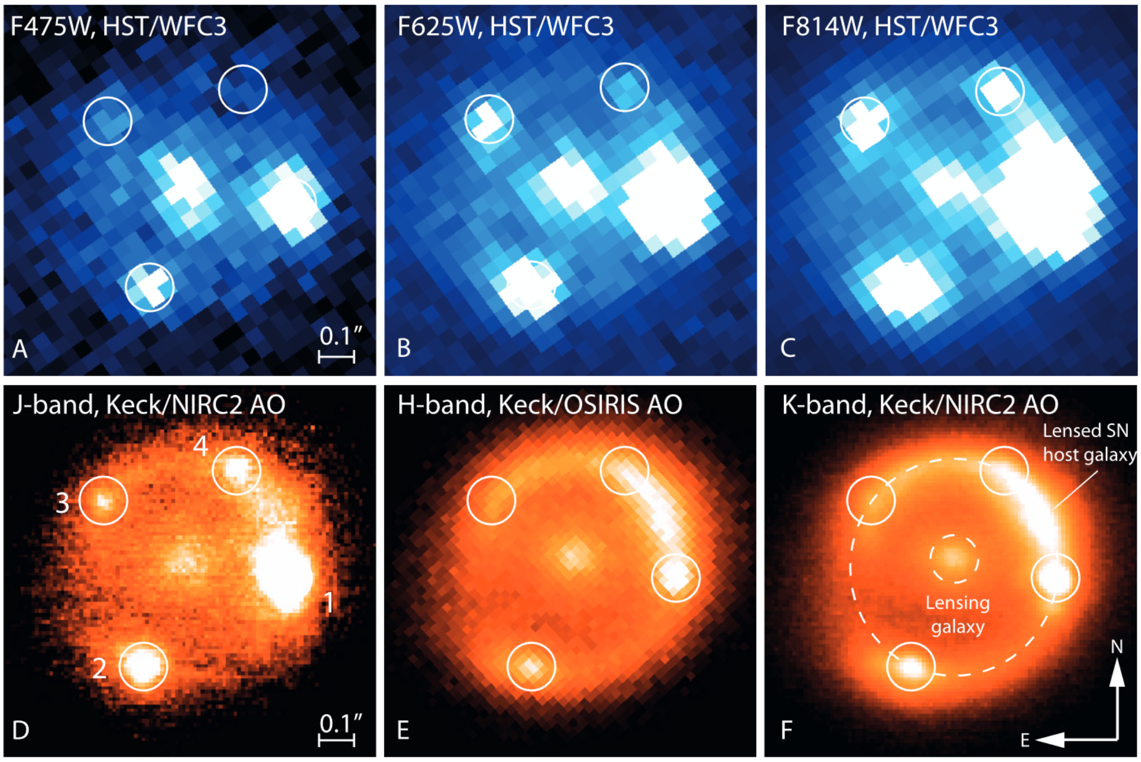


Fig. 1.4. The iPTF16geu, the first observed LSNe Ia where multiple images have been resolved. Panels A, B, and C show observations of the Hubble Space Telescope in three optical bands. Panels D, E, and F show three near-infrared bands from the Keck Observatory using Adaptive Optics. Panel B, C, and D clearly reveal four point sources. In panel A, two of the images are too faint, and panels E and F also show significant contributions of the host galaxy of iPTF16geu. Image taken from Goobar et al. (2017).

SN Requiem This LSNe has been detected by Rodney et al. (2021) in archival HST imaging of 2016, where three images of the SNe were visible in giant arcs of its host galaxy at a redshift of 1.95, strongly lensed by the galaxy cluster MACSJ0138.0-2155. A second image from 2019 shows that all three SNe images have faded away. Rodney et al. (2021) predict that a fourth and fifth image will appear. The fifth image will be highly demagnified. However, the fourth image will be of interest for future observations, but the appearance of the image will last until the year 2037 ± 2.

SN 2022riv Another LSNe was reported in mid-August 2022 by Kelly et al. (2022) in the MACS2129 galaxy-cluster field in a multiply imaged galaxy located behind the cluster at a redshift of $z = 1.52$. Lens models show that the detected SN image is the trailing or last-arriving image. The detected LSN could be used for time-delay cosmography if archival data of the previous image exists or by triggering very deep space-based follow-up observations, which might reveal the past image.

SN Zwicky / SN2022qmx A week after the SN 2022riv announcement, Goobar et al. (2022) reported a new LSNe Ia. The system was detected by the Zwicky Transient Facility (ZTF) in a g band image from 1st August 2022 and saved as a SN candidate two days later. The spectrum, obtained on 21st August, shows a perfect match to a normal SN Ia at a redshift of 0.35 but roughly three magnitudes brighter than an unlensed SNe Ia at that redshift. Follow-up observations are ongoing, but the low image separation suggests short time delays, which are not ideal for time-delay cosmography.

1.2.2.2 Rates and searches for lensed supernovae

So far, only six lensed SNe have been detected, but this will change with future surveys, such as the upcoming Rubin Observatory Legacy Survey of Space and Time (LSST; Ivezić et al., 2019). The telescope is currently under construction on the Cerro Pachón in Chile, and the 10-year monitoring survey is planned to start in 2024. The LSST survey[7] targets roughly $20000\,\mathrm{deg}^2$ of the southern hemisphere with a field of view of $9.6\,\mathrm{deg}^2$, a resolution of $0.2''$ per pixel, and six photometric bands (u, g, r, i, z, and y) covering wavelengths between ~3200 Å and ~10800 Å. Due to its rapid cadence and great depth, one expects to find up to ~900 LSNe Ia and ~2500 lensed core-collapse SNe (LCCSNe) (Quimby et al., 2014; Goldstein & Nugent, 2017; Goldstein et al., 2018; Wojtak et al., 2019) over the ten-year survey, assuming only galaxy-scale lenses where the lens modeling is much simpler in comparison to galaxy clusters. If one considers only bright and spatially resolved LSNe within LSST[8], Oguri & Marshall (2010) predict ~50 LSNe Ia and ~80 LCCSNe over ten years. Huber et al. (2019) show that depending on the LSST observing strategy, the number of LSNe Ia predicted by Oguri & Marshall (2010) varies between 40 and 100, from which 10 to 25 LSNe Ia yield accurate and precise time delays, but only if follow-up observations every second day in three filters to an LSST-like depth are triggered. If only LSST data is used without follow-up observations, only 1 to 8 LSNe provide accurate and precise time delays. The LSST observing strategy is important for the number of LSNe Ia and is currently optimized (Lochner et al., 2018; Huber et al., 2019; Lochner et al., 2022). Still, given that many different science cases depend on the LSST observing strategy, there is only limited room for improvement. Therefore, it is essential to optimize the follow-up strategy further and improve the time-delay measurement techniques. The beforementioned is of particular interest in this thesis as part of the Highly Optimised Lensing Investigations of Supernovae, Microlensing Objects, and Kinematics of Ellipticals and Spirals (HOLISMOKES) program[9]. This was initiated by

---

[7] `https://www.lsst.org/scientists/keynumbers`

[8] The OM10 catalog contains only LSNe, where the peak of the fainter image for a double or the third brightest image for a quad is brighter than 22.6 in the i band. Further, it requires a minimum image separation of $0.5''$ to assure that systems are resolved and a maximum image separation of $4''$ to exclude cluster scale lenses.

[9] `https://shsuyu.github.io/HOLISMOKES/site/`

Suyu et al. (2020) to measure the current expansion rate of our Universe with LSNe and to study early phases in the evolution of SNe.

In the following, we explain current search methods for LSNe. The first one is simply via image multiplicity, which requires deep surveys with a good resolution like the LSST, which can resolve the multiple images of a sufficient number of LSNe. For LSST, Wojtak et al. (2019) predict ~450 LSNe Ia and ~1050 LCCSNe, which can be found just from image multiplicity.

In the case of PS1-10afx, iPTF16geu, and SN Zwicky, the systems were not resolved at detection but were identified as LSNe via lensing magnification. Further, spectroscopic follow-up observations are critical to rapidly clarify the origin of a transient identified as a SN candidate. If the candidate is confirmed to be a normal SNe of a particular type, one can compare the observed brightness of the event to the expected brightness for the given redshift. If the SNe is significantly brighter than the variation expected for a specific SN type, the chances are high that the system is a LSNe. By triggering high-resolution follow-up observations, the multiple images can be resolved. Using just the lensing magnification approach Wojtak et al. (2019) predicts that LSST will find ~600 LSNe Ia and ~2200 LCCSNe. However, the lensing magnification approach in LSST will detect mainly LSNe with low image separation and, therefore, shorter time delays, which is not ideal for time-delay cosmography.

Another approach is to monitor star-forming galaxies, strongly lensed into multiple images, to wait for a SN occurrence. However, the number of potential targets is limited, although the list of lenses and lens candidates is growing quickly. By today, we already have several hundred lenses and several thousand encouraging lens candidates (e.g., Browne et al., 2003; Bolton et al., 2004; Sonnenfeld et al., 2018, 2020; Cañameras et al., 2020; Canameras et al., 2021; Shu et al., 2022). Such a large catalog of strong lenses is extremely useful for matching reported transient alerts in a survey with known strong lenses to identify potential LSNe. Shu et al. (2018) showed that targeting 128 known star-forming galaxies strongly lensed by an intervening galaxy, provides 1.23 ± 0.12 LSNe Ia and 10.4 ± 1.1 LCCSNe per year. A monitoring program with a cadence of five days with a single-visit 5σ point source depth of 24.7 mag in the r band should be able to detect per year 0.49 LSNe Ia and 2.1 LCCSNe. The advantage is that such a sample will provide large image separations (median of roughly $2''$) ideal for time-delay cosmography. At the moment, the issue is that we cannot afford to follow up a sufficient number with that cadence and depth assumed by Shu et al. (2018), but the situation will improve once LSST starts and monitors all these candidates.

1.3 Thesis outline

This thesis contains six chapters and aims to develop tools to measure time delays in LSNe Ia. This delay, in combination with lens mass modeling and reconstruction of mass perturbations along the line of sight, provides a direct way to measure the Hubble constant H_0 and, therefore, can help to address the current tension in H_0.

In Chapter 2, we describe the basics and methods this thesis is built on. Given that a strengthened H_0 tension will point towards new physics, we will explain the theoretical background of gravity and cosmology, including the exact definition of the Hubble parameter in Sect. 2.1. This is followed by a detailed introduction to gravitational lensing in 2.2. In Sect. 2.3, we explain the simulation of realistic mock LSNe Ia, used in this thesis, followed by the introduction to machine learning in Sect. 2.4, which we will use to build machine learning models to measure time delays in LSNe Ia.

In Chapter 3, we will investigate microlensing on LSNe Ia, especially using color curves as a potential source to mitigate uncertainties in time-delay measurements. This chapter contains my first-author publication Huber et al. (2021), with a brief summary. This is the third paper of the HOLISMOKES program.

In Chapter 4, we introduce two simple machine learning models, namely a fully connected neural network and a random forest, to measure time delays in LSNe Ia. Further, we investigate different sources of uncertainties and provide guidelines for future follow-up observations. This chapter contains my first-author publication Huber et al. (2022), with a brief summary. This is the seventh paper of the HOLISMOKES program.

In Chapter 5, we describe the more advanced long short-term memory network in combination with a fully connected neural network. This new approach is more flexible and has broader applications than the fully connected neural network or the random forest but still provides significantly improved performance.

Finally, we will conclude and give an outlook in Chapter 6.

2 Basics and methods

2.1 Theory of gravity and cosmology

2.1.1 Newtonian gravity

The fundamental description of our Universe is based on general relativity (GR), which emerged from the inconsistency between Newtonian gravity and special relativity (SR). In Newtonian gravity, the equation of motion in a three-dimensional space is described by

$$\vec{a} = -\nabla\Phi, \tag{2.1}$$

where $\vec{a}$ is the acceleration and Φ the gravitational potential, which is related to the mass density of a gravitational field ρ via the Poisson equation

$$\nabla^2\Phi = 4\pi G\rho, \tag{2.2}$$

where ∇ is the nabla operator, ∇^2 is the Laplace operator, and G is the Gravitational constant. Therefore, Eq. (2.2) describes the gravitational field, and Eq. (2.1) defines how matter responds to this field, and these two equations define Newtonian gravity.

To write this in a more intuitive way, we consider a case outside of a given mass distribution, where the mass density $\rho = 0$, and therefore, the Poisson equation reduces to the Laplace equation

$$\nabla^2\Phi = 0. \tag{2.3}$$

If we assume a point source with mass M_1, the Gravitational potential at a distance r is described by

$$\Phi = -\frac{GM_1}{r} \tag{2.4}$$

as the fundamental solution of Eq. (2.3) and therefore, the acceleration via Eq. (2.1) is:

$$\vec{a} = -\nabla\frac{GM_1}{r} = \frac{GM_1}{r^2}\hat{r}. \tag{2.5}$$

If we assume a second point mass particle with mass M_2 in the gravitational field of M_1 the force f_{12} acting on point mass M_2 is

$$f_{12} = M_2 a = G\frac{M_1 M_2}{r^2}, \tag{2.6}$$

which is the same force acting on M_1 coming from the gravitational field of M_2. We can see from Eq. (2.2) or Eq. (2.6) that there is no time dependency. This is in conflict with SR which states that no information can travel faster than the speed of light in vacuum $c = 2.99792458 \cdot 10^8\,\mathrm{m\,s^{-1}}$. For example, if we consider Eq. (2.6), and assume the hypothetical case that the point mass M_1 would disappear, then the force f_{12} acting on M_2 would be immediately zero, no matter how far M_2 and M_1 are apart.

2.1.2 Special relativity

Maxwell's equations (Maxwell, 1873; Heaviside, 1922) describe the classical electromagnetism and are valid in every inertial frame and, therefore, consistent with SR (Einstein, 1905), in contrast to Newtonian mechanics. An inertial frame is a reference frame that does not undergo acceleration, meaning that a particle in that frame moves with constant velocity if no force is acting on it. The consistency of classical electromagnetism with SR means that these equations are Lorentz invariant. Therefore, the equations have the same form under a Lorentz transformation, which is the change from one inertial frame to another. The Lorentz transformation is defined such that the line element $\mathrm{d}s$ does not change:

$$\mathrm{d}s^2 = c^2\mathrm{d}t^2 - \mathrm{d}x^2 - \mathrm{d}y^2 - \mathrm{d}z^2 = c^2\mathrm{d}t'^2 - \mathrm{d}x'^2 - \mathrm{d}y'^2 - \mathrm{d}z'^2, \tag{2.7}$$

where t is the time and x, y, and z are the spatial coordinates in the inertial frame IF and (t', x', y', z') are the coordinates of inertial frame IF′. Instead of space and time as independent quantities, we refer to the four-dimensional set of coordinates (t, x, y, z) as spacetime in SR. In addition to the assumption that the fundamental laws of physics take the same form in every inertial frame, SR postulates that the speed of light is constant and, therefore, the same in all inertial frames. This is fundamentally different from Newtonian mechanics, where velocities are simply added. The constant c motivates the definition of the line element as described in Eq. (2.7), given that $\mathrm{d}s = 0$ describes a lightlike event that moves at a velocity of

$$v = \sqrt{\frac{\mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2}{\mathrm{d}t^2}} = c = \sqrt{\frac{\mathrm{d}x'^2 + \mathrm{d}y'^2 + \mathrm{d}z'^2}{\mathrm{d}t'^2}} = v' \tag{2.8}$$

in both inertial frames IF and IF′. Further, we can define from Eq. (2.7) the proper time

$$\mathrm{d}\tau^2 = \frac{\mathrm{d}s^2}{c^2}, \tag{2.9}$$

which is another Lorentz invariant quantity. The proper time between two events in spacetime is the time measured by an observer who moves on a straight line between these two events. The reason is that we can find an inertial frame where this observer and its clock are at rest, and therefore, we have $\mathrm{d}x = \mathrm{d}y = \mathrm{d}z = 0$ and $\mathrm{d}\tau^2 = \mathrm{d}s^2/c^2 = \mathrm{d}t^2$.

In the same way, we can also define the proper length (also Lorentz invariant) of an object, which is the length of the object measured in the inertial frame where it is at rest:

$$\mathrm{d}\sigma^2 = -\mathrm{d}s^2. \tag{2.10}$$

SR has many interesting consequences, and probably the most fascinating one is time dilation, which, in theory, allows us to travel to the future. This can be illustrated using Eq. (2.7) for two clocks at rest in their corresponding inertial frame IF and IF′, where IF′ moves along the z-axis measured from the IF with speed $v = \mathrm{d}z/\mathrm{d}t$. The line element of the clock under motion in the IF can be written as $\mathrm{d}s^2 = c^2\mathrm{d}t^2 - \mathrm{d}z^2$ given that it only moves in the z-direction and therefore $\mathrm{d}x = \mathrm{d}y = 0$. In the IF′ the same clock is at rest. Therefore, we have

$\mathrm{d}s^2 = c^2\mathrm{d}t'^2$, leading to $c^2\mathrm{d}t^2 - \mathrm{d}z^2 = c^2\mathrm{d}t'^2$, which can be rewritten as $\mathrm{d}t = \mathrm{d}t'/\sqrt{1-v^2/c^2}$. The consequence is that $\mathrm{d}t > \mathrm{d}t'$, which means that the time interval measured from IF is longer than from the clock which moves in IF and therefore “moving clocks run slower”. The same result can be achieved if we are in IF and calculate the proper time of the moving clock $\mathrm{d}\tau^2 = \mathrm{d}t^2 - \mathrm{d}z^2/c^2 = \mathrm{d}t^2(1-v^2/c^2)$, where we see that the proper time $\mathrm{d}\tau$ corresponds to the coordinate time $\mathrm{d}t'$ where the clock is at rest:

$$\mathrm{d}t = \gamma\,\mathrm{d}\tau, \tag{2.11}$$

where γ is the Lorentz factor:

$$\gamma = \frac{1}{\sqrt{1-\frac{v^2}{c^2}}} \geq 1. \tag{2.12}$$

Confusing seems that seen from IF′, IF moves away with speed v, and therefore, the clock in the IF runs “slower”. This is not a contradiction as there is only one moment in time when both can compare their clocks directly, and that is when they pass by each other (see also the twin paradox). Therefore, as the theory states, everything is relative.

A direct consequence of the time dilation is the relativistic doppler effect, where a source that emits radiation, with wavelength λ_e, moves directly towards (negative v) or away (positive v) from an observer, who sees the wavelength λ_o:

$$\lambda_\mathrm{o} = \lambda_\mathrm{e}\sqrt{\frac{1+v/c}{1-v/c}}. \tag{2.13}$$

The observed wavelength is redshifted ($\lambda_\mathrm{o} > \lambda_\mathrm{e}$) if the source moves away from the observer and blue shifted ($\lambda_\mathrm{o} < \lambda_\mathrm{e}$) if it moves towards the observer. Given the definition of the redshift

$$1+z := \frac{\lambda_\mathrm{o}}{\lambda_\mathrm{e}}, \tag{2.14}$$

the redshift from the relativistic doppler effect is

$$z = \sqrt{\frac{1+v/c}{1-v/c}} - 1 \approx \frac{v}{c}. \tag{2.15}$$

The approximated solution comes from the Taylor expansion for low velocities corresponding to the non-relativistic doppler shift.

There are also other exciting consequences of SR; namely, moving rulers are shortened, and massive particles will always move slower than the speed of light, where theoretical predictions from SR are in excellent agreement with experiments (Mattingly, 2005).

In SR, we summarize the time and spatial coordinates of an event in a four-vector

$$x^\mu = (x^0, x^1, x^2, x^3) = (ct, x, y, z), \tag{2.16}$$

where the superscript is an index instead of the power. Further, we work in the Minkowski space, which is a four-dimensional flat spacetime defined by the Minkowski metric

$$\eta_{\mu\nu} = \mathrm{diag}(1, -1, -1, -1), \tag{2.17}$$

which is a 4 × 4 matrix where off-diagonal elements are zero. With that definition, one can write the line element from Eq. (2.7) as

$$\mathrm{d}s^2 = \sum_{\nu=0}^{3} \sum_{\mu=0}^{3} \eta_{\mu\nu} \mathrm{d}x^{\mu} \mathrm{d}x^{\nu}. \tag{2.18}$$

These notations will be especially relevant for the next section when we briefly introduce the fundamental equations of GR. To learn more about SR, see, e.g., chapter 1 by Carroll (2013).

2.1.3 General relativity

As discussed in Sect. 2.1.1 Newtonian gravity is not compatible with SR, and therefore Albert Einstein developed GR (Einstein, 1915, 1916). In GR, gravitation is not anymore a force, as in Eq. (2.6), but instead, it is a manifestation of curved spacetime caused by energy and momentum. This is motivated by the Einstein Equivalence Principle (EEP): "In small enough regions of spacetime, the laws of physics reduce to those of special relativity; it is impossible to detect the existence of a gravitational field by means of local experiment" (Carroll, 2013, p. 50), which states that no observer in a tiny box can distinguish, if the box undergoes a uniform acceleration (which can be described in SR), or if the box is in an external gravitational field. Therefore, the flat spacetime metric $\eta_{\mu\nu}$ from Eq. (2.17) needs to be replaced by a more general metric to describe curved spacetime

$$\eta_{\mu\nu} \rightarrow g_{\mu\nu}, \tag{2.19}$$

where $g_{\mu\nu}$ is the metric tensor, which replaces the gravitational field Φ as used in Eqs. (2.1) and (2.2). The line element as introduced in Eq. (2.18) is replaced by

$$\mathrm{d}s^2 = \sum_{\nu=0}^{3} \sum_{\mu=0}^{3} g_{\mu\nu} \mathrm{d}x^{\mu} \mathrm{d}x^{\nu}. \tag{2.20}$$

As a next step, it is necessary to find new formulations of Eqs. (2.1) and (2.2), to define the gravitational field and how matter responds to it. In GR, this means how matter (energy and momentum) curves spacetime and how the curvature influences matter. The procedure is that one takes a law of physics valid in SR (IF with flat spacetime) and writes it in a general coordinate-invariant form, namely a tensor equation. The first equation is the geodesic equation (Carroll, 2013, p. 152)

$$\frac{\mathrm{d}^2 x^{\mu}}{\mathrm{d}u^2} = -\sum_{\rho=0}^{3} \sum_{\sigma=0}^{3} \Gamma^{\mu}_{\rho\sigma} \frac{\mathrm{d}x^{\rho}}{\mathrm{d}u} \frac{\mathrm{d}x^{\sigma}}{\mathrm{d}u} \tag{2.21}$$

describing the curve $x^\mu(u)$ of a free particle in GR. This means that the geodesic equation describes the path, parametrized by the affine parameter u, of a particle where no force, like the electromagnetic one, is acting on it. However, the particle is still under the influence of gravity (no force in GR) as incorporated in the Christoffel symbol $\Gamma^\mu_{\rho\sigma}$, which is composed of partial derivatives of the metric tensor $g_{\mu\nu}$ (Carroll, 2013, p. 93). Eq. (2.21) therefore replaces Eq. (2.1) in Newtonian mechanics. As the second equation in GR, we have the Einstein field equations (EFEs):

$$R_{\mu\nu} - \frac{R_\mathrm{s}}{2} g_{\mu\nu} + \Lambda g_{\mu\nu} = -\frac{8\pi G}{c^4} T_{\mu\nu}, \tag{2.22}$$

where $R_\mathrm{s} = \sum_{\nu=0}^{3}\sum_{\mu=0}^{3} g^{\mu\nu} R_{\mu\nu}$ is the Ricci scalar and $R_{\mu\nu}$ is the Ricci tensor (Carroll, 2013, pp. 122, 129-130), containing combinations of Christoffel symbols and partial derivatives of Christoffel symbols. Therefore, the Ricci tensor contains second derivatives (and combinations of first-order derivatives) of the metric tensor $g_{\mu\nu}$ and quantifies the curvatures of spacetime. Further, Λ is the cosmological constant, and the energy-momentum tenor is $T_{\mu\nu}$, and therefore, the EFEs describe how the metric tensor $g_{\mu\nu}$ is influenced by energy and momentum. More details about GR can be found in, e.g., chapter 2-4 by Carroll (2013) or chapter 7 and 8 by Hobson et al. (2006).

2.1.4 Cosmology

In the following, we describe some important quantities to describe our Universe. A more detailed description can be found in, e.g., chapter 14 and 15 by Hobson et al. (2006) and chapter 8 by Carroll (2013).

2.1.4.1 Friedmann-Lemaître-Robertson-Walker metric

To construct a metric tensor $g_{\mu\nu}$ for the whole Universe, we assume that the Universe is on large scales in space homogeneous and isotropic (also known as the cosmological principle). This means that at any particular time, the Universe looks the same from all positions in space, and all directions in space at any point are equivalent[10]. This assumption is motivated by the distribution of galaxies on large scales and by observations of the CMB, where the temperature is almost the same in any direction with only tiny variations. Furthermore, we use comoving coordinates (x^1, x^2, x^3) which are constant in cosmic time t ($x^0 = ct$) and therefore

$$\frac{\mathrm{d}x^\mu}{\mathrm{d}t} = (c, 0, 0, 0). \tag{2.23}$$

These are the coordinates of a fundamental observer with no motion relative to the overall cosmological fluid. From these assumptions one can derive the Friedmann-Lemaître-Robertson-Walker (FLRW) metric (e.g., Hobson et al., 2006, chapter 14):

$$\mathrm{d}s^2 = c^2\mathrm{d}t^2 - R^2(t)\Big(\mathrm{d}\chi^2 + S^2(\chi)(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2)\Big), \tag{2.24}$$

---

[10] Homogeneity is the invariance of a metric under translation and isotropy is the invariance of a metric under rotation.

where (χ, θ, ϕ) are comoving coordinates with the limits $0 \leq \chi < \infty$, $0 \leq \theta \leq \pi$, and $0 \leq \phi \leq 2\pi$, t is the cosmic time ("time since Big Bang"), and $R(t)$ is the scale factor (not to be confused with the Ricci scalar R_s). Further,

$$S(\chi) = \begin{cases} \sin(\chi) & \text{if } k = 1;\ \text{positive spatial curvature (closed Universe)} \\ \chi & \text{if } k = 0;\ \text{zero spatial curvature (flat Universe)} \\ \sinh(\chi) & \text{if } k = -1;\ \text{negative spatial curvature (open Universe)} \end{cases}, \tag{2.25}$$

where the constant k determines the geometry of the Universe. For two fundamental observers, which can be loosely imagined as galaxies neglecting peculiar motion, the coordinate distance χ between them stays constant with cosmic time, although the physical distance changes following the scale factor $R(t)$. At this stage, we can define the Hubble parameter, which is of particular interest in this thesis and expresses the change of the scale factor with cosmic time ($\dot{R} = \mathrm{d}R/\mathrm{d}t$) in comparison to the scale factor:

$$H(t) := \frac{\dot{R}}{R}. \tag{2.26}$$

2.1.4.2 Redshift, lookback time, age of the Universe

With the geodesic Eq. (2.21), one can calculate the path of free particles in the comoving coordinates as defined in Eq. (2.24). One finds that θ and ϕ are constant (Hobson et al., 2006, pp. 365-367), meaning that light and matter move radially in the Universe if no other force or gravitational field except the cosmological "background" is present. Using $\mathrm{d}\theta = \mathrm{d}\phi = 0$ in Eq. (2.24) and assuming light ($\mathrm{d}s^2 = 0$), we get

$$\mathrm{c}^2 \mathrm{d}t^2 = R^2(t) \mathrm{d}\chi^2 \tag{2.27}$$

and therefore, if an emitter sends the light at time t_e and it reaches an observer at time t_o, we get

$$\int_{t_\mathrm{e}}^{t_\mathrm{o}} \frac{c\mathrm{d}t}{R(t)} = \int_0^{\chi_\mathrm{e}} \mathrm{d}\chi.$$

We can do the same calculation for the light one period later and get

$$\int_0^{\chi_\mathrm{e}} \mathrm{d}\chi = \int_{t_\mathrm{e}+\lambda_\mathrm{e}/c}^{t_\mathrm{o}+\lambda_\mathrm{o}/c} \frac{c\mathrm{d}t}{R(t)} = \int_{t_\mathrm{e}}^{t_\mathrm{o}} \frac{c\mathrm{d}t}{R(t)} + \int_{t_\mathrm{o}}^{t_\mathrm{o}+\lambda_\mathrm{o}/c} \frac{c\mathrm{d}t}{R(t)} - \int_{t_\mathrm{e}}^{t_\mathrm{e}+\lambda_\mathrm{e}/c} \frac{c\mathrm{d}t}{R(t)}$$

and therefore

$$\int_{t_\mathrm{e}}^{t_\mathrm{e}+\lambda_\mathrm{e}/c} \frac{c\mathrm{d}t}{R(t)} = \int_{t_\mathrm{o}}^{t_\mathrm{o}+\lambda_\mathrm{o}/c} \frac{c\mathrm{d}t}{R(t)}.$$

Given that one period for visible light is on the order of 10^{-15} s, we can assume that $R(t)$ is constant and solve the integral as $\frac{\lambda_\mathrm{e}}{R(t_\mathrm{e})} = \frac{\lambda_\mathrm{o}}{R(t_\mathrm{o})}$, yielding the relation between cosmic redshift

z and the scale factor R:

$$1+z=\frac{\lambda_{\mathrm{o}}}{\lambda_{\mathrm{e}}}=\frac{R(t_{\mathrm{o}})}{R(t_{\mathrm{e}})}=\frac{R_0}{R(t)}. \tag{2.28}$$

Here we assume that t is the age of the Universe when the photon was emitted, which we observe at our present time corresponding to the age of the Universe t_0 and therefore $R_0=R(t_0)=R(t_{\mathrm{o}})$. We can calculate the lookback time $t_{\mathrm{lb}}=t_0-t$ by using Eq. (2.28):

$$\frac{\mathrm{d}R}{\mathrm{d}z}=-\frac{R_0}{(1+z)^2},$$

which we plug into Eq. (2.26):

$$H=\frac{\mathrm{d}R}{\mathrm{d}t}\frac{1}{R}=-\frac{\mathrm{d}z}{(1+z)^2}\frac{1}{\mathrm{d}t}\frac{R_0}{R}=-\frac{\mathrm{d}z}{(1+z)}\frac{1}{\mathrm{d}t}. \tag{2.29}$$

Therefore, if one finds a relation for H as a function of z, which will be done in Sect. 2.1.4.4, one can calculate the lookback time:

$$t_{\mathrm{lb}}=t_0-t=\int_t^{t_0}\mathrm{d}t'=\int_0^z\frac{\mathrm{d}z'}{H(z')(1+z')}, \tag{2.30}$$

where the redshift at current time t_0 is zero, and the age of the Universe can be calculated by $z\to\infty$ (corresponds to $t\to 0$). From Eq. (2.27) we can further calculate the relation between the time interval when light is emitted $\mathrm{d}t_{\mathrm{e}}$ and observed $\mathrm{d}t_{\mathrm{o}}$. Here we use that the comoving coordinates are constant in cosmic time:

$$\frac{c\mathrm{d}t_{\mathrm{e}}}{R(t_{\mathrm{e}})}=\mathrm{d}\chi=\frac{c\mathrm{d}t_{\mathrm{o}}}{R(t_0)},$$

and therefore, we have

$$\frac{\mathrm{d}t_{\mathrm{o}}}{\mathrm{d}t_{\mathrm{e}}}=\frac{R_0}{R(t_{\mathrm{e}})}=1+z. \tag{2.31}$$

This equation states that SN light curves have a longer apparent timescale for an observer on Earth compared to the rest-frame timescale at the source location. This additional time helps to achieve, with monitoring, well-sampled SN light curves, ideally including the peak, to finally get the time delays from multiple well-sampled LSN Ia images.

Further, we note that the redshift, as listed in Eq. (2.28), is the cosmological redshift. Still, in reality, the total redshift will also include the redshift from the relativistic doppler effect as described in Eq. (2.15) from the peculiar motion of the light source with respect to a fundamental observer in the comoving frame (Davis & Scrimgeour, 2014). To calculate the total observed redshift z, it is necessary to first redshift the emitted wavelength λ_{e} from the inertial frame of the emitter[11] to a fundamental observer in the comoving frame located at the emitter's position, who will see the doppler-shifted wavelength λ_{c} [see Eq. (2.15)]. Further, we apply to λ_{c} the cosmological redshift [see Eq. (2.28)] leading to the total observed

---

[11] Ignoring the gravitational redshift from a light source in a gravitational potential.

redshift

$$1 + z = \frac{\lambda_\mathrm{o}}{\lambda_\mathrm{e}} = \frac{\lambda_\mathrm{c}}{\lambda_\mathrm{e}} \frac{\lambda_\mathrm{o}}{\lambda_\mathrm{c}} = (1 + z_\mathrm{doppler})(1 + z_\mathrm{cosmic}). \tag{2.32}$$

We see that for low doppler redshifts ($z_\mathrm{doppler} \ll 1$) and low cosmic redshifts ($z_\mathrm{cosmic} \ll 1$), the total redshift simplifies to the sum of both:

$$z = z_\mathrm{doppler} + z_\mathrm{cosmic}.$$

If, in addition, the cosmic redshift is much larger than the doppler redshift ($1 \gg z_\mathrm{cosmic} \gg z_\mathrm{doppler}$), the total redshift is simply the cosmic redshift:

$$z = z_\mathrm{cosmic}.$$

2.1.4.3 Distances

From the FLRW metric, we can calculate several distances. The coordinate distance or comoving distance χ has no physical meaning, and the proper distance

$$\sigma = R(t)\chi, \tag{2.33}$$

in analogy to Eq. (2.10), can not be measured in practice. Distances that can be calculated from observations are the angular diameter distance D_A and the luminosity distance D_lum. The angular diameter distance is defined as the apparent size of an object $\mathrm{d}\theta$ in comparison to the proper diameter $\mathrm{d}l$, where the latter can be determined from Eq (2.24) by using $\mathrm{d}t = \mathrm{d}\chi = \mathrm{d}\phi = 0$ and therefore $\mathrm{d}l = R(t)S(\chi)\mathrm{d}\theta = \frac{R_0 S(\chi)\mathrm{d}\theta}{(1+z)}$:

$$D_\mathrm{A} := \frac{\mathrm{d}l}{\mathrm{d}\theta} = \frac{R_0 S(\chi)}{1+z}. \tag{2.34}$$

The luminosity distance is defined via the luminosity $L = \mathrm{d}E_\mathrm{e}/\mathrm{d}t_\mathrm{e}$ of an isotropic source in comparison to the observed flux

$$F(t_0) = \left[\frac{\mathrm{d}E_\mathrm{o}}{\mathrm{d}t_\mathrm{o} A}\right](t_0).$$

The observed Energy $\mathrm{d}E_\mathrm{o}$ can be related via Eq. (2.28) to the emitted energy $\mathrm{d}E_\mathrm{e} = (1+z)\mathrm{d}E_\mathrm{o}$ given that every observed photon is redshifted in wavelength. Further, the observed time interval $\mathrm{d}t_\mathrm{o}$ can be related to the emitted one $\mathrm{d}t_\mathrm{e}$ via Eq. (2.31). The proper area A of a sphere centered around the emitter and reaching out to the observer can be calculated from the metric tensor as defined by the FLRW metric in Eq. (2.24) in comparison to Eq. (2.20):

$$A = \int_0^\pi \mathrm{d}\theta \int_0^{2\pi} \mathrm{d}\phi \sqrt{|g_{\theta\theta} g_{\phi\phi}|} = \int_0^\pi \mathrm{d}\theta \int_0^{2\pi} \mathrm{d}\phi \sqrt{R^4(t_0) S^4(\chi) \sin^2(\theta)} = 4\pi R^2(t_0) S^2(\chi).$$

Therefore we get

$$F(t_0) = \left[\frac{\mathrm{d}E_\mathrm{o}}{\mathrm{d}t_\mathrm{o}A}\right](t_0) = \left[\frac{\mathrm{d}E_\mathrm{e}}{\mathrm{d}t_\mathrm{e}(1+z)^2 4\pi R^2(t_0) S^2(\chi)}\right](t_0) = \frac{L}{4\pi R_0^2 S^2(\chi)(1+z)^2},$$

and the luminosity distance can be calculated via:

$$D_\mathrm{lum} := \sqrt{\frac{L}{4\pi F(t_0)}} = R_0 S(\chi)(1+z). \tag{2.35}$$

From the comparison of Eqs. (2.34) and (2.35) one finds the relation:

$$D_\mathrm{lum} = D_\mathrm{A}(1+z)^2. \tag{2.36}$$

Both distances are only comparable for low redshifts, given that the deviation is already around 2% for a redshift of $z = 0.01$. The luminosity and angular diameter distance depend on the comoving distance χ, and it is necessary to relate χ to observable quantities. Using Eqs. (2.27), (2.28), and (2.29) we find:

$$\chi = c\int_{t}^{t_*} \frac{\mathrm{d}t'}{R(t')} = \frac{c}{R_0}\int_{z_*}^{z} \frac{\mathrm{d}z'}{H(z')}, \tag{2.37}$$

where we have $z > z_*$ corresponding to $t_* > t$. Typically we have t_* equal to our present time and, therefore, $z_* = 0$. Still, we can also use this formula to calculate distances between different objects located at redshifts z and z_* along our line of sight, which will be relevant for gravitation lensing in Sect. 2.2.

2.1.4.4 Friedmann equations

All expressions derived or motivated so far rely only on the FLRW metric from Eq. (2.24) and the geodesic Eq. (2.21). However, we can learn more by solving the EFEs from Eq. (2.22) under the assumption of an energy-momentum tensor. Our Universe is composed of matter and radiation, and as shown in Eq. (2.22) it also allows for the cosmological constant Λ, which can be interpreted as the energy density of vacuum as we will see later. We model for now matter (index $i = \mathrm{m}$) and radiation (index $i = \mathrm{r}$) as a perfect fluid, for which we have an energy-momentum tensor of

$$T_i^{\mu\nu} = (\rho_i + \frac{p_i}{c^2})u^\mu u^\nu - p_i g^{\mu\nu}, \tag{2.38}$$

where $u^\mu = (c, 0, 0, 0)$ is the 4-velocity, ρ_i the proper density, and p_i the pressure. ρ_i and p_i are only functions of cosmic time t given that we assume a homogeneous and isotropic Universe. Further, we assume that the components of the cosmological fluid, for now, matter and radiation, do not interact with each other except due to their gravitational field. Therefore the total density ρ and total pressure p is simply the sum of the individual components i. By using now the FLRW metric in combinations with the EFEs for a perfect fluid, one can derive the cosmological field equations (Hobson et al., 2006, pp. 376-381), describing the evolution of the scale factor $R := R(t)$:

$$\ddot{R} = -\frac{4\pi G}{3}\sum_i\left(\rho_i + \frac{3p_i}{c^2}\right)R + \frac{1}{3}\Lambda c^2 R \tag{2.39}$$

$$\dot{R}^2 = \frac{8\pi G}{3}\sum_i \rho_i R^2 + \frac{1}{3}\Lambda c^2 R^2 - c^2 k. \tag{2.40}$$

These have first been derived by Friedmann (1922, 1924) and are therefore called the Friedmann equations. If one takes the derivative of Eq. (2.40) and plugs the resulting $\ddot{R}$ into Eq. (2.39), one gets for component i:

$$\dot{\rho}_i = -3\frac{\dot{R}}{R}\left(\rho_i + \frac{p_i}{c^2}\right). \tag{2.41}$$

Further, by assuming that our components of the cosmological fluid follow the equation of state

$$p_i = w_i \rho_i c^2, \tag{2.42}$$

the differential Eq. (2.41) has the solution

$$\rho_i = \rho_{i,0}\left(\frac{R_0}{R}\right)^{3(1+w_i)}, \tag{2.43}$$

where $\rho_{0,i}$ is the present-day proper density of component i, where we have $w_\mathrm{m} = 0$ for matter (pressureless dust) and $w_\mathrm{r} = 1/3$ for radiation (Hobson et al., 2006, pp. 386-387). Further, we can also relate the cosmological constant Λ to a constant energy density by using $w_\Lambda = -1$ and therefore, we can interpret the term $\frac{1}{3}\Lambda c^2$ from Eq. (2.40) as the energy density of the vacuum

$$\rho_\Lambda = \rho_{\Lambda,0} = \frac{\Lambda c^2}{8\pi G}, \tag{2.44}$$

which reduces Eqs. (2.39) and (2.40) to:

$$\ddot{R} = -\frac{4\pi G}{3}(\rho + \frac{3p}{c^2})R, \tag{2.45}$$

$$\dot{R}^2 = \frac{8\pi G}{3}\rho R^2 - c^2 k, \tag{2.46}$$

where the total density is

$$\rho = \rho_\mathrm{m} + \rho_\mathrm{r} + \rho_\Lambda = \rho_{\mathrm{m},0}\frac{R_0^3}{R^3} + \rho_{\mathrm{r},0}\frac{R_0^4}{R^4} + \rho_{\Lambda,0}. \tag{2.47}$$

We see that $\rho_\mathrm{m} \propto R^{-3}$, which is the expected behavior, namely, that the matter density decreases with the volume in an expanding Universe. For radiation, we have $\rho_\mathrm{r} \propto R^{-4}$ which has, in addition to the third power from the expanding volume, the factor $1/R$ coming from the cosmological redshift of the photons as described by Eq. (2.28). Therefore, in an expanding Universe, the early times are dominated by radiation and the late ones by the vacuum, where matter is the dominant property between these phases. We assumed here a Universe where the dark energy is equal to the cosmological constant. However, in reality, we could have, for example, a constant $w_{\rm de} \neq w_\Lambda = -1$, and therefore, our density would require the more general term $\rho_{\rm de}\,(R_0/R)^{3(1+w_{\rm de})}$. Further, Eq. (2.47) is valid for cold dark matter, meaning that the dark matter moves slowly compared to the speed of light and is therefore represented in $\Omega_{\rm m}$, but in principle, dark matter could have $w_{\rm dm} \neq w_{\rm m} = 0$, or several different forms of dark matter might exist. Anyway, this is beyond the scope of this thesis, and we continue with Eq. (2.47), which represents any ΛCDM cosmology supported by current observations (Riess et al., 1998; Perlmutter et al., 1999; Planck Collaboration, 2014, 2016, 2020).

Motivated by Eqs. (2.26) and (2.46), one can define the dimensionless density parameters:

$$\Omega_i := \frac{8\pi G}{3H^2}\rho_i \tag{2.48}$$

and

$$\Omega_k := -\frac{c^2 k}{H^2 R^2}, \tag{2.49}$$

which simplifies Eq. (2.46) to

$$1 = \Omega_{\rm m} + \Omega_{\rm r} + \Omega_\Lambda + \Omega_k. \tag{2.50}$$

Further, we can use

$$\Omega_{i,0} = \frac{8\pi G \rho_{i,0}}{3H_0^2} \qquad \text{and} \qquad \Omega_{k,0} = -\frac{c^2 k}{H_0^2 R_0^2}, \tag{2.51}$$

in combination with Eqs. (2.28) and (2.47) to rewrite Eq. (2.50) as:

$$H^2 = H_0^2\Big(\Omega_{\rm m,0}(1+z)^3 + \Omega_{\rm r,0}(1+z)^4 + \Omega_{\Lambda,0} + \Omega_{k,0}(1+z)^2\Big). \tag{2.52}$$

We see that the evolution of our Universe, described by the Hubble parameter $H = \dot{R}/R$, depends on five parameters, namely the current value of the Hubble parameter H_0, also referred to as the Hubble constant, and the current values of the four density parameters for matter, radiation, vacuum energy, and curvature. However, it is enough to determine three of them given the relation from Eq. (2.50). Therefore our cosmological model as derived here is fully described by $(H_0, \Omega_{\rm m,0}, \Omega_{\rm r,0}, \Omega_{\Lambda,0})$ and it is of great interest in observational cosmology to measure these parameters.

2.2 Gravitational lensing

Following up on the introduction to gravitational lensing in Sect. 1.2, we summarize the most important points from a more theoretical perspective. A more detailed summary of the topic can be found in Schneider et al. (2006).

2.2.1 Lens equation and deflection angle

In Fig. 2.1, a gravitational lensing event is shown, where light from a source gets deflected by a massive lens under the angle $\hat{\alpha}$, such that the source appears at the image position $\vec{\theta}$, although it is located at the source position $\vec{\beta}$. Source and image position are connected via the lens equation:

$$\vec{\beta} = \vec{\theta} - \vec{\alpha}(\vec{\theta}). \tag{2.53}$$

The scaled deflection angle $\vec{\alpha}$ can be related to the deflection angle $\hat{\alpha}$ via simple geometry from Fig. 2.1, assuming small angles[12] and using Eq. (2.53):

$$\vec{\alpha}(\vec{\theta}) = \frac{D_{\mathrm{ds}}}{D_{\mathrm{s}}}\hat{\alpha}. \tag{2.54}$$

The distances D_{ds}, D_{s}, and D_{d}, as shown in Fig. 2.1, are angular diameter distances from the lens to the source, the observer to the source, and the observer to the lens. The angular diameter distance between two redshifts z_1 and z_2, where $z_2 > z_1$ can be calculated from Eqs. (2.34), (2.37), and (2.52) substituting $z \rightarrow z_2$ and $z_* \rightarrow z_1$:

$$D_{\mathrm{A}}(z_2, z_1) = \frac{R_0}{1+z_2} S\left(\frac{c}{R_0}\int_{z_1}^{z_2} \frac{\mathrm{d}z}{H_0\sqrt{\Omega_{\mathrm{m},0}(1+z)^3 + \Omega_{\mathrm{r},0}(1+z)^4 + \Omega_{\Lambda,0} + \Omega_{k,0}(1+z)^2}}\right). \tag{2.55}$$

We, as an observer, are located at redshift 0, the source is at z_{s}, and the lens at z_{d}. Eq. (2.55) is valid in any ΛCDM model (cold dark matter + dark energy in its simplest form, the cosmological constant). If we assume in addition a flat model, meaning $\Omega_{k,0} = 0$ and $S(\chi) = \chi$, the angular diameter distance simplifies to

$$D_{\mathrm{A}}(z_2, z_1) = \frac{c}{H_0(1+z_2)}\int_{z_1}^{z_2} \frac{\mathrm{d}z}{\sqrt{\Omega_{\mathrm{m},0}(1+z)^3 + \Omega_{\mathrm{r},0}(1+z)^4 + \Omega_{\Lambda,0}}}. \tag{2.56}$$

The deflection angle $\hat{\alpha}$ can be calculated from GR under the assumption of a flat Minkowski metric $\eta_{\mu\nu}$ [see Eq. (2.17)] with a small perturbation $h_{\mu\nu}$, where $|h_{\mu\nu} \ll 1|$, and therefore, the metric tensor is

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}. \tag{2.57}$$

From this, one can calculate the linearised field equations (Hobson et al., 2006, pp. 467-472), starting from the EFEs in Eq. (2.22). For the Newtonian limit, one considers static sources, where the particles of the source are at rest, and the only component of the energy-momentum tensor different from zero is the rest energy $T^{00} = \rho c^2$, where ρ is the proper density distribution of the source. For static sources, the linearised field equations provide the solution that the only non-zero elements of the perturbation $h_{\mu\nu}$ are $h_{00} = h_{11} = h_{22} = h_{33} = 2\Phi/c^2$ (Hobson et al., 2006, pp. 483-486), where Φ is the gravitational potential as defined in Eq. (2.2). For a static point mass M, the gravitational potential is $\Phi = -GM/r$

---

[12] Small angle approximations, e.g., $\vec{\beta} \approx \tan(\vec{\beta}) = \frac{\vec{\eta}}{D_{\mathrm{s}}}$ is valid, given that the distance to the source is much larger than the offset of the source from the line of sight connecting observer and lens.

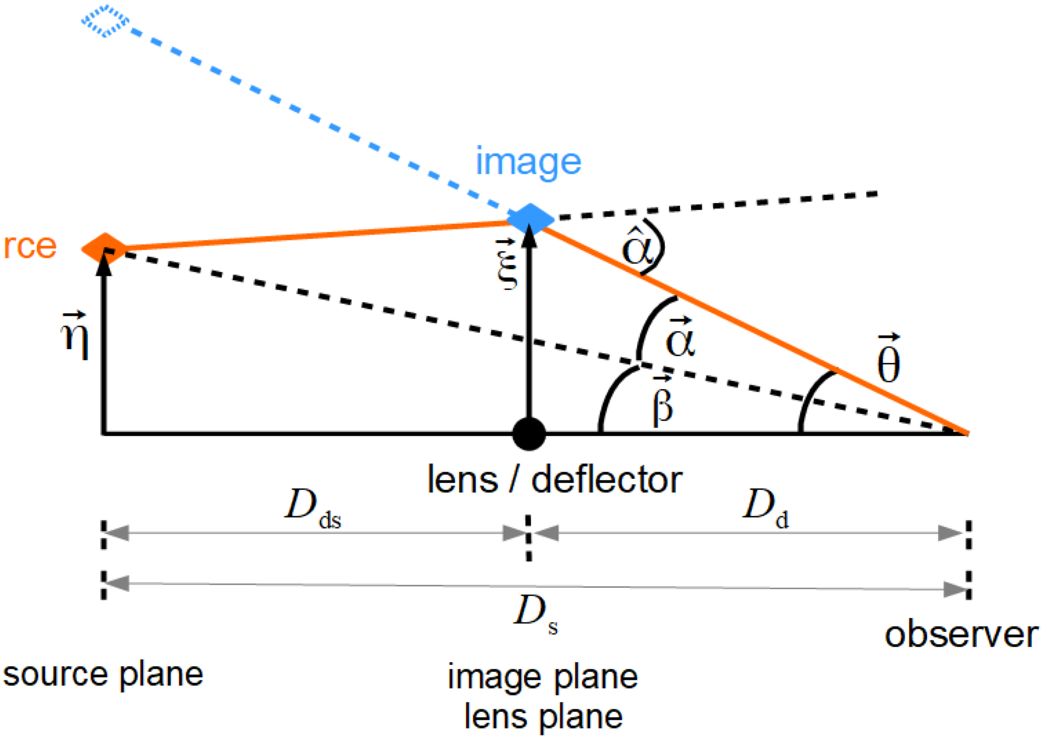


Fig. 2.1. Illustration of gravitational lensing where light from the source is deflected by the lens by an angle $\hat{\alpha}$. An observer sees the image of the source at position $\vec{\theta}$, different from the source position $\vec{\beta}$.

[see Eq. (2.4)], which leads to the following line element in spherical coordinates:

$$\mathrm{d}s^2 = \left(1 - \frac{2GM}{c^2 r}\right) c^2\,\mathrm{d}t^2 - \left(1 + \frac{2GM}{c^2 r}\right)\left(\mathrm{d}r^2 + r^2\,\mathrm{d}\theta^2 + r^2 \sin^2\theta\,\mathrm{d}\phi^2\right). \tag{2.58}$$

From the geodesic Eq. (2.21), one can calculate the path of light in such a metric to determine the deflection angle $\hat{\alpha}$ of an object passing by a point mass M at a distance $\vec{\xi}$, as shown in Fig. 2.1:

$$\hat{\alpha} = \frac{4GM}{c^2 \xi^2}\vec{\xi} = 1.75'' \frac{M}{M_\odot}\frac{R_\odot}{\xi^2}\vec{\xi} \tag{2.59}$$

(Carroll, 2013, pp. 286-293), where $M_\odot \approx 2.0 \times 10^{30}\,\mathrm{kg}$ is the solar mass, and $R_\odot \approx 7.0 \times 10^{8}\,\mathrm{m}$ the solar radius. The deflection angle can also be derived from the Schwarzschild metric (Hobson et al., 2006, pp. 233-234), where one finds the requirement that ξ needs to be much greater than the Schwarzschild radius of a mass $R_\mathrm{s} = 2GM/c^2$, such that the weak field approximation is justified (Schneider et al., 2006, pp. 18-19).

The predicted value for $\hat{\alpha}$ from GR is twice that of Newtonian gravity. However, it is fundamentally unclear how mass should deflect light in Newtonian gravity, in contrast to GR, where it is a natural consequence of curved spacetime. The discrepancy between Newtonian physics and GR led to the famous expedition from Eddington (1919), who observed the position of stars close to the sun during a solar eclipse. According to Eq. (2.59), the maximum deflection of the background stars is 1.75″ in GR. The measurement done by Eddington (1919) was more consistent with GR than Newtonian gravity, which made Einstein and his new theory of gravity famous. Later, doubts were raised if the measurement was exact enough to distinguish between Newtonian physics and GR. However, in a reanalysis, Harvey (1979) came to $1.83'' \pm 0.13''$ in excellent agreement with GR but more than a factor of two away from Newtonian gravity. Until today GR has passed many precision tests, and gravitational lensing is a well-established tool in astronomy, as already highlighted in Sect. 1.2.

2.2.2 Lens potential, convergence, shear, and magnification

We considered a point mass in Eq. (2.59). Now, if we have an extended source, the total deflection is the sum of the individual deflections of the mass distribution, as long as we are in the weak field approximation, as defined by Eq. (2.57). A three-dimensional lens-mass distribution, as shown in Fig. 2.2, can be divided into small mass elements dm connected to the volume element dV via the mass density ρ:

$$\mathrm{d}m = \rho(\vec{r}\,')\,\mathrm{d}V = \rho(\vec{r}\,')\,\mathrm{d}r_3'\,\mathrm{d}^2\xi'.$$

The location of the small mass element is described by the vector $\vec{r}\,' = (\vec{\xi}\,', r_3')$, and the light ray passing by the mass distribution is represented by the vector $\vec{r} = (\vec{\xi}(u), r_3(u))$. The two-dimensional plane $\vec{\xi}$ is perpendicular to the line of sight, meaning that light close to the observer propagates in the direction where the r_3 coordinate points. If the deflection is small, we can approximate the light ray as a straight line. Therefore, $\vec{\xi}(u) = \vec{\xi}$ and a deflector satisfying this is called a geometrically-thin lens (Schneider et al., 2006, p. 19). In this approximation, the impact parameter connecting a mass element dm and the light ray is independent of r_3' and, thus, simply $\vec{\xi} - \vec{\xi}\,'$. Replacing in Eq. (2.59) the impact parameter and the mass distribution, we get

$$\hat{\alpha}(\vec{\xi}) = \frac{4G}{c^2} \int \mathrm{d}^2\xi' \int \mathrm{d}r_3'\,\rho(\vec{\xi}\,', r_3')\,\frac{\vec{\xi} - \vec{\xi}\,'}{|\vec{\xi} - \vec{\xi}\,'|^2}. \tag{2.60}$$

This motivates the definition of the surface mass density

$$\Sigma(\vec{\xi}) = \int \mathrm{d}r_3\,\rho(\vec{\xi}, r_3), \tag{2.61}$$

in units of kg m^{-2}. The scaled deflection angle following Eq. (2.54) and using $\vec{\theta} = \vec{\xi}/D_\mathrm{d}$ is then

$$\begin{aligned}\vec{\alpha}(\vec{\theta}) &= \frac{D_\mathrm{ds}}{D_\mathrm{s}}\,\hat{\alpha}(D_\mathrm{d}\vec{\theta}) = \frac{4G}{c^2}\frac{D_\mathrm{ds}}{D_\mathrm{s}} \int \mathrm{d}^2\theta'\,D_\mathrm{d}^2\Sigma(D_\mathrm{d}\vec{\theta}\,')\,\frac{\vec{\theta} - \vec{\theta}\,'}{|\vec{\theta} - \vec{\theta}\,'|^2}\,\frac{D_\mathrm{d}}{D_\mathrm{d}^2} \\ &= \frac{4G}{c^2}\frac{D_\mathrm{ds}D_\mathrm{d}}{D_\mathrm{s}} \int \mathrm{d}^2\theta'\,\Sigma(D_\mathrm{d}\vec{\theta}\,')\,\frac{\vec{\theta} - \vec{\theta}\,'}{|\vec{\theta} - \vec{\theta}\,'|^2}.\end{aligned}$$

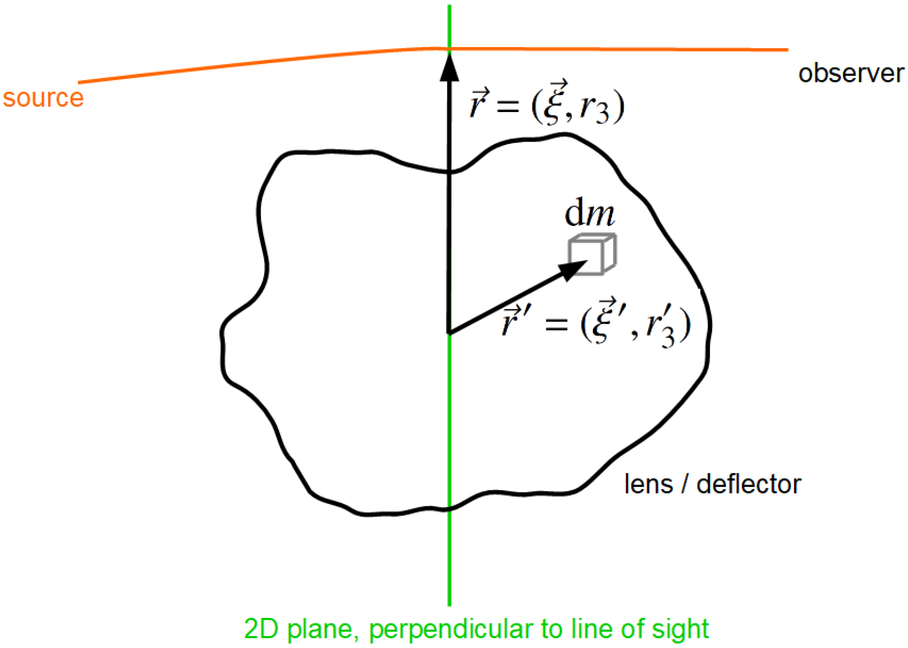


Fig. 2.2. Gravitational lensing by an extended lens. The total deflection is the sum of all the individual deflections from all mass elements dm.

Following this, we define the “dimensionless surface mass density” or “convergence”:

$$\kappa(\vec{\theta}) = \frac{\Sigma(D_\mathrm{d}\vec{\theta})}{\Sigma_\mathrm{c}}, \tag{2.62}$$

where the critical surface mass density is

$$\Sigma_\mathrm{c} = \frac{c^2}{4\pi G}\frac{D_\mathrm{s}}{D_\mathrm{ds}D_\mathrm{d}}, \tag{2.63}$$

simplifying the scaled deflection angle to

$$\vec{\alpha}(\vec{\theta}) = \frac{1}{\pi}\int \mathrm{d}^2\theta'\,\kappa(\vec{\theta}\,')\,\frac{\vec{\theta}-\vec{\theta}\,'}{|\vec{\theta}-\vec{\theta}\,'|^2}. \tag{2.64}$$

Eq. (2.64) motivates the definition of the lens potential as

$$\Psi(\vec{\theta}) = \frac{1}{\pi}\int \mathrm{d}^2\theta'\,\kappa(\vec{\theta}\,')\,\ln|\vec{\theta}-\vec{\theta}\,'| \tag{2.65}$$

leading to

$$\vec{\alpha} = \nabla_\theta \Psi(\vec{\theta}). \tag{2.66}$$

From using $\nabla^2 \ln|\vec{\theta}| = 2\pi\delta_\mathrm{D}(\vec{\theta})$, where $\delta_\mathrm{D}(\vec{\theta})$ is the Dirac delta, we find the Poisson equation in two dimensions:

$$\nabla_\theta^2 \Psi(\vec{\theta}) = 2\kappa(\vec{\theta}), \tag{2.67}$$

which relates the convergence to the lens potential:

$$\kappa = \frac{1}{2}\left(\frac{\partial^2\Psi(\vec{\theta})}{\partial\theta_1^2} + \frac{\partial^2\Psi(\vec{\theta})}{\partial\theta_2^2}\right). \tag{2.68}$$

From the lens potential, one can also define the shear

$$\gamma = \gamma_1 + \mathrm{i}\gamma_2 = |\gamma|\mathrm{e}^{2\mathrm{i}\phi}, \tag{2.69}$$

where

$$\gamma_1 = \frac{1}{2}\left(\frac{\partial^2\Psi}{\partial\theta_1^2} - \frac{\partial^2\Psi}{\partial\theta_2^2}\right) \qquad \text{and} \qquad \gamma_2 = \frac{\partial^2\Psi}{\partial\theta_1\,\partial\theta_2}. \tag{2.70}$$

As a next step, we calculate the Jacobian matrix of the lens equation [see Eq. (2.53)]:

$$A(\vec{\theta}) = \begin{pmatrix} \frac{\partial\beta_1}{\partial\theta_1} & \frac{\partial\beta_1}{\partial\theta_2} \\ \frac{\partial\beta_2}{\partial\theta_1} & \frac{\partial\beta_2}{\partial\theta_2} \end{pmatrix} = \begin{pmatrix} 1 - \frac{\partial^2\Psi}{\partial\theta_1^2} & -\frac{\partial^2\Psi}{\partial\theta_2\partial\theta_1} \\ -\frac{\partial^2\Psi}{\partial\theta_1\partial\theta_2} & 1 - \frac{\partial^2\Psi}{\partial\theta_2^2} \end{pmatrix} = \begin{pmatrix} 1-\kappa-\gamma_1 & -\gamma_2 \\ -\gamma_2 & 1-\kappa+\gamma_1 \end{pmatrix}. \tag{2.71}$$

With this matrix, one can demonstrate the impact of convergence κ and shear γ on the image of a source. If $\det(A) \neq 0$, A is invertible, and one can calculate the impact of a small variation on the source plane, onto the image plane, via $\mathrm{d}\vec{\theta} = A^{-1}\mathrm{d}\vec{\beta}$. For a small circular source with radius R, one finds that the image gets distorted into an ellipse, as shown in Fig. 2.3, with semi-major and semi-minor axis equal to the radius R times the absolute value of the eigenvalues $\tilde{a}_1$ and $\tilde{a}_2$ of the inverse Jacobian matrix A^{-1}:

$$\tilde{a}_1 = \frac{1}{1-\kappa-|\gamma|}, \qquad \tilde{a}_2 = \frac{1}{1-\kappa+|\gamma|}. \tag{2.72}$$

The orientation of the elliptical image with respect to the circular source is determined by ϕ, as listed in Eq. (2.69). For $\gamma = 0$, we have $\tilde{a}_1 = \tilde{a}_2$, and therefore, the circular source is mapped to a circular image, although its size changed. In general, the area Ω_{ellipse} covered by the ellipse will be different from the area Ω_{circle} covered by the circular source, and the ratio between these two quantities is defined as the magnification

$$\mu = \frac{\Omega_{\text{ellipse}}}{\Omega_{\text{circle}}} = \frac{\pi R\tilde{a}_1 R\tilde{a}_2}{\pi R^2} = \tilde{a}_1\tilde{a}_2 = \frac{1}{(1-\kappa)^2 - |\gamma|^2}. \tag{2.73}$$

An important point to note is that lensing conserves surface brightness, respectively, the specific intensity, which follows from the Liouville theorem, and the absence of absorption or emission of photons during the deflection of light (Schneider et al., 2006, p. 23).

As a reminder, the specific intensity I_λ is defined as the differential energy $\mathrm{d}E$ going through an effective detector area[13] $\mathrm{d}A\cos\theta$ in the time interval $\mathrm{d}t$ coming from a solid angle $\mathrm{d}\Omega$ of a source in the sky in the wavelength range $[\lambda, \lambda+\mathrm{d}\lambda]$:

$$I_\lambda = \frac{\mathrm{d}E}{\mathrm{d}A\cos\theta\,\mathrm{d}t\,\mathrm{d}\Omega\,\mathrm{d}\lambda}. \tag{2.74}$$

---

[13] θ is the angle between the normal of $\mathrm{d}A$ and the direction to the source.

The units of the specific intensity are $\mathrm{J\,s^{-1}\,m^{-2}\,\AA^{-1}\,ster^{-1}}$. The surface brightness is then defined as:

$$S_\mathrm{B} = \int \mathrm{d}\lambda\, I_\lambda. \tag{2.75}$$

We can calculate the radiation flux[14] F_λ via

$$F_\lambda = \int_{\Omega_0} \mathrm{d}\Omega\, I_\lambda \cos\theta, \tag{2.76}$$

for a source covering the solid angle Ω_0. The flux F is then

$$F = \int \mathrm{d}\lambda\, F_\lambda = \int_{\Omega_0} \mathrm{d}\Omega\, S_\mathrm{B} \cos\theta. \tag{2.77}$$

Given that gravitational lensing conserves surface brightness, a larger apparent area of the image, compared to the source, leads to an increased flux. Therefore the image is brighter due to lensing. From our example, in Eq. (2.73), we see that the last term corresponds to one over the determinate of the Jacobian matrix. In general, we can define the magnification factor of a small source via

$$\mu = \frac{\Omega_\mathrm{image}}{\Omega_\mathrm{source}} = \frac{F_\mathrm{o}}{F_\mathrm{e}} = \frac{1}{\det(A)} = \frac{1}{(1-\kappa)^2 - |\gamma|^2}, \tag{2.78}$$

where Ω_image and Ω_source are the areas covered on the sky by the image, respectively source. The observed flux is F_o, and the emitted one is F_e. Fig. 2.3 shows that the shape distortion comes only from the shear γ, where magnification (or demagnification if $|\mu| < 1$) is caused by convergence and shear. Gravitational lensing is achromatic because light deflection [see Eq. (2.59)] does not depend on the wavelength of the light. Therefore, the magnification μ, as defined for a small source in Eq. (2.78), is also achromatic. However, an extended source will cover regions where the magnification factor can vary. If the specific intensity I_λ of the source varies on similar scales for different wavelengths, the overall magnification of the source will be wavelength-dependent. We will learn more about that in Sec 2.3.2. Anyway, the magnification of lensing can be a helpful feature for observing faint sources at high redshifts.

2.2.3 Fermat potential and multiple images

Multiple images are produced if, for a single source position β, multiple image positions θ solve the lens equation [Eq. (2.53)]. A useful quantity to distinguish between different image types is the Fermat potential:

$$\tau(\vec{\theta},\vec{\beta}) = \frac{1}{2}(\vec{\theta}-\vec{\beta})^2 - \Psi(\vec{\theta}), \tag{2.79}$$

which is defined such that $\nabla_\theta \tau = 0$ yields the lens equation. This is motivated by Fermat's principle which states that physical light rays travel along paths where the light travel time

---

[14] F_λ has units of $\mathrm{J\,s^{-1}\,m^{-2}\,\AA^{-1}}$.

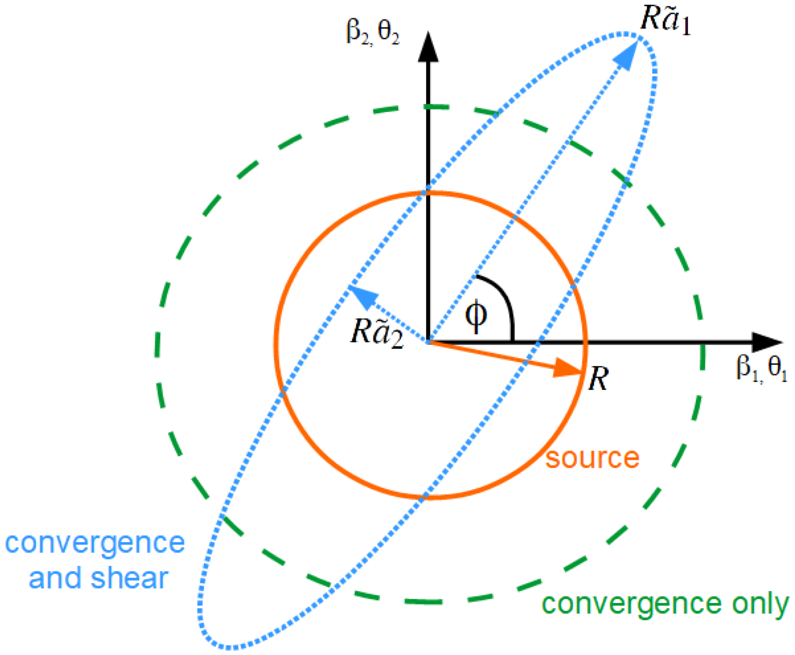


Fig. 2.3. Impact of convergence κ and shear γ on a small circular source with radius R (solid orange). The convergence only leads to a magnification (dashed green) but shear in addition causes a circular source to appear elliptical (dotted blue) with semi-major axis $R\tilde{a}_1$ and semi-minor axis $R\tilde{a}_2$.

t is stationary, meaning that $\nabla t = 0$. Therefore, the Fermat potential τ is proportional to the light travel time t, and physical images of the source occur where $\nabla_\theta \tau = 0$. It is straightforward to show that the Hessian matrix of the Fermat potential ($\tau_{ij} = \frac{\partial^2 \tau}{\partial\theta_i \partial\theta_j}$) is the Jacobian Matrix, as listed in Eq. (2.71). The Hessian matrix, in general, tells us the image type. If the Hessian matrix is at a stationary point positive-definite (all eigenvalues > 0), it is a local minimum, if it is negative-definite (all eigenvalues < 0), it is a local maximum, and if it is indefinite (eigenvalues have different signs), it is a saddle point. In our two-dimensional case, the determinant and trace of the Hessian matrix define the image type. Further, the eigenvalues of the matrix in Eq. (2.71) are:

$$a_1 = 1 - \kappa - |\gamma|, \qquad a_2 = 1 - \kappa + |\gamma|. \tag{2.80}$$

From this, one can define three different image types with the following properties:

- Type I image defined as minimum of $\tau \Leftrightarrow a_i > 0 \Leftrightarrow \det(\mathrm{A}) > 0$ and $\mathrm{trace}(\mathrm{A}) > 0$.

  This leads to $|\gamma| < 1 - \kappa \leq 1$ and $0 \leq \kappa < 1$, where $\kappa \geq 0$ comes simply from the definition of convergence in Eq. (2.62). For the magnification, we find $\mu > 1$; therefore, minimum images are always magnified and have a positive parity (since $\mu > 0$).

- Type II image defined as saddle point of $\tau \Leftrightarrow a_1 > 0 > a_2$ or $a_2 > 0 > a_1 \Leftrightarrow \det(\mathrm{A}) < 0$.

  This leads to the relation $(1-\kappa)^2 < |\gamma|^2$; therefore, saddle images can be magnified ($|\mu| > 1$) or demagnified ($|\mu| < 1$), but they always have a negative parity ($\mu < 0$).

- Type III image defined as maximum of $\tau \Leftrightarrow a_i < 0 \Leftrightarrow \det(\mathrm{A}) > 0$ and $\mathrm{trace}(\mathrm{A}) < 0$.

From this we find $(1-\kappa)^2 > |\gamma|^2$ and $\kappa > 1$. The absolute value of the magnification of a maximum image can take any value; however, in many cases, it is demagnified. Further, maximum images always have a positive parity ($\mu > 0$).

For a lens with a realistic mass distribution, the gravitational deflection will vanish for large $|\vec{\theta}|$ and, therefore, the Jacobian matrix converges to diag(1,1), which has det(A) > 0 and trace(A) > 0, corresponding to a minimum image. Thus, gravitational lensing produces at least one minimum image independent of the source position.

For all locations where $\det(A) \neq 0$, the Jacobian matrix is invertible. Thus, a small shift of the source position within that region leads only to a small shift of the image position ($\mathrm{d}\vec{\theta} = A^{-1}\mathrm{d}\vec{\beta}$), and therefore, the number of images does not change. Consequently, a changing number of images is related to the crossing of $\det(A) = 0$. The curves on the image plane related to $\det(A) = 0$ are called "critical curves", and they divide the image plane into regions of positive parity ($\mu > 0$) and negative parity ($\mu < 0$). The critical curves can be mapped onto the source plane via the lens equation and are called "caustics". Critical curves are closed and smooth curves, but caustics can form "cusps". The smooth part of caustics is called "fold". One can show that for gravitational lensing, the odd number theorem (Burke, 1981; McKenzie, 1985) is valid; for a lens with a smooth surface mass density, that decreases for $|\theta| \to \infty$ faster than $1/|\theta|$, the number of images with negative parity (saddle images) equals the number of images with positive parity (minimum or maximum) - 1 (Schneider et al., 2006, pp. 27-28). Therefore, if the source crosses a caustic on the source plane, two images with opposite parity are created or vanish on the image plane. This depends on the direction of the caustic crossing on the source plane. In Fig. 2.4, we can see five different lensing systems and, depending on the source position, the number and position of the multiple images.

2.2.4 Weak lensing

In the previous subsection, we discussed the occurrence of multiple images. If the source is located in the outer regime of a lens such that no caustics have been crossed, only one image of the source appears, which is distorted and magnified. In this case, we are talking about weak lensing (Mellier, 1999; Bartelmann & Schneider, 2001). Given that the source's intrinsic shape and brightness are unknown and that the amplitude of the weak-lensing distortion is generally small for a single galaxy, this effect can only be observed statistically. As discussed in Sect. 2.2.2, a small circular source gets mapped onto an ellipse. The semi-major axis of the image is tangentially orientated with respect to the lens center. However, background source galaxies are not intrinsically round, and the observed ellipticity is a combination of lensing and intrinsic ellipticity. By assuming that the intrinsic ellipticity of many galaxy images around a lens is random, one can determine the impact of weak lensing by averaging over all these images and, thus, get a map of the gravitational potential.

Such an approach is used to constrain the total mass distribution in the merging galaxy cluster 1E 0657-558 (Tucker et al., 1998), also known as Bullet Cluster, with two primary galaxy concentrations moving away from each other after the collision. In a galaxy cluster,

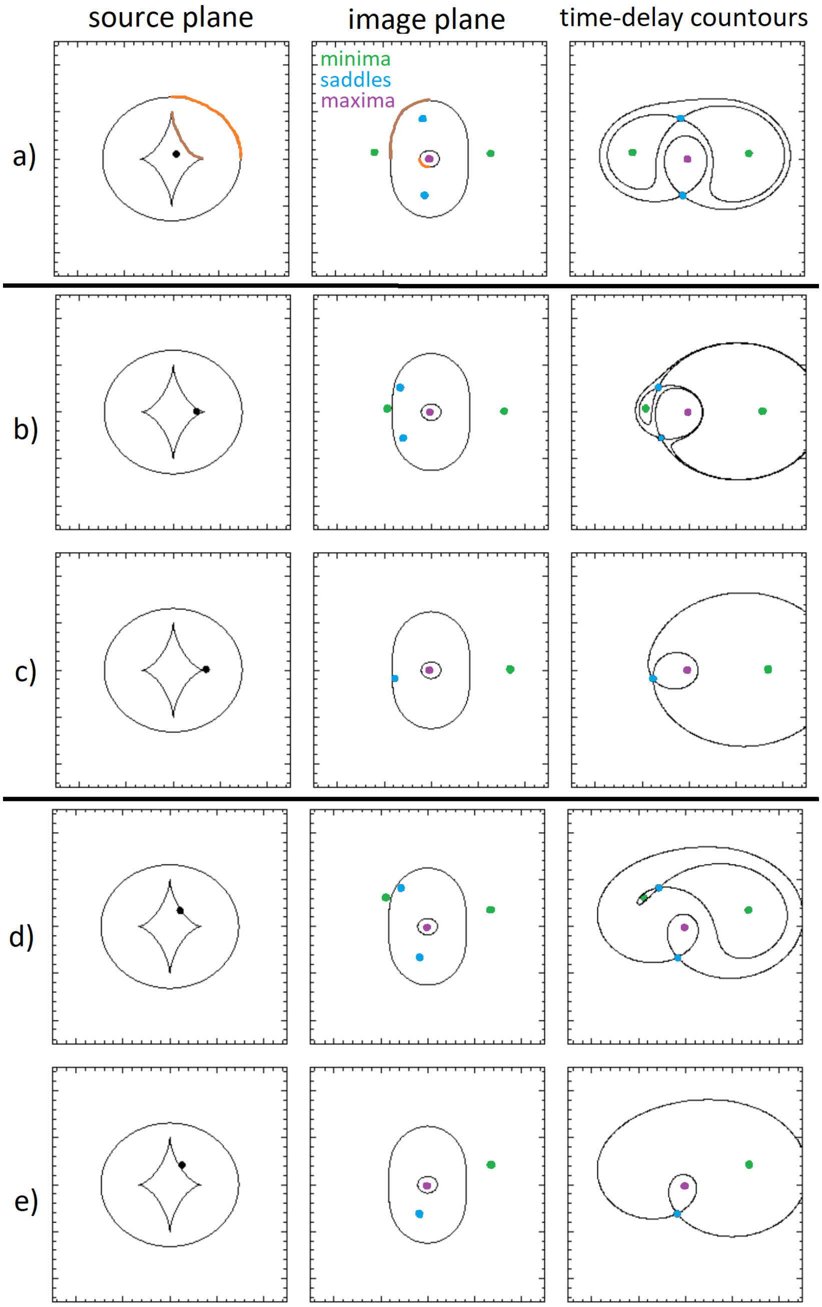


Fig. 2.4. Panels in the left column show source positions and caustics, panels in the middle column show image positions and critical curves, and panels in the right column show the time-delay contours on the image planes. In row a), we see a system producing five multiple images. The lens is located in the center, and the maximum image there is often highly demagnified. Therefore, the maximum image is typically not observable, and we see just four multiple images, which we refer to as a quad system. The orange and brown segments show the relation between critical curves from the image plane mapped onto caustics on the source plane. Row b) shows a cusp quad system. Row c) is very similar to Row b), but the source position changed such that a caustic was crossed, and two images disappeared. Also, the central maximum is typically not observable for such an image system, and we refer to it as a double system. Panel d) shows a fold quad system, and Panel e) shows the same system with a slightly moved source position, where two images disappeared. The source planes have a scale of half that of the image planes. Image taken from Courbin et al. (2002) with colored modifications.

the mass of the X-ray emitting plasma (Allen et al., 2002; Vikhlinin et al., 2006) is roughly five times higher than the stellar mass within the galaxies of the cluster (Kochanek et al., 2003). During the collision of two galaxy clusters, the plasma experiences ram pressure, unlike the galaxies, which behave like collisionless particles. Therefore, the plasma spatially decouples from the galaxies during such a merger. Clowe et al. (2006) created a κ map for the Bullet Cluster using weak gravitational lensing, showing that the gravitational potential traces the galaxies instead of the significantly more massive plasma, providing a piece of strong empirical evidence for the existence of collisionless dark matter.

There are also many other studies where weak lensing is used to constrain mass distributions of galaxy clusters and detect substructures in the cluster (e.g., Geiger & Schneider, 1998; King & Schneider, 2001; Cypriano et al., 2004; Jee et al., 2004). Further, one can search for weak lensing in random patches of the sky to determine the cosmic shear and, therefore, directly measure the large-scale structures in the Universe to probe the distribution of dark matter (e.g., Wittman et al., 2000; Bacon et al., 2000; Hamana et al., 2003). Large-scale structures between the last scattering surface and an observer also impact the CMB. In general, all gravitational potential wells along the line of sight lead to a weak lensing event, and their net impact is represented in the weak lensing power spectrum. In the case of the CMB, this is represented in the CMB lensing power spectrum, which can constrain H_0 with an external determination of Ω_m (e.g., Baxter & Sherwin, 2021). This can provide a measurement based on CMB observations independent of the sound horizon[15] r_s, which is interesting given that, for example, Bernal et al. (2016); Aylor et al. (2019) pointed out that H_0 probes, providing a lower H_0 value, assume that r_s from the standard cosmological model is correct.

2.2.5 Strong lensing and microlensing

From the lens equation for a point mass lens, one can estimate the image separation between two images at position $\vec{\theta}_1$ and $\vec{\theta}_2$ as

$$\begin{aligned}\Delta\vec{\theta} = \vec{\theta}_2 - \vec{\theta}_1 = \vec{\alpha}_2 - \vec{\alpha}_1 = \frac{D_\mathrm{ds}}{D_\mathrm{s}}(\hat{\alpha}_2 - \hat{\alpha}_1) = \frac{4GM}{c^2}\frac{D_\mathrm{ds}}{D_\mathrm{s}}\left(\frac{\vec{\xi}_2}{\xi_2^2} - \frac{\vec{\xi}_1}{\xi_1^2}\right) \\ = \frac{4GM}{c^2}\frac{D_\mathrm{ds}}{D_\mathrm{s}D_\mathrm{d}}\left(\frac{\vec{\theta}_2}{\theta_2^2} - \frac{\vec{\theta}_1}{\theta_1^2}\right) = \theta_\mathrm{Ein}^2\left(\frac{\vec{\theta}_2}{\theta_2^2} - \frac{\vec{\theta}_1}{\theta_1^2}\right)\end{aligned}, \tag{2.81}$$

---

[15] The sound horizon is the comoving distance a sound wave traveled between the Big Bang and the last scattering surface. The Universe before the CMB emission was a plasma composed of baryons and electrons, where photons were basically trapped because of their short free streaming length. This led to oscillations in overdense regions of the baryon-photon fluid due to gravitational contraction and counteracting radiation pressure. These oscillations are imprinted in the CMB power spectrum, which shows temperature fluctuations of the CMB as a function of multipole moments l (angular separation). The angular size of the sound horizon is represented by the first main peak in the power spectrum at $l \approx 200$, where all overdensities of the baryon-photon fluid of a specific size just reached its maximal density.

where in the last step, the Einstein radius is defined:

$$\theta_{\mathrm{Ein}} = \sqrt{\frac{4GM}{c^2}\frac{D_{\mathrm{ds}}}{D_{\mathrm{s}}D_{\mathrm{d}}}}, \tag{2.82}$$

which sets the order of magnitude for the image separation. We can rewrite the Einstein radius as

$$\theta_{\mathrm{Ein}} = 2.85'' \sqrt{\frac{M}{10^{12}M_{\odot}}\frac{D_{\mathrm{ds}}1\mathrm{Gpc}}{D_{\mathrm{s}}D_{\mathrm{d}}}} = 0.901'' \times 10^{-3} \sqrt{\frac{M}{M_{\odot}}\frac{D_{\mathrm{ds}}10\mathrm{kpc}}{D_{\mathrm{s}}D_{\mathrm{d}}}}. \tag{2.83}$$

2.2.5.1 Strong lensing

The first expression of Eq. (2.83) shows that on extragalactic scales, the image separation is on the order of arcseconds for galaxy lenses with masses comparable to the Milky Way ($1.15 \times 10^{12} M_{\odot}$, Carlesi et al., 2022). A galaxy cluster is typically 100 to 1000 times more massive than a galaxy leading to image separations 10 to 30 times larger than galaxy scale lenses. Of course, the point mass approximation is only a rough estimate, but it gives the order of magnitude of the image separation. For galaxy lenses, the image separation is around arcseconds, and for galaxy cluster lenses, around tens of arcseconds. This is, for example, in good agreement with the observation shown in Fig. 1.3. Panel b) shows a galaxy scale lens producing images S1 to S4, and the comparison of image SX to S1 to S4, shows the image separation on cluster lensing scales. In the case where we have galaxies or galaxy clusters as lenses, we refer to it as strong lensing.

Strong gravitational lensing is a powerful tool for measuring the total mass of the lens. As we saw in our example of a point mass lens, one can measure the Einstein radius [see Eq. (2.81)] from the position of two images. This is further related to the mass M and the distances $D_{\mathrm{s}}, D_{\mathrm{d}}$, and D_{ds} [see Eq. (2.82)]. From measuring the redshift of the source and the lens, one can calculate the angular diameter distances via Eq. (2.55) under the assumption of a cosmological model. Thus, one can determine the mass M of the point mass lens. Of course, this is very simplified, given that real lenses are not point-like, and we can have multiple lensing events between the source and the observer.

In a real application, the multiple images are used to construct a lens mass model using well-established modeling software packages like `GLEE` (Suyu & Halkola, 2010; Suyu et al., 2012) or `glafic` (Oguri, 2010, 2021). Thus, strong lensing can typically constrain the mass within the Einstein radius, though the definition of θ_{Ein} deviates from Eq. (2.82), dependent on the lens model. Still, as soon as the mass model of the lens is set, one can further constrain the dark matter distribution by assuming a mass-to-light ratio (e.g., Suyu et al., 2012; Grillo et al., 2013; Schuldt et al., 2019).

Another powerful application is time-delay cosmography, as highlighted already in Sect. 1.2, which we will further elaborate on in Sect. 2.2.6.

2.2.5.2 Microlensing

Microlensing is similar to strong lensing, but lenses are of lower mass such that the image separation is much smaller and typically not observable. We see from Eq. (2.83) that a sun-like star, as a lens, would produce on extragalactic scales image separations on the order of micro-arcseconds. Therefore, this is referred to as microlensing. On scales where sources are in our galaxy or the nearby neighborhood, the image separations are on the order of milli-arcseconds. Although multiple images are typically not resolved, microlensing is still observable due to additional magnification caused by the microlenses. This phenomenon was first observed in the QSO 2237+0305 (Huchra et al., 1985), a strongly lensed quasar with four multiple images, where uncorrelated brightness variations in the images could be assigned to a microlens with mass $\leq 0.1 M_{\odot}$ (Irwin et al., 1989).

Further, microlensing can be used to test if Massive Astrophysical Compact Halo Objects (MACHOs), like faint stars or black holes, could account for the missing matter (dark matter) in our Milky Way, as proposed by Paczynski (1986). For this purpose, the brightness fluctuations of millions of stars in the Magellanic Clouds have been observed over several years to search for characteristic time-variable magnifications due to microlensing events caused by MACHOs moving past our line of sight to the sources. However, results by Tisserand et al. (2007) showed that these events are too rare, and MACHOs in the mass range from $0.6 \times 10^{-7} M_{\odot}$ to $15 M_{\odot}$ can not be the primary occupants of the Milky Way Halo. Similar studies by Niikura et al. (2017) ruled out primordial black holes in the mass range from $10^{-14} M_{\odot}$ to $10^{-9} M_{\odot}$ as dark matter candidates by searching for microlensing signatures in tens of millions of stars in the Andromeda Galaxy.

Microlensing is not just a valuable probe to search for astrophysical dark matter. It is also used to detect extrasolar planets. If a foreground star passes very close to our line of sight to a more distant background star, the light of the background star will be temporarily magnified. Further, if the foreground star hosts a planet, this planet can act as an additional microlens, resulting in a characteristic short-lived signature (e.g., Bond et al., 2004; Nikolaus & Hundertmark, 2018).

As pointed out in Sect. 1.2.2, for the example of iPTF16geu, microlensing effects need to be accounted for in LSNe Ia studies (More et al., 2017; Yahalomi et al., 2017; Mörtsell et al., 2020), given that stars in the lensing galaxy can distort light curves from the multiple images differently and, therefore, add uncertainties to the time-delay measurement (Goldstein et al., 2018; Huber et al., 2019). More on microlensing of LSNe Ia will be discussed in Sect. 2.3.

2.2.6 Time-delay cosmography

In Sect. 1.2, we already introduced time-delay cosmography to measure H_0. The method is based on three pillars, a time-delay measurement of a strongly lensed variable source, a lens mass model, and a mass reconstruction along the line of sight. In the following, we will explain time-delay cosmography based on the lensing formalism introduced in this section. As shown in Sect. 2.2.3, the Fermat potential τ is proportional to the light travel time t, and the exact relation between two images, A and B, is (Cooke & Kantowski, 1975; Kayser &

Refsdal, 1983):

$$\Delta t = \frac{D_{\Delta \mathrm{t}}}{c}[\tau(\vec{\theta}_2,\beta) - \tau(\vec{\theta}_1,\beta)], \tag{2.84}$$

where the time-delay distance is

$$D_{\Delta \mathrm{t}} = \frac{D_{\mathrm{d}} D_{\mathrm{s}}}{D_{\mathrm{ds}}}(1 + z_{\mathrm{d}}). \tag{2.85}$$

Using the definition of the Fermat potential from Eq. (2.79), we get for the time-delay:

$$\Delta t = \frac{D_{\Delta \mathrm{t}}}{c}\left[\frac{1}{2}(\vec{\theta}_2 - \vec{\beta})^2 - \frac{1}{2}(\vec{\theta}_1 - \vec{\beta})^2 + \Psi(\vec{\theta}_1) - \Psi(\vec{\theta}_2)\right]. \tag{2.86}$$

The first delay $\frac{D_{\Delta \mathrm{t}}}{c}[\frac{1}{2}(\vec{\theta}_2 - \vec{\beta})^2 - \frac{1}{2}(\vec{\theta}_1 - \vec{\beta})^2)]$ is the "geometric delay" from the different path lengths of images A and B. The second delay $\frac{D_{\Delta \mathrm{t}}}{c}[\Psi(\vec{\theta}_1) - \Psi(\vec{\theta}_2)]$ is the "gravitational delay" or "Shapiro delay" (Shapiro, 1964) from the different gravitational time dilations for images A and B.

The image positions $\vec{\theta}_1$ and $\vec{\theta}_2$ are known from the observation. From the lens mass model with a reconstruction of mass perturbations along the line-of-sight, one gets from the multiple image positions the source position $\vec{\beta}$ and the lens potential $\Psi(\vec{\theta})$. Therefore by measuring the time delay between images A and B, which will be the main scope of this thesis, one determines the time-delay distance $D_{\Delta \mathrm{t}}$.

The time-delay distance contains combinations of the three angular diameter distances, as listed in Eq. (2.55) for any ΛCDM model and Eq. (2.56) for a flat ΛCDM cosmology. Suppose a large sample of measured time-delay distances is available. In that case, one can determine the cosmological parameters $(H_0, \Omega_{\mathrm{m},0}, \Omega_{\mathrm{r},0}, \Omega_{\Lambda,0})$, that describe our Universe's evolution as listed in Eq. (2.52). It is enough to determine three density parameters $\Omega_{i,0}$, given that from the relation in Eq. (2.50), the fourth one follows automatically via $\Omega_{k,0} = 1 - \Omega_{\mathrm{m},0} - \Omega_{\mathrm{r},0} - \Omega_{\Lambda,0}$. Further, for an open or closed Universe, it is not necessary to sample the current scale factor R_0, as listed in Eq. (2.55). For an open ($k = -1$, $\Omega_{k,0} > 0$) universe, the scale factor is set by $R_0 = \frac{c}{H_0 \sqrt{\Omega_{k,0}}}$, and for an closed Universe ($k = 1$, $\Omega_{k,0} < 0$), the relation is $R_0 = \frac{c}{H_0 \sqrt{-\Omega_{k,0}}}$ [see Eq. (2.51)].

Although the time-delay distance depends on four parameters, it is most sensitive to H_0. To illustrate this, we assume our current flat ΛCDM model with $H_0 \approx 70\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, $\Omega_{\mathrm{m},0} \approx 0.3$, $\Omega_{\Lambda,0} \approx 0.7$, and $\Omega_{\mathrm{r},0} \approx 0$. For a flat cosmological model, we have $D_{\Delta t} \propto H_0^{-1}$. However, $\Omega_{\mathrm{m},0}$ in the integrant of Eq. (2.56), is connected to the redshift, and therefore, we probe typical source redshifts and lens redshifts. For SNe Ia strongly lensed by galaxies we expect sources within $z_{\mathrm{s}} \in [0.17, 1.34]$ and lenses within $z_{\mathrm{d}} \in [0.026, 0.912]$ (Oguri & Marshall, 2010). For each pair of z_{s}, z_{d} ($z_{\mathrm{s}} > z_{\mathrm{d}}$, step size 0.02), we calculate the time-delay distance $D_{\Delta t}$ for different H_0 and $\Omega_{\mathrm{m},0}$, where $\Omega_{\mathrm{r},0} = 0$ and $\Omega_{\Lambda,0} = 1 - \Omega_{\mathrm{m},0}$. The result is shown in Fig. 2.5. The red line shows the case for $\Omega_{\mathrm{m},0} = 0.3$ and different H_0 values; in contrast the blue line represents $H_0 = 70\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ for different $\Omega_{\mathrm{m},0}$. Each value is normalized by the time-delay distance for $H_0 = 70\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ and $\Omega_{\mathrm{m},0} = 0.3$. The spread in the blue curve shows the contribution of the different z_{s} and z_{d} pairs, which cancels out for

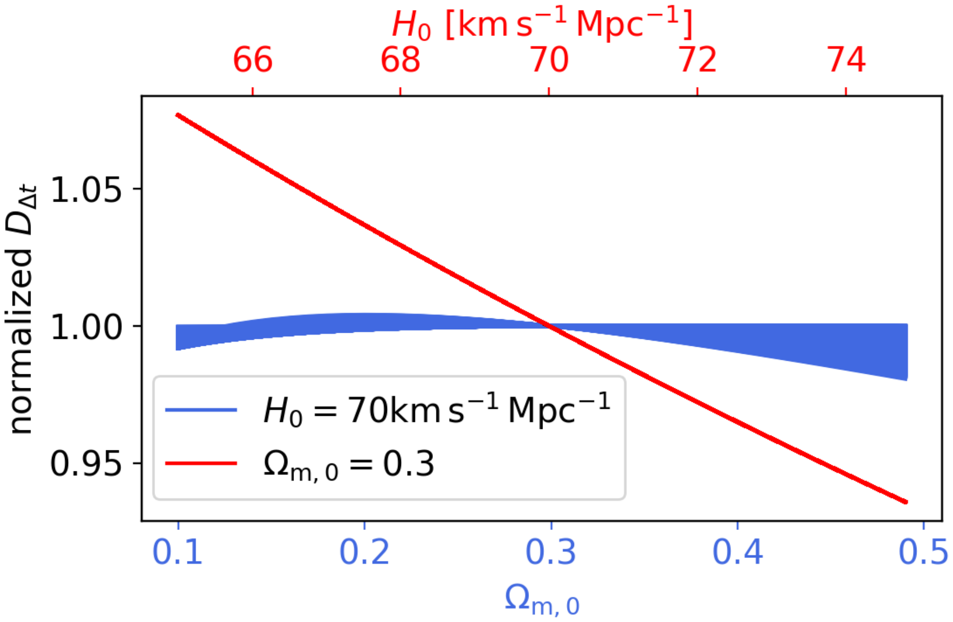


Fig. 2.5. Normalized time-delay distance $D_{\Delta t}$ in flat ΛCDM cosmology for different H_0 and $\Omega_{m,0}$ values ($\Omega_{\Lambda,0} = 1 - \Omega_{m,0}$, $\Omega_{r,0} = 0$). $D_{\Delta t}(z_s, z_d)$ is normalized by the corresponding $D_{\Delta t}(z_s, z_d)$ value from $H_0 = 70\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ and $\Omega_{m,0} = 0.3$. Further, we probe different pairs of z_s and z_d, where the spread is present in the case of varying $\Omega_{m,0}$ (blue). For the case of different H_0 values (red), the relation is $D_{\Delta t} \propto H_0^{-1}$ because the vertical offset cancels out due to normalization. We see that $\Omega_{m,0}$ influences the time-delay distance far less than the H_0 value. Therefore, time-delay cosmography is mostly sensitive to H_0 and only weakly dependent on other cosmological parameters.

the red line given that the curves are normalized and $D_{\Delta t} \propto H_0^{-1}$. We see that the time-delay distance $D_{\Delta t}$ is primarily sensitive to H_0 and depends only weakly on $\Omega_{m,0}$. Thus, time-delay cosmography is an appropriate way to determine H_0. Further, with a large sample of lensed transients, as will be provided by LSST, one could constrain other cosmological parameters.

2.2.7 Mass-sheet degeneracy

One issue of gravitational lensing is that a lens mass model fitting the observation well is not a unique model. Suppose one finds a $\kappa(\vec{\theta})$ that provides an excellent fit to the observation, meaning that image multiplicity, positions, and flux ratios are produced very well. Then all κ_λ fulfilling

$$\kappa_\lambda(\vec{\theta}) = \lambda + (1 - \lambda)\kappa(\vec{\theta}), \tag{2.87}$$

lead to the same observables. This is referred to as mass-sheet degeneracy (Falco et al., 1985; Schneider & Sluse, 2014). The first term corresponds to a constant dimensionless surface mass density $\kappa_c = \lambda$, and the second term is a rescaling of the dimensionless surface mass density $\kappa(\vec{\theta})$. We can show that, for κ_λ, the scaled source position

$$\vec{\beta}_\lambda = (1 - \lambda)\vec{\beta}, \tag{2.88}$$

will produce the same observables. Changes on the source plane will not be detectable. We are only interested in whether or not the image plane observables are affected by the mass-sheet transformation, as we can measure only these quantities. From Eqs. (2.67) and (2.87), we find that

$$\nabla_\theta^2 \Psi_\lambda(\vec{\theta}) = 2\kappa_\lambda(\vec{\theta}) = 2\lambda + 2(1-\lambda)\kappa(\vec{\theta}) = 2\lambda + (1-\lambda)\nabla_\theta^2 \Psi(\vec{\theta}).$$

If we take the antiderivative of this equation[16] and use Eq. (2.66) we get

$$\vec{\alpha}_\lambda(\vec{\theta}) = \nabla_\theta \Psi_\lambda(\vec{\theta}) = \lambda\vec{\theta} + (1-\lambda)\nabla_\theta \Psi(\vec{\theta}) = \lambda\vec{\theta} + (1-\lambda)\vec{\alpha}(\vec{\theta}), \tag{2.89}$$

where we neglect any integration constant $\vec{c}$, which would be compensated with an additional, not observable shift on the source plane $[\vec{\beta}_\lambda = \vec{\beta}(1-\lambda) - \vec{c}]$. Therefore, using the lens equation, we find that

$$\begin{aligned} \vec{\theta}_\lambda &= \vec{\beta}_\lambda + \vec{\alpha}_\lambda = (1-\lambda)\vec{\beta} + \lambda\vec{\theta} + (1-\lambda)\vec{\alpha}(\vec{\theta}) = (1-\lambda)[\vec{\beta} + \vec{\alpha}(\vec{\theta})] + \lambda\vec{\theta} = \\ &= (1-\lambda)\vec{\theta} + \lambda\vec{\theta} = \vec{\theta}. \end{aligned}$$

This means that the observable image positions do not change under the mass-sheet transformation defined in Eq. (2.87). Further, given that

$$\frac{\partial \beta_{\lambda,i}}{\partial \theta_j} = (1-\lambda)\frac{\partial \beta_i}{\partial \theta_j},$$

we find that the Jacobian matrix from Eq. (2.71) scales as $A_\lambda = (1-\lambda)A$. Therefore, we have $\det(A_\lambda) = (1-\lambda)^2 \det(A)$, which leads according to Eq. (2.78) to

$$\mu_\lambda = \frac{1}{(1-\lambda)^2}\mu. \tag{2.90}$$

Directly observable is the brightness ratio of two images, A and B, but this is not influenced by the mass-sheet transformation given that $\frac{\mu_{\lambda \mathrm{A}}}{\mu_{\lambda \mathrm{B}}} = \frac{\mu_\mathrm{A}}{\mu_\mathrm{B}}$. From the determinant of the Jacobian matrix and Eq. (2.78), we find the transformation of the shear:

$$|\gamma_\lambda| = (1-\lambda)|\gamma|. \tag{2.91}$$

Fig. 2.3 shows that we can measure the semi-major and semi-minor axes for a spatially resolved elliptical image. However, this depends on the assumption of a circular source with radius R, as well as on κ and γ, as shown in Eq. (2.72). Therefore, we can break the mass-sheet degeneracy if we know the intrinsic shape and size of the source. Further, from Eq. (2.89), we find that

$$\Psi_\lambda(\vec{\theta}) = \frac{\lambda}{2}\vec{\theta}^2 + (1-\lambda)\Psi(\vec{\theta}). \tag{2.92}$$

---

[16] $\nabla_\theta \cdot \vec{\theta} = 2$.

Thus, after a rather long calculation, one finds for the Fermat potential [Eq. (2.79)]:

$$\tau_\lambda(\vec{\theta},\vec{\beta}) = (1-\lambda)\tau(\vec{\theta},\vec{\beta}) - (1-\lambda)\frac{1}{2}\lambda\vec{\beta}^2. \tag{2.93}$$

If we had considered the integration constants, we would have more constant terms and terms depending on β. However, these terms cancel out in the difference and therefore:

$$\Delta\tau_\lambda = (1-\lambda)\Delta\tau. \tag{2.94}$$

From Eq. (2.84), we see that the time delay between two images is related to the difference in the Fermat potential via:

$$\Delta t = \frac{D_{\Delta t}}{c}\Delta\tau. \tag{2.95}$$

The time delay Δt will be directly measured from the observation, and $\Delta\tau$ will be determined from lens mass modeling, although, according to Eq. (2.94), it is not unique. If we assume a cosmological model with fixed (H_0, $\Omega_{\mathrm{m},0}$, $\Omega_{\mathrm{r},0}$, $\Omega_{\Lambda,0}$), the time-delay distance $D_{\Delta t}$ is determined for a given system and one would break the mass-sheet degeneracy. However, our goal is to measure especially H_0. Thus, the mass-sheet degeneracy transitions directly into the uncertainty of H_0 via

$$\delta H_0 \propto \frac{1}{\delta D_{\Delta t}} \propto (1-\lambda), \tag{2.96}$$

for a flat cosmological model. Therefore, it is essential to break the mass-sheet degeneracy. We mentioned already that this would be possible if we knew the intrinsic shape and size of the source, which is almost impossible in practice. Another way is, that we measure κ directly from stellar kinematics (e.g., Barnabè et al., 2011a; Yıldırım et al., 2017; Shajib et al., 2018; Yıldırım et al., 2020) at a position $\vec{\theta}$ where $\kappa(\vec{\theta}) \neq 1$, because $\kappa = 1 \Rightarrow \kappa_\lambda = 1$. Further, we can break the mass-sheet degeneracy if we know the absolute luminosity of the source [see Eq. (2.90)]. Therefore, if microlensing uncertainties are negligible, LSNe Ia are promising to break the mass sheet degeneracy given that SNe Ia are standardizable candles.

2.3 Simulated light curves of strongly lensed Type Ia supernovae

2.3.1 Type Ia supernovae

SNe are transient events originating from luminous explosions of stars where observables, like spectra or light curves, depend on the progenitor of the SN. Spectral features historically led to a classification scheme (e.g., Minkowski, 1941; Filippenko, 1997; Turatto, 2003), where type II SNe show hydrogen lines and the lack of those lines points towards type I SNe, as shown in Fig. 2.6. Further criteria on the presence of silicon and helium lines subdivide type I SNe in SNe Ia, Ib, and Ic. Today supernovae are mostly distinguished according to their progenitor. In a core-collapse SNe the released energy is the gravitational binding energy due to a collapse of a massive star, and SNe Ib, Ic, and II are of that type.

In contrast, SNe Ia are thermonuclear explosions, although the exact progenitor system

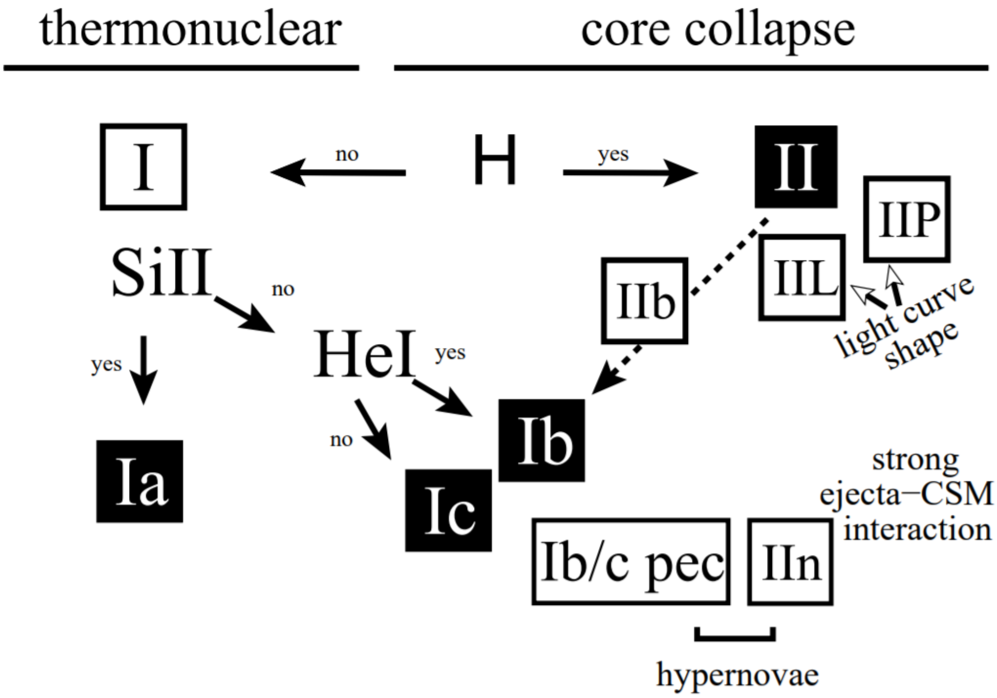


Fig. 2.6. Classification scheme of SNe depending on spectral features. Image taken from Turatto (2003).

is unknown. From observations, one finds that SNe Ia occur in young and old stellar populations, but so far, no compact remnant of a normal SN Ia has been detected. Further, a few pre-explosion images of nearby SNe Ia, such as SN 2011fe and SN 2014J, constrain the luminosity of the SN progenitor (Li et al., 2011; Kelly et al., 2014; Nielsen et al., 2014). These observational facts support theoretical predictions that SNe Ia have their origin in a thermonuclear explosion of a carbon-oxygen (CO) white dwarf (WD) in a binary system (e.g., Hoyle & Fowler, 1960; Nomoto et al., 1984). The classical textbook scenario is that a WD accretes material from a main-sequence star or a red giant until it reaches the Chandrasekhar mass limit of $\sim 1.4 M_{\odot}$ (Whelan & Iben, 1973; Nomoto, 1982). If a non-rotating WD approaches that limit, the WD becomes unstable, and a thermonuclear runaway occurs. However, several alternative scenarios have been suggested and are nowadays considered to be equally viable.

The observation of a SN is the radioactive afterglow triggered by the explosion. During the thermonuclear burning, carbon and oxygen are fused to mostly ^{56}Ni (Pankey, 1962; Truran et al., 1967; Colgate & McKee, 1969a; Kuchner et al., 1994). Then, the radioactive ^{56}Ni decays with a half-life time of 6.1 days to ^{56}Co via different decay channels, where γ-ray photons with a mean energy of 1.72 MeV are emitted. The consecutive decay of ^{56}Co to ^{56}Fe occurs at a half-life time of 77 days, where γ-ray photons with a mean energy of 3.59 MeV are emitted, coming from multiple decay channels (Colgate & McKee, 1969b). In addition to the γ-photons, these decay processes release positrons (e^{+}) and electron neutrinos (ν_{e}). While neutrinos can escape and form a potentially detectable signal (Kunugise & Iwamoto, 2007), the γ-ray photons and positrons deposit their energy in the SN ejecta. During the interaction of the radiation with the ejecta, the high energy photons are converted to wavelengths covering the ultraviolet to the infrared regime. This radiation can escape the expanding SN at some point and form characteristic observables like spectra and light curves.

2.3.1.1 Theoretical models and radiative transfer

To calculate the effect of microlensing on LSNe Ia, as motivated by Goldstein et al. (2018), we require theoretical models to infer the specific intensity because the spatial variation of the SN is needed. Different theoretical models are available, originating in the SNe Ia's unknown progenitor system. Therefore, in this thesis, we use four of these models to increase the diversity of progenitor scenarios and the spread in the predicted light curve shapes. The first model is the one-dimensional and spherically symmetric W7 explosion model (Nomoto et al., 1984). This model is not a first principle model, but still, it has been quite successful in reproducing key observables in SNe Ia (e.g., Jeffery et al., 1992; Gall et al., 2012). The second model is a detonation of a centrally ignited sub-Chandrasekhar (sub-Ch) mass WD with $1.06 M_{\odot}$ taken from Sim et al. (2010). This model belongs to a class where the explosion of the WD is triggered before reaching the Chandrasekhar mass limit of $\sim 1.4 M_{\odot}$. A double detonation can, for example, trigger such an explosion. In this scenario, a CO WD accretes a helium-rich outer shell from a companion star. The helium shell becomes unstable and detonates if it becomes sufficiently massive. This leads to a shock wave which may produce a secondary detonation in the CO core of the WD (Woosley & Weaver, 1986; Fink et al., 2007; Fink et al., 2010; Moll & Woosley, 2013). As a third model, we use a merger model of two CO WDs with 0.9 and 1.1 $M_{\odot}$ from Pakmor et al. (2012). In such a process, the secondary WD becomes unstable at some point and is tidally disrupted. The primary WD violently accretes the debris of the secondary which initiates a detonation. Our fourth model is the N100 model from Seitenzahl et al. (2013), which is a delayed detonation model of a CO WD around the Chandrasekhar mass limit. In this model, 100 randomly distributed ignition spots trigger a subsonic deflagration, which transitions into a detonation later. More about different explosion models and SNe Ia in general can be found in, e.g., Hillebrandt & Niemeyer (2000), Hillebrandt et al. (2013), and Livio & Mazzali (2018).

The radiative transfer equation, which describes how energy is transferred due to absorption, emission, and scattering, must be solved to produce synthetic observables for our four models. For this, the radiative transfer code `ARTIS` (Kromer & Sim, 2009), which stands for Applied Radiative Transfer In Supernovae, is used. This is a Monte Carlo-based approach, following methods developed by Lucy (1999, 2002, 2003, 2005). Here, monochromatic γ-ray packets are produced following the different decay channels of ^{56}Ni to ^{56}Co and ^{56}Co to ^{56}Fe. These γ-ray packets are then converted to ultraviolet-optical-infrared radiation (UVOIR) packets due to interaction with the ejecta via Compton scattering, absorption, or pair production. The full Monte Carlo radiative transport procedure is then used to propagate the UVOIR packets through the ejecta. Here free-free interactions (free electrons emit or absorb photons), bound-free interactions (photon emission or absorption due to electron

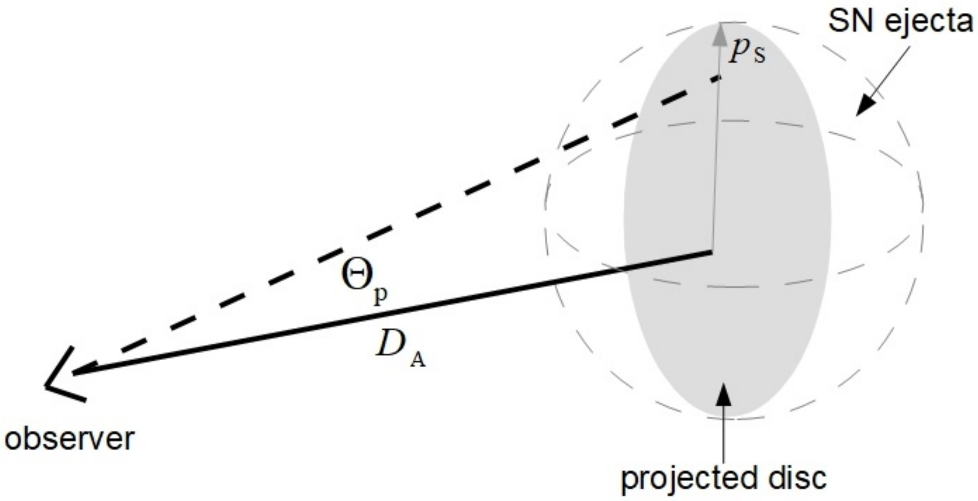


Fig. 2.7. Projected SN onto a disk perpendicular to the line of sight to the observer, assuming spherical symmetry. The specific intensity is represented by the gray disc, which is a function of time, wavelength, and impact parameter p, where $p = 0$ corresponds to the center of the disc and p_S is the radius of the disc. The SN is at an angular diameter distance of D_A. Image taken from Huber et al. (2019).

transitions between bound states and ionization), and especially bound-bound interactions (photon emission or absorption due to electron transition between two bound states in an atom) are considered. Once a photon packet escapes from the simulation box surrounding the SN ejecta, the position, time, energy, wavelength, and propagation direction are stored. For more details, see Kromer & Sim (2009). These packets are typically binned in wavelength and time to produce spectra and light curves. For our case, the spatial information of the projected SN emission, perpendicular to the line-of-sight to the observer, is essential, as illustrated in Fig. 2.7, to calculate microlensing effects. The projection is described in detail by Huber et al. (2019), where spherical symmetry of the SN ejecta is assumed. The result is the emitted specific intensity $I_{\lambda,e}(t,p)$ as a function of wavelength λ, time after explosion t, and impact parameter p, i.e., the projected distance from the ejecta center, as shown in Fig. 2.7. The assumption on spherical symmetry is exact for the one-dimensional W7 model and a good approximation for the N100 and sub-Ch model, which produce nearly spherically symmetric ejecta (Sim et al., 2010; Seitenzahl et al., 2013; Sim et al., 2013). However, merger models (Pakmor et al., 2010, 2011, 2012, 2013) have strongly asymmetric ejecta, and the approximation is less accurate. We will investigate the effects of asymmetries in Sect. 3.4.3.

2.3.1.2 Synthetic observables

The calculation of the observed radiation flux from the specific intensity is given from Eq. (2.76), and we get for our case

$$F_{\lambda,o}(t) = \int_0^{2\pi} \mathrm{d}\phi \int_0^{\theta_S} \mathrm{d}\theta\, \theta \cos\theta\, I_{\lambda,o}(t,p), \tag{2.97}$$

where $\theta_{\rm S}$ is the angular radius of the SN ejecta projected onto a disc. Further, we assume parallel light rays and small angles meaning that $\cos\theta \approx 1$ and $\theta \approx \tan\theta = p/D_{\rm A}$, which is justified by the large distance $D_{\rm A}$ in comparison to the source radius $p_{\rm S}$. Therefore we get

$$F_{\lambda,\rm o}(t) = \frac{2\pi}{D_{\rm A}^2} \int_0^{p_{\rm S}} {\rm d}p\, p\, I_{\lambda,\rm o}(t,p). \tag{2.98}$$

Now we can relate the observed specific intensity $I_{\lambda,\rm o}$ to the emitted one $I_{\lambda,\rm e}$ using Liouville's theorem which states that I_ν/ν^3 is a conserved quantity in free space (Mihalas & Mihalas, 1984, p. 414). Here the specific intensity is defined via the frequency ν and $I_\nu \propto {\rm d}\nu^{-1}$. From $\nu = c\lambda^{-1}$ follows $\frac{{\rm d}\nu}{{\rm d}\lambda} = -c\lambda^{-2}$ and, therefore, $I_\lambda = I_\nu {\rm d}\nu {\rm d}\lambda^{-1} = I_\nu c \lambda^{-2}$, where we dropped the minus sign because I_λ and I_ν are per definition positive. From this, we learn that $I_\lambda \lambda^5$ is conserved and, therefore

$$I_{\lambda,\rm o}(t,p) = I_{\lambda,\rm e}(t,p)\frac{\lambda_{\rm e}^5}{\lambda_{\rm o}^5} = \frac{I_{\lambda,\rm e}(t,p)}{(1+z)^5}, \tag{2.99}$$

where we used Eq. (2.28) in the last step. Therefore, we get the radiation flux

$$F_{\lambda,\rm o}(t) = \frac{2\pi}{D_{\rm A}^2(1+z)^5} \int_0^{p_{\rm S}} {\rm d}p\, p\, I_{\lambda,\rm e}(t,p) = \frac{2\pi}{D_{\rm lum}^2(1+z)} \int_0^{p_{\rm S}} {\rm d}p\, p\, I_{\lambda,\rm e}(t,p). \tag{2.100}$$

From the radiation flux, we can calculate the AB-magnitudes in the filter X by multiplying it with the transmission function $S_{\rm X}(\lambda)$ and integrating over the band with[17]

$$m_{\rm AB,X}(t) = -2.5\log_{10}\left(\frac{\int {\rm d}\lambda\, \lambda S_{\rm X}(\lambda)\, F_{\lambda,\rm o}(t)}{\int {\rm d}\lambda\, S_{\rm X}(\lambda)\, c\lambda^{-1}} \times {\rm cm}^2\, {\rm erg}^{-1}\right) - 48.6 \tag{2.101}$$

(Bessell & Murphy, 2012). Fig. 2.8 shows the transmission function $S_{\rm X}(\lambda)$ for the nine filters / bands investigated in this thesis. Most relevant are the six LSST filters covering the optical to near-infrared. A zoomed-in version of the six filters is shown in Fig. 3.1.

From Eq. (2.101), we can calculate the light curves for all filter bands (*ugrizyJHK*) from the four theoretical SN Ia models. This is shown in Fig. 2.9 for rest-frame light curves, meaning that $z = 0$, $D_{\rm lum} = D_{\rm A} = 10\,{\rm pc}$. The four models show substantial offsets, although they cover the same trend. In Fig. 4.1 and 5.4 we also see a comparison to empirical SN Ia light curves, which show quite some variations as well. Overall, the theoretical models' variations encapsulate the empirical models' variations. We will see in Secs. 4 and 5 that the variations in the theoretical models are helpful to cover intrinsic variations of real SNe Ia when we train our machine learning models to measure time delays in LSNe Ia.

From Eq. (2.101), we can define color curves, which will be especially relevant for Sect. 3. A color curve is defined as the difference between two bands X and Y:

$$x - y := m_{\rm AB,X}(t) - m_{\rm AB,Y}(t). \tag{2.102}$$

---

[17] $1\,{\rm erg} = 10^{-7}\,{\rm J}$

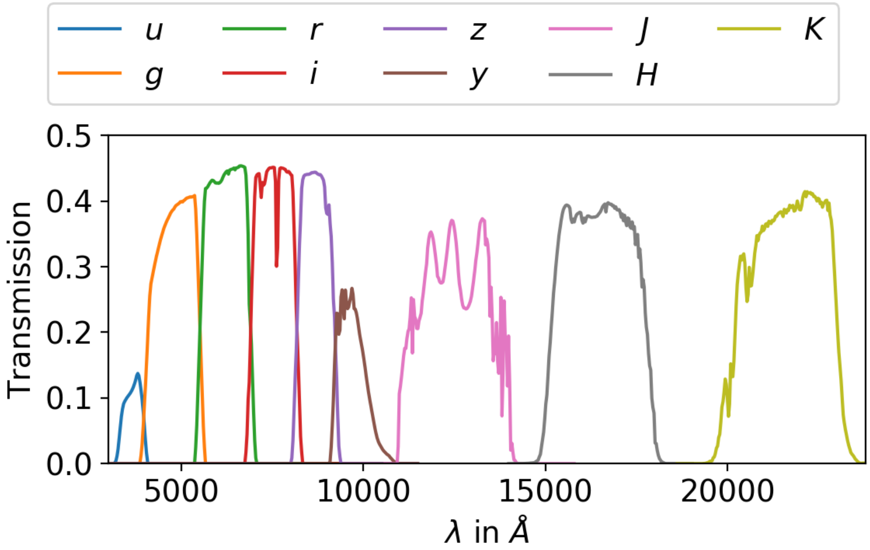


Fig. 2.8. Transmission functions $S_X(\lambda)$ for nine filter bands used in this thesis. The effective wavelengths are $\lambda_{\mathrm{eff},u} = 3671\,\text{Å}$, $\lambda_{\mathrm{eff},g} = 4827\,\text{Å}$, $\lambda_{\mathrm{eff},r} = 6223\,\text{Å}$, $\lambda_{\mathrm{eff},i} = 7546\,\text{Å}$, $\lambda_{\mathrm{eff},z} = 8691\,\text{Å}$, $\lambda_{\mathrm{eff},y} = 9712\,\text{Å}$, $\lambda_{\mathrm{eff},J} = 12\,513\,\text{Å}$, $\lambda_{\mathrm{eff},H} = 16\,467\,\text{Å}$, and $\lambda_{\mathrm{eff},K} = 21\,641\,\text{Å}$. A zoomed-in version showing just the LSST bands is shown in Fig. 3.1.

2.3.2 Microlensing on SNe Ia

We explained the basic concepts of microlensing already in Sect. 2.2.5.2. In Sect. 1.2.2.1, we showed, for the example of iPTF16geu, that microlensing might explain observed brightness fluctuations. The calculation of microlensing on LSNe Ia with specific intensity profiles was first implemented by Goldstein et al. (2018) using the W7 model, where synthetic observables have been calculated with the radiative transfer code `SEDONA` (Kasen et al., 2006). As mentioned already in Sect. 2.3.1, we use in this thesis four theoretical models where synthetic observables are calculated with `ARTIS`, where the W7 model was already used in our previous work (Huber et al., 2019).

Since lensing conserves surface brightness, adding a magnification factor μ to our equations introduced in Sect. 2.3.1 is straightforward, by substituting $I_{\lambda,\mathrm{e}}(t,p) \rightarrow \mu I_{\lambda,\mathrm{e}}(t,p)$. The magnification factor μ can come from macrolensing only, as listed in Eq. (2.78), for a strong lens like a galaxy, or it can be a combination of macrolensing and microlensing, taking the microlenses, like stars within a foreground lens galaxy, into account. In the case where microlensing is accounted for, the magnification is generally a function of cartesian coordinates (x,y). The factor $\mu(x,y)$ yields the magnification coming from stars of the lensing galaxy projected to the source plane. By taking $\mu(x,y)$ into account, we get the radiation flux:

$$
\begin{aligned}
F_{\lambda,\mathrm{o}}(t) &= \frac{1}{D_\mathrm{A}^2 (1+z)^5} \int \mathrm{d}x \int \mathrm{d}y\, I_{\lambda,\mathrm{e}}(t,x,y)\mu(x,y) = \\
&= \frac{1}{D_\mathrm{lum}^2 (1+z)} \int \mathrm{d}x \int \mathrm{d}y\, I_{\lambda,\mathrm{e}}(t,x,y)\mu(x,y).
\end{aligned} \tag{2.103}
$$

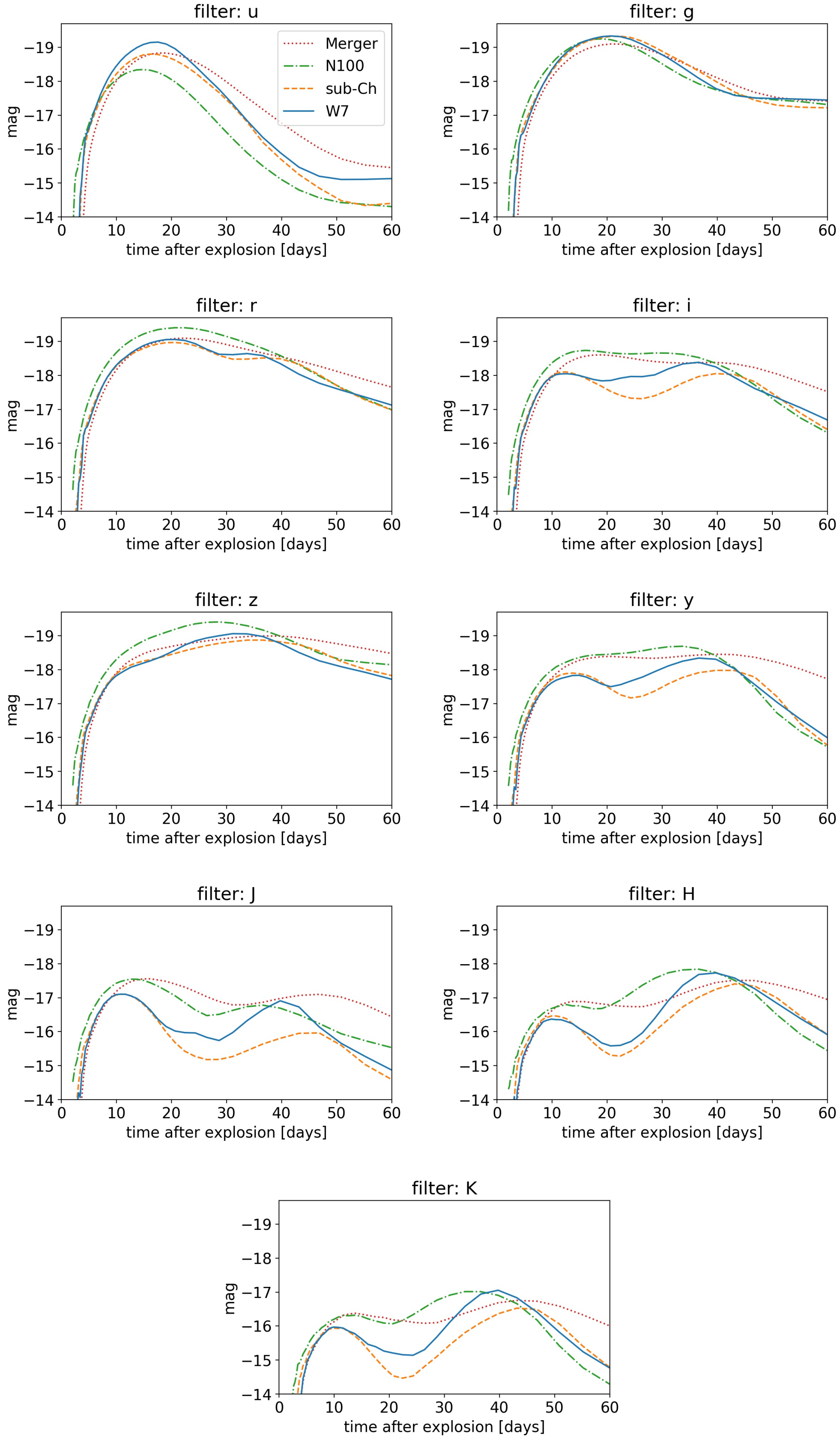


Fig. 2.9. Rest-frame light curves for nine different filter bands and four theoretical models. The various theoretical models show quite some variation but overall have the same trend. In Figs. 4.1 and 5.4, we see a comparison of theoretical models to empirical models, which also show substantial variations but are encapsulated by the spread of the theoretical models mostly.

Here we replaced $2\pi \int_0^{p_S} \mathrm{d}p\, p \rightarrow \int \mathrm{d}x \int \mathrm{d}y$ to switch from polar to cartesian coordinates. Therefore, also $I_{\lambda,\mathrm{e}}(t,p)$ is interpolated to the cartesian grid, which we describe in detail in Huber et al. (2019).

The magnification maps we use in this thesis are based on `GERLUMPH` (Vernardos & Fluke, 2014; Vernardos et al., 2014, 2015). The main parameters of the microlensing map are the convergence κ and the shear γ. These two parameters are determined by the macro modeling of the strongly lensed system. In addition to κ and γ, also the smooth matter fraction s is important, which is defined as

$$s = \frac{\kappa_\mathrm{s}}{\kappa} = 1 - \frac{\kappa_*}{\kappa}. \tag{2.104}$$

Here, κ_s is the smooth matter convergence, κ_* the convergence of the stars (microlenses), and $\kappa = \kappa_\mathrm{s} + \kappa_*$ the total convergence. For the case where only smooth matter is present ($\kappa_\mathrm{s} = \kappa$), the whole map would be smooth and represent the macro magnification. Each image of a strongly lensed system is represented by a different microlensing map, depending on the values of κ, γ, and s of the image position. For the creation of a magnification map, hundreds of stars are distributed randomly with a mean mass of $\langle M \rangle$ following a initial mass function, such that the mean of the magnification map would represent the macrolensing magnification set by κ and γ. The exact number of microlenses depends, among others, on κ_* and $\langle M \rangle$ (e.g., Vernardos & Fluke, 2013). Methods to determine the microlensing effect of hundreds of stars are often based on the inverse ray-shooting technique (e.g., Kayser et al., 1986; Schneider, 1987; Wambsganss et al., 1992; Vernardos & Fluke, 2013). This approach works by propagating light rays backward from the observer through the lens plane to the source plane. At the lens plane, the rays are deflected by finding approximate solutions of the lens equation. The source plane is then divided into a grid of pixels to count the number of rays N_micro that reach each pixel. Comparing this to the number of rays N that would reach the pixel if no lensing happens yields the value of the magnification factor for each pixel (N_micro/N). Another important quantity is the Einstein Radius on the source scale, which determines the scale of the microlensing map. Therefore, by multiplying the angular Einstein radius for a point mass lens [see Eq. (2.82)] with D_s and replacing M by the mean mass of the microlenses $\langle M \rangle$, we get

$$R_\mathrm{Ein} = \sqrt{\frac{4G\langle M \rangle}{c^2} \frac{D_\mathrm{s} D_\mathrm{ds}}{D_\mathrm{d}}}. \tag{2.105}$$

Therefore, also the source redshift z_s and z_d are parameters relevant to the microlensing effect. A set of different microlensing maps will be shown in Fig. 3.12.

Fig. 2.10 shows a SN Ia placed in a magnification map for 14.9 and 39.8 rest-frame days after explosion assuming the W7 model. The displayed size of the SN in cyan in the first and second row is defined such that the area contains 99.9 % of the total specific intensity $\int \mathrm{d}\lambda \int \mathrm{d}x \int \mathrm{d}y\, I_{\lambda,\mathrm{e}}(t,x,y)$. The second row shows the normalized specific intensity profiles in the six LSST bands. The dashed black line marks the radial coordinate in the first row, separating low and high-magnification regions. The third row shows the spectrum from the corresponding positions in the first row (solid red line), compared to the case of no microlensing where $\mu(x, y) = 1$ (dashed black line). The spectra are calculated using Eq. (2.103). The fourth row shows the wavelength-dependent magnification factors. In the case of 14.9 days after the explosion, the SN is in a rather homogeneous region of demagnification. Therefore the spectrum appears fainter due to lensing, but it is not distorted. For the case of 39.8 days after the explosion, the SN Ia has expanded and crossed a caustic, separating low and high magnification regions. Given that the specific intensity profiles highly depend on the wavelength, as shown in panel f), regions from the u, g, and z bands are more magnified than the other bands leading to a highly distorted spectrum due to microlensing. The wavelength-dependent magnification, even though lensing itself is achromatic, was also mentioned in Sect. 2.2.2.

In Fig. 2.11, we show an example of the impact of microlensing on light curves. The presented curves are calculated using Eqs. (2.103) and (2.101). Both light curves (g and z band) show significant distortion due to microlensing. Given that each image of a LSNe Ia will have a random position in its own microlensing map, the microlensing effect on each image will be different. Therefore, the microlensing will add large uncertainties to the time-delay measurement in the case presented here. The cases in Fig. 2.10 and 2.11 are examples of significant microlensing, not ideal for breaking the mass-sheet degeneracy via SNe Ia as standardizable candles.

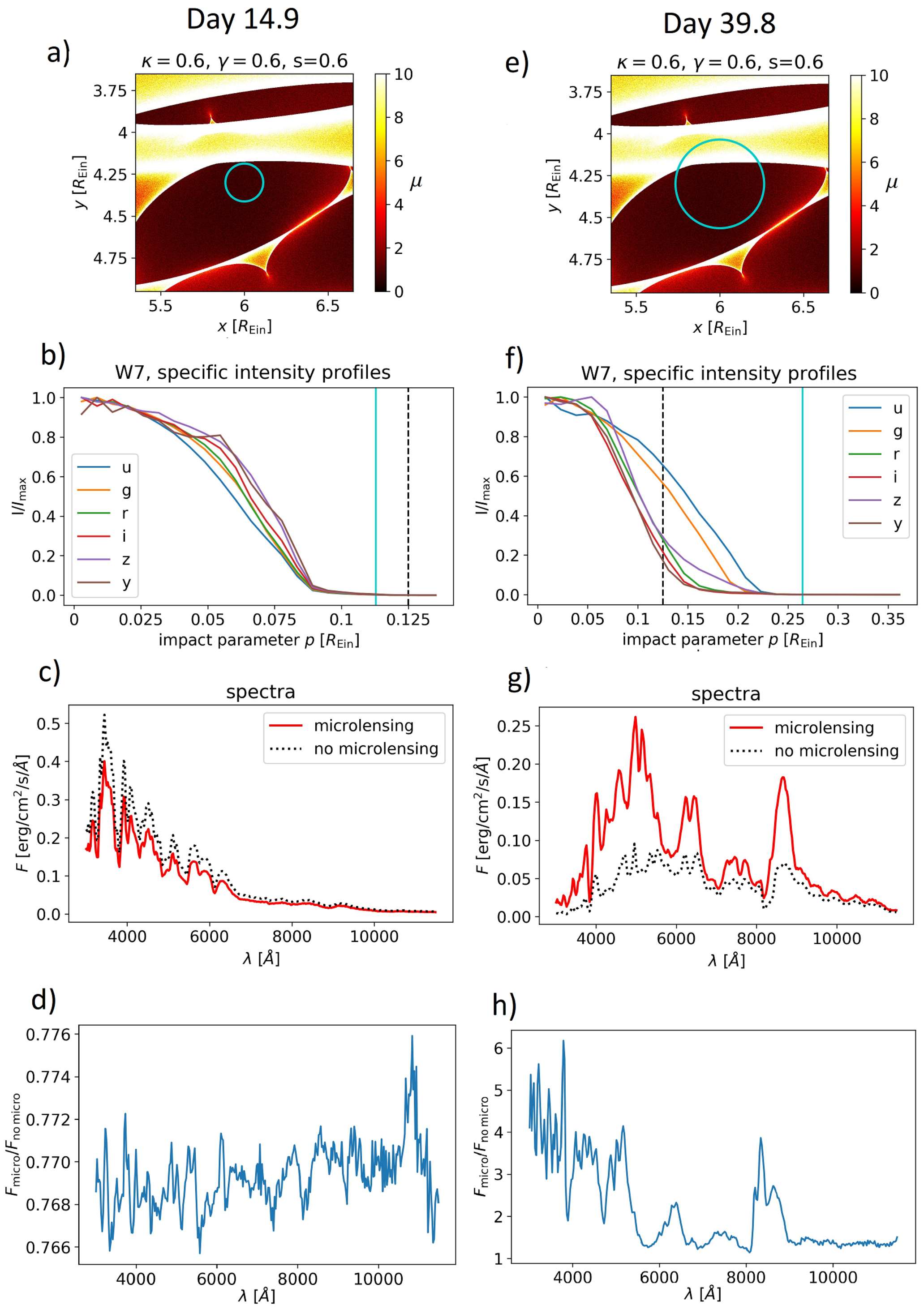


Fig. 2.10. Microlensing on rest-frame spectra of LSNe Ia for 14.9 and 39.8 rest-frame days after explosion assuming the W7 model. The first row shows a SN Ia in a magnification map, where the Einstein Radius $R_{\rm Ein} = 7.2 \times 10^{-3}$ pc $= 2.2 \times 10^{16}$ cm sets the physical scale. The second row shows the normalized specific intensity profiles as a function of the radial impact parameter. The third row shows the corresponding spectra with $\mu(x,y)$ as displayed in the magnification map (solid red line) in comparison to the case of $\mu(x,y) = 1$ (dashed black line). The fourth row shows the wavelength-dependent magnification caused by the specific intensity profiles in combination with the spatial-dependent magnification factor, although lensing itself is achromatic. The case in the first column shows only little wavelength-dependent magnification in contrast to the second column, where the spectrum is highly distorted. Image taken from Huber et al. (2019).

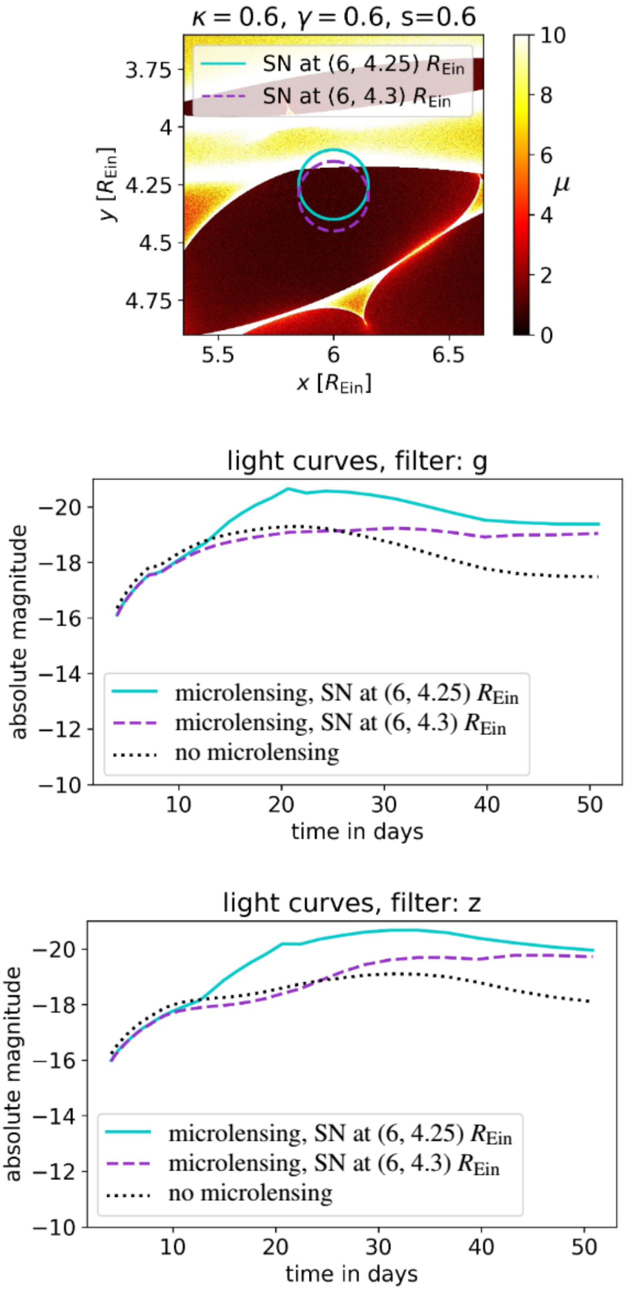


Fig. 2.11. Two microlensed rest-frame LSN Ia light curves (g and z band) based on the W7 model for two different positions (solid cyan and dashed magenta) in a magnification map ($R_{\mathrm{Ein}} = 7.2 \times 10^{-3}\,\mathrm{pc}$). We see that microlensing can highly distort light curves compared to the case of no microlensing (dotted black). Each image of a LSNe Ia will have a random position within its own magnification map, so each image's distortion will be different; therefore, microlensing can add large uncertainties to the time delay measurement. Image taken from Huber et al. (2019) with minor modifications.

2.3.3 Mock LSNe Ia

A mock lens catalog is required to simulate realistic LSNe Ia expected with an upcoming survey like LSST. For this purpose, we use the Oguri & Marshall (2010, hereafter OM10) lens catalog, which contains mock quasars and supernovae strongly lensed by galaxies. The catalog is ten times oversampled and contains 417 LSNe Ia, meaning that the prediction for the 10-year LSST survey is on the order of 40 LSNe Ia. This number depends on the assumptions that LSST will cover 20000deg^2 with a season length of three months over ten years. Another criterion for the catalog is that the multiple images are resolved, meaning that the image separation is larger than $0.5''$. Further, the 10σ point source limiting magnitude is 23.3 for a single visit in the i band. A system counts as detected if, for a double, the peak of the fainter image is 0.7 mag brighter than the limiting magnitude in the i band and falls in an observing season. The same criterium applies for detecting quads, but here the reference is the third brightest image. The estimates in the OM10 catalog are rather conservative, and the actual number should exceed the prediction, as discussed in Sect. 1.2.2.2. However, the systems listed in OM10 are the most relevant ones for our science case because unresolved systems with small image separation will provide short time delays Δt that are not ideal for time-delay cosmography. The catalog assumes as a lens mass model the singular isothermal ellipsoid (Kormann et al., 1994) and provides for all LSNe Ia systems the time-delays, redshifts (z_s and z_d), and (κ, γ) parameters. Further, we probed in Chapters 3, 4, and 5 different s values and assumed a Salpeter initial mass function (IMF) with $\langle M \rangle = 0.35 M_\odot$, where the assumed IMF has negligible impact on our studies (Chan et al., 2021). We use these parameters to calculate the corresponding microlensing maps, as described in Sect. 2.3.2, and determine 10000 microlensed light curves per map following Eqs. (2.101) and (2.103). We also include observational uncertainties via

$$m_\mathrm{X}(t) = m_\mathrm{AB,X}(t) + r_\mathrm{norm}\sigma_\mathrm{X}(t), \tag{2.106}$$

where r_norm is a random number following a Gaussian distribution with zero mean and a standard deviation of one. The 1σ standard deviation of $m_\mathrm{X}(t)$ is $\sigma_\mathrm{X}(t)$, which depends mainly on $m_\mathrm{AB,X}(t)$ relative to the 5σ depth of the filter X [more on that in LSST Science Collaboration (2009), Sect. 3.5, p.67, and Sect. 4.10.1].

2.4 Machine learning

Machine learning, especially deep learning, got incorporated into many different applications and research fields in the last few years. These techniques are used, for example, in spam detection (e.g., Crawford et al., 2015), video recommendation (Deldjoo et al., 2016), and cancer diagnosis (e.g, Benhammou et al., 2020; Wulczyn et al., 2020), but they also find more applications in the field of gravitational lensing, for example, in the search for lenses (e.g., Petrillo et al., 2019; Jacobs et al., 2019; Jacobs et al., 2019; Cañameras et al., 2020; Canameras et al., 2021; Shu et al., 2022) and lens modeling (e.g., Schuldt et al., 2022a,b). Machine learning summarizes algorithms that can learn independent tasks they were not explicitly designed to learn. Deep learning is a part of machine learning where multilayered neural networks learn from a huge data set to extract certain features in the data set. This type of architecture is initially inspired by the human brain, where neurons are the fundamental units responsible for converting input from the environment to thoughts and reactions.

Machine learning is divided into three main categories; supervised, unsupervised, and partially supervised learning. In the case of supervised learning, the training data is labeled, meaning that the labeled input data is directly related to a labeled ground truth output. While the input data is processed through a machine learning model, the predicted outcomes are compared to the ground truth values. The model is optimized so that the predictions match the corresponding ground truths. There are two sub-types of this approach, classification, and regression. In the case of classification, the outputs are discrete or categorical, and for regression, the outcomes are continuous. The main challenges in this approach are to generalize to test data not used in the training process and to assemble large-enough training sets. In the case of unsupervised learning, the training data is not labeled, and the machine learning models learn significant features to identify relationships in the training data, such as clustering (e.g., Saeed et al., 2020). In the case of partially supervised learning, parts of the data are labeled, and others are not. The advantage is that one can minimize the amount of labeled data. However, on the other side, certain not relevant input features could lead to wrong decisions (Chapelle et al., 2006). A detailed review of machine learning and deep learning can be found, for example, in Alzubaidi et al. (2021).

In this thesis, we use supervised learning on a regression problem. For this, we apply two deep learning methods, a feed-forward fully connected neural network and a long short-term memory network (Hochreiter & Schmidhuber, 1997; Sak et al., 2014; Sherstinsky, 2020, LSTM;). In addition, we will use a random forest (RF; Breiman, 2001), which is a machine-learning method built on many random decision trees. However, the underlying problem is the same for all three approaches. We have a given input X, also referred to as one single sample, for which we like to produce an output y:

$$X = (x_1, x_2, \dots x_N) \quad \rightarrow \quad \text{machine learning model} \quad \rightarrow \quad y.$$

In our case, X contains the light curves of a LSN Ia, where each element x_i represents a single observed data point at time t_i for a given image of the LSN Ia. The network output is the time delay Δt. If X contains more than two images of a LSN Ia, the output will provide multiple time delays. However, in most cases, X will contain the data of two images, and therefore, the outcome will be a single time delay. Further, we require multiple LSN Ia systems to train a machine learning model, and thus, we have several thousand to millions of samples like X. Given that almost no observational data for LSNe Ia is available, we have to simulate LSNe Ia following Sect. 2.3, and we will present the whole process as well as the machine learning methods in detail in Sects. 4 and 5.

In the following, we explain some basic concepts of machine learning using a simple fully connected neural network (FCNN). We assume that our data set contains samples,

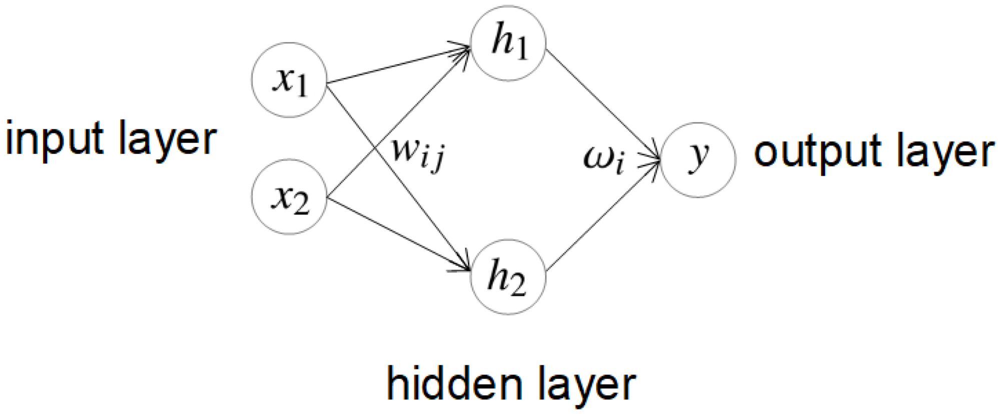


Fig. 2.12. Simple fully connected neural network with one hidden layer.

where each sample has two input values $X = (x_1, x_2)$ and an corresponding ground truth value y_{true}. Our network, shown in Fig. 2.12, has an input and hidden layer with two nodes, and an output layer with one node. These nodes are also referred to as neurons. Of course, such a network can only solve straightforward tasks, and it is one of the challenges to build a machine learning environment complex enough to address the problem. Nevertheless, the nodes from the different layers are connected by weights w (connecting input and hidden layer) and ω (connecting hidden and output layer), which have to be learned during the training process starting from some initial values. The nodes of the hidden layers are calculated via[18]

$$h_1 = A(w_{11}x_1 + w_{12}x_2) \qquad \text{and} \qquad h_2 = A(w_{21}x_1 + x_{22}x_2). \tag{2.107}$$

From this one can calculate the output of the FCNN as

$$y = \tilde{A}(\omega_1 h_1 + \omega_2 h_2). \tag{2.108}$$

The functions A and $\tilde{A}$ are activation functions which can be used to introduce nonlinearities to the network (e.g., Glorot et al., 2010; Maas et al., 2013; Nwankpa et al., 2018). A few activation functions, like rectified linear units (ReLU), sigmoid, and tanh are shown in Fig. 2.13.

Now we process N training samples through the network, where we denote a single sample with l, such that our inputs are $X_l = (x_{1,l}, x_{2,l})$. Typically N is a small subset of the whole training data. However, using X_l, we get a predicted value y_l for each sample l. To compare this to the ground through value $y_{\text{true},l}$ one calculates a loss L_l. The form of the loss function depends on the specific problem to solve, but a straightforward one is the mean squared error loss which is suitable for our example:

$$L = \frac{1}{N}\sum_{l=1}^{N} L_l = \frac{1}{N}\sum_{l=1}^{N} (y_l - y_{\text{true},l})^2 \tag{2.109}$$

---

[18] We did not introduce any bias term here. For example, the equation for h_1 with bias would be $h_1 = A(w_{11}x_1 + w_{12}x_2 + b_1)$. These biases are learned during the training process, similar to the weights.

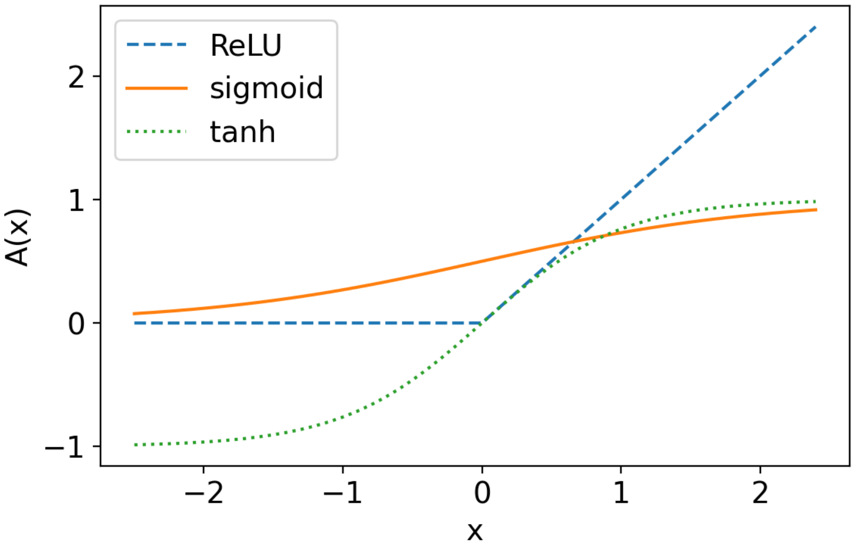


Fig. 2.13. A few activation functions as used in deep learning to introduce nonlinearities. The argument x of the activation function $A(x)$ is a linear combination of weights and inputs, respectively hidden states, as listed in Eqs. (2.107) and (2.108)

As a next step, one uses backpropagation on the loss L to optimize the weights. This means that one calculates the partial derivatives of the loss with respect to the weights using the chain rule:

$$\frac{\partial L_l}{\partial \omega_i} = \frac{\partial L_l}{\partial y_l}\frac{\partial y_l}{\partial \omega_i} \qquad \text{and} \qquad \frac{\partial L_l}{\partial w_{ij}} = \frac{\partial L_l}{\partial y_l}\frac{\partial y_l}{\partial h_i}\frac{\partial h_i}{\partial w_{ij}}. \tag{2.110}$$

The partial derivatives are then evaluated for each sample l at the corresponding values (y_l, X_l), and are used to update the weights:

$$\omega_i' = \omega_i - \alpha \frac{1}{N}\sum_{l=0}^{N} \frac{\partial L_l}{\partial \omega_i}(y_l, X_l) \qquad \text{and} \qquad w_{ij}' = w_{ij} - \alpha \frac{1}{N}\sum_{l=0}^{N} \frac{\partial L_l}{\partial w_{ij}}(y_l, X_l), \tag{2.111}$$

where α is the learning rate, which has to be adjusted to the specific problem. A high learning rate can lead to diverging weights, whereas the network will take too long to fit the training data properly for a low learning rate. The optimization method presented in Eq. (2.111) is called gradient descent. However, in practice, one often uses improved optimizers for faster convergence (e.g., Sun et al., 2019). Nevertheless, after the weights are updated, the process continues (using the new weights) with the next N samples of the training set until the whole training data is processed through the network, which marks the end of a training epoch. After a training epoch is finished, one continues with the next epoch, and if the network is set up properly, the training loss will decrease over the training process. However, to avoid overfitting the training data, one uses, in addition, a validation data set that is drawn from the same distribution as the training data, but not used to optimize the weights. For the validation loss one calculates after each epoch the loss of the validation set. After a sufficient number of training epochs, one will find a minimum in the validation loss while the training loss is still decreasing. This is the so-called generalization gap, and the final network for applications will be stored at that stage. In Fig. 5.8, we show such a training process, where we reach the generalization gap after 7313 training epochs.

The FCNN, as used to introduce basic deep learning concepts, will be applied in Chapter 4, and Chapter 5 in combination with an LSTM network which will be introduced in detail in Sect. 5.2.1. In addition to these two deep learning approaches, we will investigate the RF as a third machine learning method in Sect. 4. However, before we develop machine learning models to measure time delays in LSNe Ia, we first investigate qualitatively color curves in Chapter 3.

3 Achromatic phase of strongly lensed Type Ia supernovae

To determine the Hubble constant from LSNe Ia a time-delay measurement is required. In principle, this can be achieved by matching clear features in time coming from spectra, light curves, or color curves of the multiple images of a LSNe Ia. However, microlensing can distort spectra and light curves and add significant uncertainties to the time-delay measurement. Therefore, color curves seem to be very promising in reducing microlensing uncertainties. If two light curves from different filters are microlensed similarly, the effect will cancel out in the corresponding color curve. This was indeed found by Goldstein et al. (2018) for the spherically symmetric W7 model (Nomoto et al., 1984), where synthetic observables have been calculated with the radiative transfer code `SEDONA` (Kasen et al., 2006). The results of this study show that the microlensing effect in color curves is almost negligible up to three rest-frame weeks after the explosion, which is referred to as the achromatic phase.

The question remained if the achromatic phase is only present in the W7 model or if it can be confirmed in other SN Ia models. Therefore, we performed a detailed analysis on four theoretical models, namely W7, N100 (Seitenzahl et al., 2013), a sub-Ch (Sim et al., 2010), and a merger model (Pakmor et al., 2012), where we also tested the effects of asymmetries in the SN ejecta. To calculate synthetic observables, we used the radiative transfer code `ARTIS` (Kromer & Sim, 2009). Microlensed light curves are calculated as described in Sect. 2.3.2 based on Huber et al. (2019). The color curves follow from Eq. (2.102). We investigate the microlensing effect on color curves for the four theoretical models by probing typical values of κ, γ, z_s, and z_d following the OM10 catalog and a broad range of potential s values.

Our findings support the results from Goldstein et al. (2018), as we find, on average, an achromatic phase around three rest-frame weeks or longer for all investigated SN Ia models. Furthermore, asymmetries in the supernova ejecta do not have a significant impact. However, we find that the duration of the achromatic phase varies substantially for different color curves and microlensing magnification maps. Overall the achromatic phase is longer for higher s values and lower macro-magnification factors. The trend that larger R_ein provide a longer achromatic phase is not significant.

The main issue we have to report is that strong features (clear extreme points or turning points) are only present in three out of five independent rest-frame LSST color curves containing combinations from *ugri* bands. This would require follow-up observations for typical LSN Ia redshifts (0.5 to 1.0) in *rizyJH* bands. Further, we do not find color curves with promising features in the LSST bands from the rest-frame UV region. Given that the shapes of the color curves are somewhat uncertain, we advocate triggering follow-up observations at least in *r, i,* and *z* band. With just three bands, it is likely that no promising color curve will be available, but still, we get good-quality light curves that can be used as well for measuring time delays. Only if follow-up resources for more bands are available observations in *y, J,* and *H* band should be triggered in addition.

In Chapter 4, we will see that reducing the observational noise by targeting good-quality bands is significantly more important than lowering microlensing uncertainties. Further, we will see a more quantitative analysis of potential bands for follow-up observations if light curves are used for the time-delay measurement.

The rest of this chapter contains our publication Huber et al. (2021) with all the details related to this summary. I developed the code to calculate microlensing on SNe Ia in Huber et al. (2019) and applied it here to four theoretical models provided by Markus Kromer and Stuart Sim. Ulrich Noebauer developed the code to extract spherically symmetric specific intensity profiles from the theoretical models for Huber et al. (2019), which I modified to test the effects of asymmetries. Further, I did all the analysis described in the paper, created all figures, respectively tables, and am the paper's main author.

HOLISMOKES - III. Achromatic phase of strongly lensed type Ia supernovae

S. Huber, S. H. Suyu, U. M. Noebauer, J. H. H. Chan, M. Kromer, S. A. Sim, D. Sluse, and
S. Taubenberger

ABSTRACT

To use strongly lensed Type Ia supernovae (LSNe Ia) for cosmology, a time-delay measurement between the multiple supernova (SN) images is necessary. The sharp rise and decline of SN Ia light curves make them promising for measuring time delays, but microlensing can distort these light curves and therefore add large uncertainties to the measurements. An alternative approach is to use color curves where uncertainties due to microlensing are significantly reduced for a certain period of time known as the achromatic phase. In this work, we investigate in detail the achromatic phase, testing four different SN Ia models with various microlensing configurations. We find on average an achromatic phase of around three rest-frame weeks or longer for most color curves, but the spread in the duration of the achromatic phase (due to different microlensing maps and filter combinations) is quite large and an achromatic phase of just a few days is also possible. Furthermore, the achromatic phase is longer for smoother microlensing maps and lower macro-magnifications. From our investigations, we do not find a strong dependency on the SN model or on asymmetries in the SN ejecta. We find that six rest-frame LSST color curves exhibit features such as extreme points or turning points within the achromatic phase, which make them promising for time-delay measurements; however, only three of the color curves are independent. These curves contain combinations of rest-frame bands *u, g, r,* and *i,* and to observe them for typical LSN Ia redshifts, it would be ideal to cover (observer-frame) filters *r, i, z, y, J,* and *H*. If follow-up resources are restricted, we recommend *r, i,* and *z* as the bare minimum for using color curves and/or light curves since LSNe Ia are bright in these filters and observational uncertainties are lower than in the infrared regime. With additional resources, infrared observations in *y, J,* and *H* would be useful for obtaining color curves of SNe, especially at redshifts above ~0.8 when they become critical.

Credit: Huber et al., A&A 646, A110, 2021, published in A&A with Open Access MPI agreement ©Huber.

3.1 Introduction

On the one hand, there is a tension in the Hubble constant H_0 of at least 4σ (Verde et al., 2019) between the early Universe measurements (Planck Collaboration, 2020) and late Universe measurements from the Cepheids distance ladder (e.g., Riess et al., 2016, 2018, 2019). On the other hand, the Hubble constant from the distance ladder using the tip of the red giant branch (Freedman et al., 2019, 2020) or surface brightness fluctuations for the calibration of Type Ia supernovae (SNe Ia, Khetan et al., 2021) is consistent with both the Planck results and the Cepheids. To assess this tension and whether new physics is required to reconcile it, independent methods and measurements of H_0 are important. Lensing time-delay cosmography is a powerful tool for measuring H_0 in a single step (Refsdal, 1964), independent of other probes. The time delay can be inferred from a variable source, strongly lensed into multiple images by an intervening galaxy or galaxy cluster. This technique has been applied successfully to six lensed quasars to measure H_0 with 2.4 percent precision (e.g., Suyu et al., 2017; Courbin et al., 2018; Bonvin et al., 2018; Birrer et al., 2019; Sluse et al., 2019; Rusu et al., 2019; Chen et al., 2019; Wong et al., 2020), and more systems are being analyzed (Shajib et al., 2020; Millon et al., 2020; Birrer et al., 2020).

Instead of quasars, strongly lensed type Ia supernovae (LSNe Ia) are promising for measuring H_0 given that: (1) characteristic supernova (SN) light curve shapes make time-delay measurements possible on shorter time scales, (2) SNe fade away, facilitating measurements of the dynamics of the lens (Barnabè et al., 2011a; Yıldırım et al., 2017; Shajib et al., 2018; Yıldırım et al., 2020) to break model degeneracies, for example the mass-sheet degeneracy (Falco et al., 1985; Schneider & Sluse, 2014), and (3) SNe Ia are standardizable candles that allow us to break model degeneracies, independently of dynamics, for lens systems whose lensing magnifications are well characterized (Oguri & Kawano, 2003; Foxley-Marrable et al., 2018). So far two LSNe with resolved multiple images have been observed, namely SN "Refsdal" (Kelly et al., 2016a,b) and iPTF16geu (Goobar et al., 2017), but we expect to find 500 to 900 LSNe Ia (Quimby et al., 2014; Goldstein & Nugent, 2017; Goldstein et al., 2018; Wojtak et al., 2019) with the upcoming Rubin Observatory Legacy Survey of Space and Time (LSST).

To measure time delays between different images of a LSN Ia, one could use spectra, light curves and/or color curves. Problems arise from microlensing; this phenomenon is similar to strong lensing but instead of galaxies and galaxy clusters, the compact objects (for example, stars) located in the main lens also deflect the light. Due to the low masses of the microlenses, multiple images are typically not resolvable and for these cases microlensing is only observable as additional magnification. Originally predicted by Chang & Refsdal (1979), this phenomenon was unambiguously detected for the first time by Irwin et al. (1989) in the quasar QSO 2237+0305 as uncorrelated brightness variations between the four multiple images. More information on quasar microlensing is available in, for example, Schmidt & Wambsganss (2010) and Moustakas et al. (2019), and a general overview of microlensing is available in, for example, Wambsganss (2006) and Mediavilla et al. (2016). For our case, the additional magnification from stars in the lens galaxy influences images independently from one another and therefore adds uncertainties to the delay measurement (Yahalomi et al., 2017; Goldstein et al., 2018; Foxley-Marrable et al., 2018; Huber et al., 2019; Pierel & Rodney, 2019).

While the influence of microlensing on spectra and light curves is strong in certain configurations, color curves have the following advantage: If microlensing affects light curves from different filters in a similar way, it cancels out in the color curves. This was first investigated by Goldstein et al. (2018), who show that microlensed color curves of LSNe Ia are "achromatic," in other words, their color curves are independent of microlensing for up to three rest-frame weeks after explosion, and therefore time-delay uncertainties are reduced from approximately 4% to 1% if color curves are used instead of light curves. Huber et al. (2019) investigated this further using the spherically symmetric W7 model (Nomoto et al., 1984) (similar to Goldstein et al. (2018)) for a specific magnification map (with lensing convergence $\kappa = 0.6$, shear $\gamma = 0.6$, and smooth matter fraction $s = 0.6$). Huber et al. (2019) also find the presence of an achromatic phase, but only for color curves where the specific intensity profiles are similar. In this paper, we explore the achromatic phase further, notably to test if it is only present in the W7 model or if other SN Ia models, including multidimensional and asymmetric ones, also show an achromatic phase. In addition, we investigate the dependency of the duration of the achromatic phase on different microlensing maps and color curves.

This paper is organized as follows. In Sect. 3.2, we present the different SN Ia models investigated in this work. The calculation of microlensed SN Ia light curves is shown in Sect. 3.3, and results are presented in Sect. 3.4. We conclude in Sect. 3.5.

3.2 SN Ia models and color curves

In Sect. 3.2.1 we give a short introduction to SNe Ia and the four theoretical models we use in this work. In Sect. 3.2.2 we compare the color curves from the theoretical models to an empirical model.

3.2.1 Theoretical SN Ia models

SNe Ia most likely have their origin in a thermonuclear explosion of a carbon-oxygen white dwarf (WD) (e.g., Hoyle & Fowler, 1960), but details about the progenitor and explosion mechanism are still unknown. The classical textbook scenario is the single degenerate case where a non-rotating WD is stable until it approaches the Chandrasekhar mass limit, $M_{\mathrm{Ch}} \approx 1.4 M_{\odot}$ (Whelan & Iben, 1973; Nomoto, 1982), due to the accretion from a main sequence star or a red giant. Today this classical scenario has lost some of its relevance and there are other mechanisms considered where the WD explodes before the Chandrasekhar mass is reached, which are typically called sub-Chandrasekhar (sub-Ch) explosions (Sim et al., 2010). Furthermore, for the thermonuclear burning, one distinguishes between detonation (supersonic shock) and deflagration (subsonic heat conduction). Another approach for SNe Ia is the violent merger mechanism that belongs to the family of double-degenerate scenarios in which the companion is another WD (see Pakmor et al., 2010, 2011, 2012, 2013). At the moment it is not clear which of the scenarios is the right one to describe SNe Ia and whether multiple scenarios or just a single one can explain the observed SN Ia explosions. More details can be found in, for example, Hillebrandt & Niemeyer (2000), Hillebrandt et al. (2013), and Livio & Mazzali (2018).

In this work we investigate four different SN Ia models to test the dependency of the achromatic phase on these models. The first one is the W7 model (Nomoto et al., 1984), a (parameterized) 1D deflagration of a carbon-oxygen white dwarf close to the Chandrasekhar mass (M_{Ch} CO WD) with $0.59 M_{\odot}$ of ^{56}Ni, which is known to reproduce key observables of normal SNe Ia, for example, spectra (Jeffery et al., 1992; Nugent et al., 1997; Baron et al., 2006; Gall et al., 2012). Furthermore, we investigate the N100 model (Seitenzahl et al., 2013), which is a delayed detonation model, in other words, the burning starts as 3D deflagration and transitions into a detonation later, of a M_{Ch} CO WD. In this particular delayed detonation model, the explosion produced $0.6 M_{\odot}$ of ^{56}Ni. The third model is a sub-Ch model with a carbon-oxygen WD of $1.06 M_{\odot}$ producing $0.56 M_{\odot}$ of ^{56}Ni (Sim et al., 2010). In addition, we consider a merger model from Pakmor et al. (2012) where two carbon-oxygen WDs of $0.9 M_{\odot}$ and $1.1 M_{\odot}$ collide and ignite a detonation in which $0.62 M_{\odot}$ of ^{56}Ni are produced.

From our theoretical SN models we can get the observed flux $F_{\lambda,\mathrm{o}}(t)$ from which we can calculate AB magnitudes following

$$m_{\mathrm{AB,X}}(t) = -2.5 \log_{10}\left(\frac{\int \mathrm{d}\lambda\, \lambda S_{\mathrm{X}}(\lambda) F_{\lambda,\mathrm{o}}(t)}{\int \mathrm{d}\lambda\, S_{\mathrm{X}}(\lambda)\, c\, \lambda^{-1}} \times \mathrm{cm}^2\, \mathrm{erg}^{-1} \right) - 48.6 \tag{3.1}$$

(Bessell & Murphy, 2012), where c is the speed of light and $S_{\mathrm{X}}(\lambda)$ is the transmission function[19] for the LSST filter X (that can be u, g, r, i, z, or y), which are illustrated in Fig. 3.1.

3.2.2 Color curves of different SN Ia models

In this work we are especially interested in color curves and the dependence of the achromatic phase on the assumed theoretical SN Ia model. Therefore, we first compare in Fig. 3.2 the color curves of the SN Ia models used in this work to the empirical SN Ia model `SNEMO15` from Saunders et al. (2018) for six LSST color curves. We pick `SNEMO15` instead of `SNEMO2` or `SNEMO7` as this provides the largest variety in colors and therefore represents best the scatter of the colors based on real observations. To produce the median and 2σ (97.5th percentile − 2.5th percentile) curves of `SNEMO15`, we consider all 171 SNe Ia from Saunders et al. (2018) used for training and validation of the empirical SN model. Furthermore, we use the flux from the empirical model to calculate the LSST color curves using Eq. (3.1). We note that the data cover only 3305 Å to 8586 Å, and therefore color curves containing the z and y bands cannot be calculated. Considering the u band, we find that the filter transmission becomes relevant around 3200 Å. Therefore, our presented u band light curve is an approximation, but a reasonably accurate one since flux drops further in the UV

---

[19] `https://github.com/lsst/throughputs/tree/master/baseline`

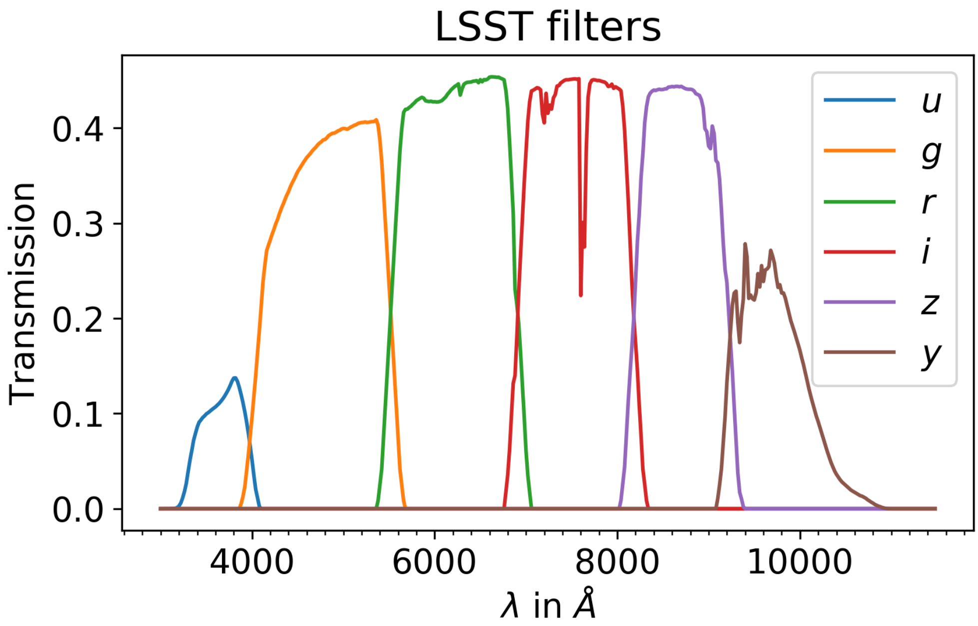


Fig. 3.1. Transmission for all six LSST filter bands. The effective wavelengths are $\lambda_{\mathrm{eff},u}$ = 3671 Å, $\lambda_{\mathrm{eff},g}$ = 4827 Å, $\lambda_{\mathrm{eff},r}$ = 6223 Å, $\lambda_{\mathrm{eff},i}$ = 7546 Å, $\lambda_{\mathrm{eff},z}$ = 8691 Å, and $\lambda_{\mathrm{eff},y}$ = 9712 Å.

region at short wavelengths and the LSST filter transmission in the missing 100 Å region is low.

The comparison of the W7 model from this work to the one presented in Goldstein et al. (2018), where a different radiative transfer code has been used, shows a different trend in the color evolution. Details about uncertainties in the light curve shapes due to different ionization treatments in `ARTIS` can be found in Kromer & Sim (2009). A spectral comparison of different SN Ia models shows that differences are larger in early stages (~8 days) in comparison to later epochs (~30 days), but Suyu et al. (2020) point out as well that the exact spectral shapes depend on approximations that are used in the radiative transfer calculations (Dessart et al., 2014; Noebauer et al., 2017).

As we see in Fig. 3.2, none of the theoretical models are able to predict the color curves perfectly, but theses models are required to investigate impacts of microlensing on color curves and it is particularly interesting, how the microlensing signal differs between models. Furthermore, even though the values of the color curves for different models are offset, the overall color evolution do match the trend of `SNEMO15`. Although features are stronger in the theoretical models, they are still present in the color curves from `SNEMO15`. To measure time delays from color curves, features like extreme points or turning points would be crucial, to mitigate color differences between images due to differential dust extinction (Eliasdottir et al., 2006). Furthermore, these extreme or turning points should be located within the achromatic phase, to reduce uncertainties due to microlensing.

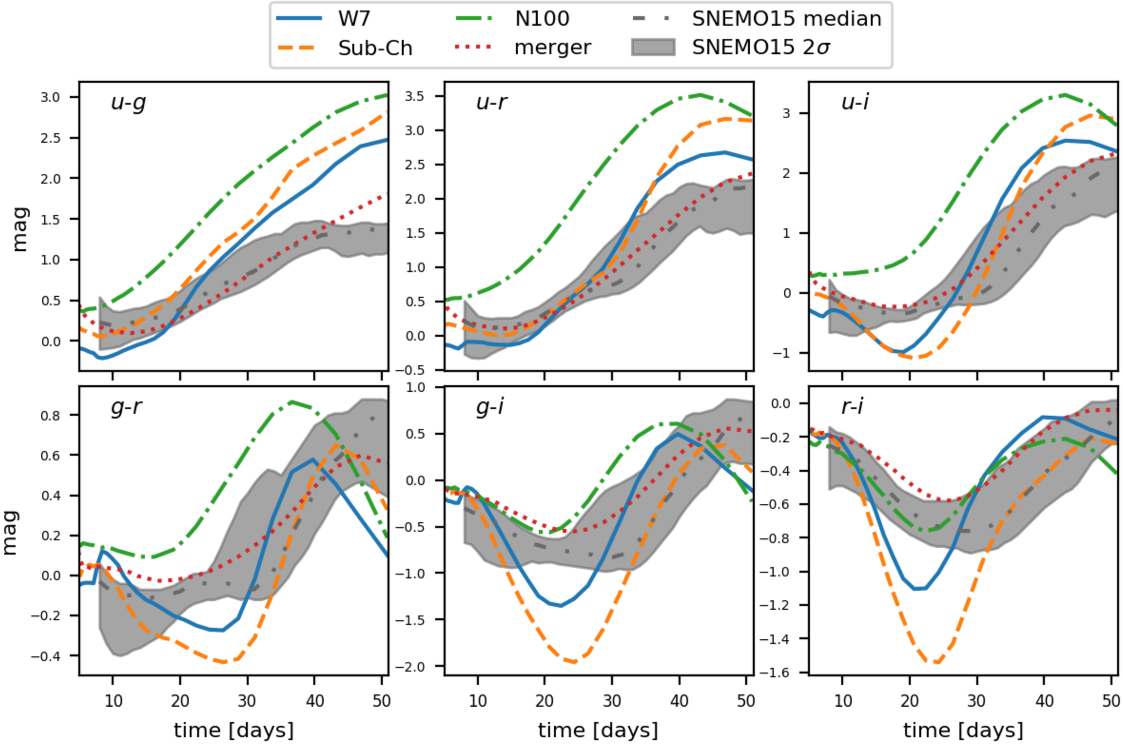


Fig. 3.2. Comparison of the four theoretical models: W7, N100, sub-Ch, and the merger to the empirical model `SNEMO15`.

3.3 Microlensing on SNe Ia

The calculation of microlensing on SNe Ia, which we use in this work, is described in detail by Huber et al. (2019), who where motivated by the work of Goldstein et al. (2018). While both assumed the W7 model for the SNe Ia, Huber et al. (2019) calculated synthetic observables using the radiative transfer code `ARTIS` (Kromer & Sim, 2009) whereas Goldstein et al. (2018) used `SEDONA` (Kasen et al., 2006). In the following, we briefly summarize the general idea.

To calculate microlensed light curves, we use the emitted specific intensity $I_{\lambda,e}(t,p)$ at the source plane calculated via `ARTIS` for a given SN model, where $I_{\lambda,e}(t,p)$ is a function of wavelength λ, time since explosion t, and impact parameter p, which is the projected distance from the ejecta center. We combine $I_{\lambda,e}(t,p)$ with magnification maps from `GERLUMPH` (Vernardos et al., 2015, J. H. H. Chan in preparation), which uses the inverse ray-shooting technique (e.g., Kayser et al., 1986; Wambsganss et al., 1992; Vernardos & Fluke, 2013) yielding the magnification factor $\mu(x,y)$ as a function of cartesian coordinates x and y on the source plane[20].

Throughout this work, the specific intensity is treated in spherical symmetry and has therefore just a 1D spatial dependency on p. This approximation is exact for the W7 model and the sub-Ch model of Sim et al. (2010), and good for the N100 model, which produces nearly spherically symmetric ejecta, but results inferred from the asymmetric merger model are questionable. To estimate the impact of viewing angle effects we investigate the asym-

---

[20] We note that μ denotes the magnification factor and not $\cos\theta$ as usually in radiative transfer equations.

metries in the merger model by using only photons from one half of the ejecta, for example, only averaging over photons that emerge in positive x-direction.

Magnification maps are determined by three main parameters, namely the convergence κ, the shear γ, and the smooth matter fraction $s = 1 - \kappa_*/\kappa$, where κ_* is the convergence of the stellar component. Furthermore, we assume a Salpeter initial mass function with a mean mass of the point mass microlenses of $\langle M \rangle = 0.35 M_\odot$. Details of the initial mass function have negligible impact on our studies (J. H. H. Chan in preparation). Our maps have a resolution of 20000 × 20000 pixels with a total size[21] of $10 R_\mathrm{Ein} \times 10 R_\mathrm{Ein}$. The Einstein Radius is a characteristic size of the map and can be calculated via

$$R_\mathrm{Ein} = \sqrt{\frac{4G\langle M \rangle}{c^2} \frac{D_\mathrm{s} D_\mathrm{ds}}{D_\mathrm{d}}}, \tag{3.2}$$

where D_s, D_d and D_ds are the angular diameter distances from the observer to the source, from the observer to the lens (deflector), and between the lens and the source, respectively. To calculate these distances we assume a flat ΛCDM cosmology where we neglect the contribution of radiation ($H_0 = 72\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ and $\Omega_m = 0.26$ as assumed by Oguri & Marshall (2010, hereafter OM10), our reference for typical LSN Ia images used in this work).

To calculate the observed microlensed flux we place the SNe Ia in the magnification map and solve:

$$F_{\lambda,\mathrm{o}}(t) = \frac{1}{{D_\mathrm{lum}}^2 (1 + z_\mathrm{s})} \int \mathrm{d}x \int \mathrm{d}y \, I_{\lambda,\mathrm{e}}(t, x, y) \mu(x, y), \tag{3.3}$$

where D_lum is the luminosity distance to the source. Furthermore, we interpolate the emitted specific intensity $I_{\lambda,\mathrm{e}}$ onto a 2D cartesian grid (x, y) with the same spatial resolution as the microlensing map and integrate over the whole size of the SN (Huber et al., 2019), which depends on the time after explosion. We then obtain the AB-magnitudes via Eq. (3.1).

3.4 Achromatic phase of LSNe Ia

In this section, we investigate the dependency of the achromatic phase on different parameters. Sect. 3.4.1 probes the four different SN Ia models (W7, N100, sub-Ch model, and the merger models) under the assumption of spherical symmetry. In addition, the dependency on the smooth matter fraction s and different image configurations (varying κ and γ) is discussed. Sect. 3.4.2 shows the dependency on the scale of the microlensing map defined via R_Ein, which depends on the source and lens redshifts, z_s and z_d, respectively. In Sect. 3.4.3, effects of asymmetries in the merger model are investigated. While Sect. 3.4.1 to 3.4.3 assume rest-frame color curves, redshifted color curves are investigated in Sect. 3.4.4. In Table 3.1, we summarize the number of models and different magnification maps that are taken into account in the different sections.

---

[21] As a cross-check we investigated for a few cases larger maps with a total size of $20 R_\mathrm{Ein} \times 20 R_\mathrm{Ein}$, and our conclusions drawn in this work do not depend on the size of the maps.

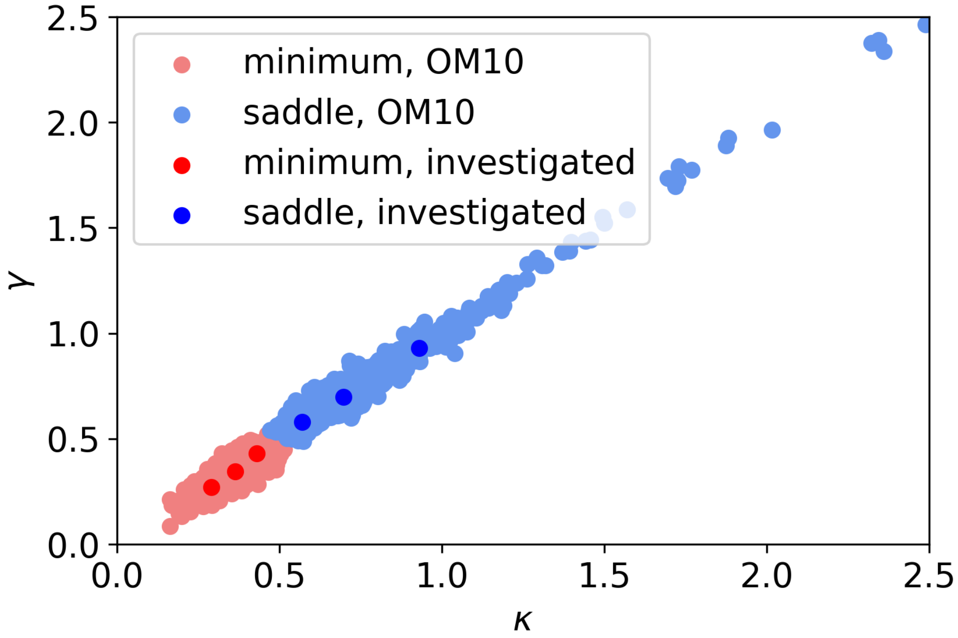


Fig. 3.3. Distribution of κ and γ values for two different image types (minimum and saddle, in red and blue, respectively), showing the sample of the OM10 catalog and the six pairs we have investigated ($\kappa, \gamma = (0.29, 0.27), (0.36, 0.35), (0.43, 0.43), (0.57, 0.58), (0.70, 0.70)$, and $(0.93, 0.93)$). There are a few saddle points not shown in the plot going up to κ and γ values of around 10.

Section	models	asymmetries	z_s, z_d	κ, γ	s	color curves
3.4.1	4	no	1	6	5	15
3.4.2	4	no	4	2	5	15
3.4.3	1	6	1	2	5	15
3.4.4	4	no	3	2	5	15

Table 3.1. Summary of the number of different parameters investigated in Sects. 3.4.1 to 3.4.4. In Sect. 3.4.1 the dependency on different models, smooth matter fractions and image configurations is investigated. Sect. 3.4.2 exhibits the dependency on the scale of the magnification map, Sect. 3.4.3 contains the investigations of asymmetries for the merger model, and Sect. 3.4.4 shows redshifted color curves.

3.4.1 SN Ia models, smooth matter fraction, and image configuration

To investigate the achromatic phase of LSNe Ia, we pick typical lens and image configurations from the OM10 catalog (Oguri & Marshall, 2010). For the source and lens redshifts we assume $z_\mathrm{s} = 0.77$ and $z_\mathrm{d} = 0.32$, which are the median values of the OM10 catalog. We use redshifts in Sects. 3.4.1 to 3.4.3 only to calculate the scale of the microlensing maps, hence R_Ein, and therefore the color curves discussed in these sections are in the rest-frame. The investigated κ and γ values are also based on OM10 and shown in Fig. 3.3 as six dark points for two different image types (minimum and saddle, in red and blue, respectively). For each of the two image types, the investigated points correspond to the median values and the 16th and 84th percentiles of the OM10 sample, taken separately for κ and γ. For each of the six pairs of κ and γ, five different s values (0.1, 0.3, 0.5, 0.7, and 0.9) are considered, which cover typical s values at image positions of galaxy-scale lenses (e.g., Schechter et al., 2014; Chen et al., 2018; Bonvin et al., 2019). Therefore, we have in total 30 magnification maps (from 6×5). Magnification maps for most configurations we investigate in this work are shown in Appendix 3.6.1.

In comparison to the 30 different magnification maps probed in this work, Huber et al. (2019) have analyzed only a single magnification map ($\kappa, \gamma, s = 0.6$). Goldstein et al. (2018) investigated a much larger sample of LSNe Ia with 78184 multiple images, but the sample is dominated by small-image-separation systems that will not be resolvable from ground-based monitoring, whereas we focus on spatially resolvable systems for cosmography by using the OM10 sample of strong lenses. Although our sample of LSNe Ia is much smaller, we probe per map 10000 random positions instead of just one as by Goldstein et al. (2018); therefore, we investigate a much larger sample of microlensed SN Ia color curves, which further allows us to probe dependencies such as the duration of the achromatic phase on different image configurations.

In our analysis, we draw for each of the 30 magnification maps 10000 random positions, where we calculate for each position all six LSST microlensed light curves following Eq. (3.1) and from these the 15 LSST color curves (obtained through all possible pair-wise combinations of the light curves). For each color curve, we consider the 1σ band ("full-width," corresponding to the difference between the 84th percentile and 16th percentile) and 2σ band (full-width, corresponding to the difference between the 97.5th percentile and 2.5th percentile) from the 10000 random positions, which is shown in Fig. 3.4 for *g*-*i* for $\kappa = 0.36, \gamma = 0.35$ and $s = 0.5$ for all 4 SN models investigated in this work. We use 32 time bins, covering rest-frame days 4.2 to 50.8 after explosion of the SNe Ia.

The black vertical lines in Fig. 3.4 correspond to the duration of the achromatic phase t_achro which we define as the first of two neighboring time bins where the 2σ band becomes larger than the threshold

$$\Delta_\mathrm{threshold} = \max(0.05\,\mathrm{mag}, \Delta_{3\%}), \tag{3.4}$$

where $\Delta_{3\%} = 0.03\,(\max(\mathrm{color\,curve}) - \min(\mathrm{color\,curve}))$. The reasoning behind these definitions is as follows. Typically the 2σ band increases with time but for rare cases it can exceed

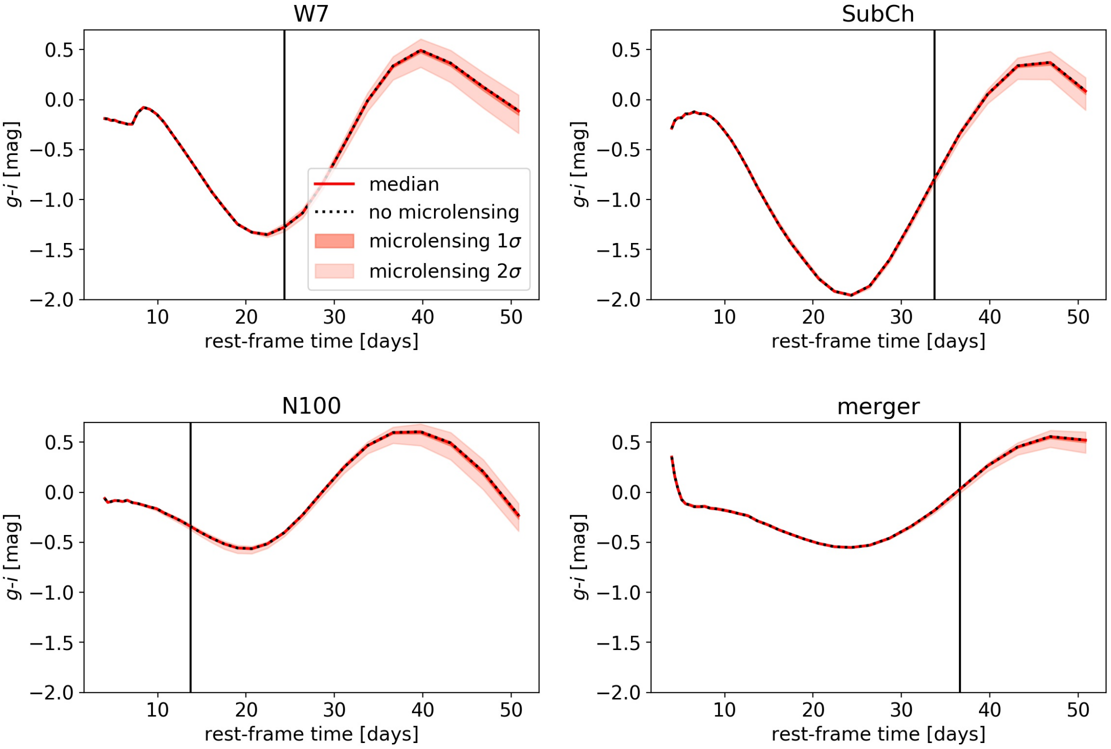


Fig. 3.4. Rest-frame LSST color curves for 10000 random SN positions in the magnification map with $\kappa = 0.36, \gamma = 0.35$ and $s = 0.5$, comparing microlensed color curves (with median in solid red, and 1σ and 2σ band in different shades) to non-microlensed ones (dotted black). The vertical black lines indicate the duration of the achromatic phase. We find different durations of the achromatic phase for different SN Ia models, where the N100 model in this specific case has the shortest and the merger model the longest duration. This is just a specific case to illustrate the dependency of the achromatic phase on the specific intensity profiles. More general conclusions can be drawn from Figures 3.6 and 3.7.

$\Delta_{\rm threshold}$ for a single time bin and then drop below that limit again. With our definition of $t_{\rm achro}$, including more than one time bin that exceeds $\Delta_{\rm threshold}$, we skip such outliers. If multiple time bins are within one day we require that all time bins exceed the $\Delta_{\rm threshold}$ in the 2σ band and set $t_{\rm achro}$ as the lowest time bin[22].

To motivate the 0.05 mag criterion, we look at a LSNe Ia at the median source redshift of OM10, $z_{\rm s} = 0.77$, and assume as an image the median minimum corresponding to $(\kappa,\gamma) = (0.36, 0.35)$ leading to a macro-magnification of 3.5. According to Huber et al. (2019), follow-up observations going between 1 and 2 mag deeper than LSST is ideal to measure time delays of LSNe Ia. Assuming observations of around 1.5 mag deeper than the LSST-like 5σ depth, we expect the 2σ uncertainties at the light curve peak of around 0.025 mag in r and i bands and higher uncertainties in other bands (LSST Science Collaboration, 2009). This 2σ uncertainty corresponds to 0.05 mag in the 2σ band, and therefore our first limit in Eq. (3.4) assures that microlensing uncertainties within the achromatic phase are smaller than typical observational uncertainties. This is a very conservative estimate, especially for color curves having much larger changes than 0.05 mag over time. This motivates the use of $\Delta_{3\%}$ as the second criterion in Equation (3.4). This means that for color curves covering more than ~1.7 mag over time the 0.05 mag criterion is replaced by the slightly larger value $\Delta_{3\%}$. Variations of color curves are always within 3 mag so uncertainties due to microlensing are kept within 0.09 mag in the achromatic phase.

As an illustration, the g-i color curve in the upper-right-hand panel (sub-Ch) of Fig. 3.4 would have an achromatic phase of ~15 d with only the first criterion in Equation (3.4), but an achromatic phase of ~35 d with both requirements, which is more appropriate given the large color variation. As a cross-check, we compared different definitions of the achromatic phase, namely, using only $\Delta_{\rm threshold} = 0.05$ or taking the mean of the first two time bins instead of requiring multiple neighboring bins. Although this changes the duration of the achromatic phase for some cases, our general conclusions from averaging over many different microlensing maps, color curves or models, are not influenced. While our definition of the achromatic phase is arbitrary to some extent, it is justified by looking at many plots as shown in Fig. 3.4, and we keep the definition consistent over the whole work, which allows us to compare different models and microlensing parameters.

In Fig. 3.4, we see that the duration of the achromatic phase for this example depends on the SN Ia model. While the model N100 has a short achromatic phase of ~15 d, the merger and sub-Ch models show a significantly longer duration of ~35 d and the W7 model is in between with ~25 d. This is related to the specific intensity profiles, shown in Fig. 3.5, for the radial distribution of the radiation in the six LSST filters. The influence of different specific intensity profiles on spectra and light curves is presented in detail in Appendix A of Huber et al. (2019).

The profiles for filters g and i for the sub-Ch model and N100 at day 14.9 show larger differences than W7 and the merger model, explaining the much shorter achromatic phase for the N100 model. The sub-Ch model reaches the 0.05 mag at a similar time as the N100 model but given the large range covered by the color evolution, the achromatic phase of the

---

[22] Only relevant for first few time bins since time is binned logarithmically.

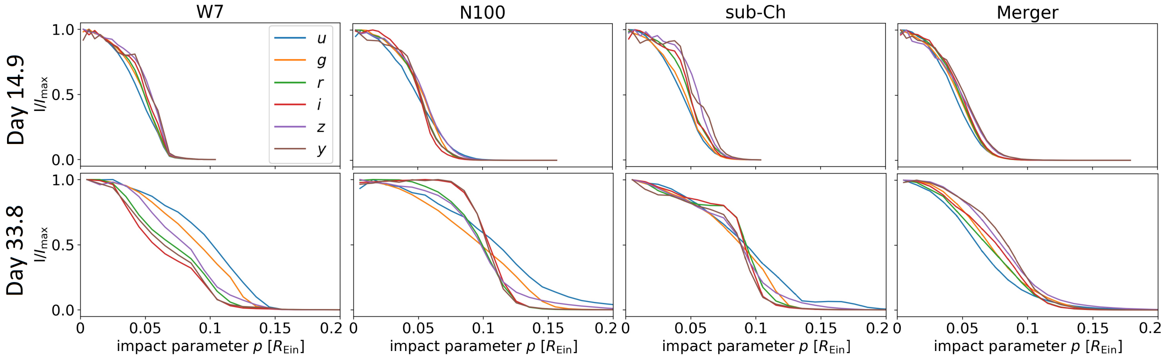


Fig. 3.5. Radial specific intensity profiles for four SN Ia models (W7, N100, the sub-Ch, and the merger model in the four labeled columns) for the six LSST filters at rest-frame day 14.9 (top row) and day 33.8 (bottom row) after explosion, where $R_{\mathrm{Ein}} = 2.9 \times 10^{16}$ cm for the median source and lens redshifts ($z_{\mathrm{s}}, z_{\mathrm{d}} = 0.77, 0.32$) based on OM10. Different SN Ia models yield different specific intensity profiles, leading to different durations of the achromatic phase. The wiggles for low p values at Day 14.9 are most likely due to Monte Carlos noise and we do not expect them to be physical. For our microlensing calculations, differences at higher p values are more relevant, since events where micro caustics are crossed become more likely with larger radius.

sub-Ch model is much longer. In the case of the merger model, the long achromatic time is explained by the intensity profiles because for later times (day 33.8), only the merger model still shows quite similar intensity profiles in g and i explaining the longer duration of the achromatic phase in the color curve g-i.

To draw a more general conclusion, we look at all color curves for the 6 × 5 magnification maps (see Table 3.1) for different SN Ia models as shown in Fig. 3.6 and 3.13. For each of the 30 magnification maps, we have per color curve and model a t_{achro} from the 10000 random positions as shown in Appendix 3.6.2. The vertical lines in Fig. 3.6 and 3.13 mark the mean values from the 30 t_{achro}, and are matched in color and linestyles to that of the SN Ia models. In Fig. 3.6, six color curves are shown, which we refer to as useful color curves; these color curves have, for at least three SN models, substantially non-linear features like extreme points and turning points for delay measurements within their corresponding achromatic phases. The remaining nine LSST color curves are shown in Fig. 3.13. The most promising colors are rest-frame u-i, g-r, g-i, and r-i. The colors u-g and u-r are also encouraging but only if early features are captured. Therefore, to target the most promising color curves, rest-frame filters u, g, r, and i are necessary, which lead to three independent color curves. If one looks at the median color curves of the empirical SN model `SNEMO15` in Fig. 3.2, g-r and g-i are almost featureless within the achromatic phase in comparison to the theoretical SNe models. Nevertheless the message of the filters to target is not influenced since the other four promising color curves include filters g, r, and i anyway.

If one considers the median source redshift of the OM10 of $z_{\mathrm{s}} = 0.77$, the r band would be shifted from around 6200 Å to 11 000 Å and therefore mostly not observed in the LSST bands. This means that from the six promising color curves just rest-frame u-g will be fully observed within the LSST bands. Assuming the nearby iPTF16geu system ($z_{\mathrm{s}} = 0.409$, Goo-

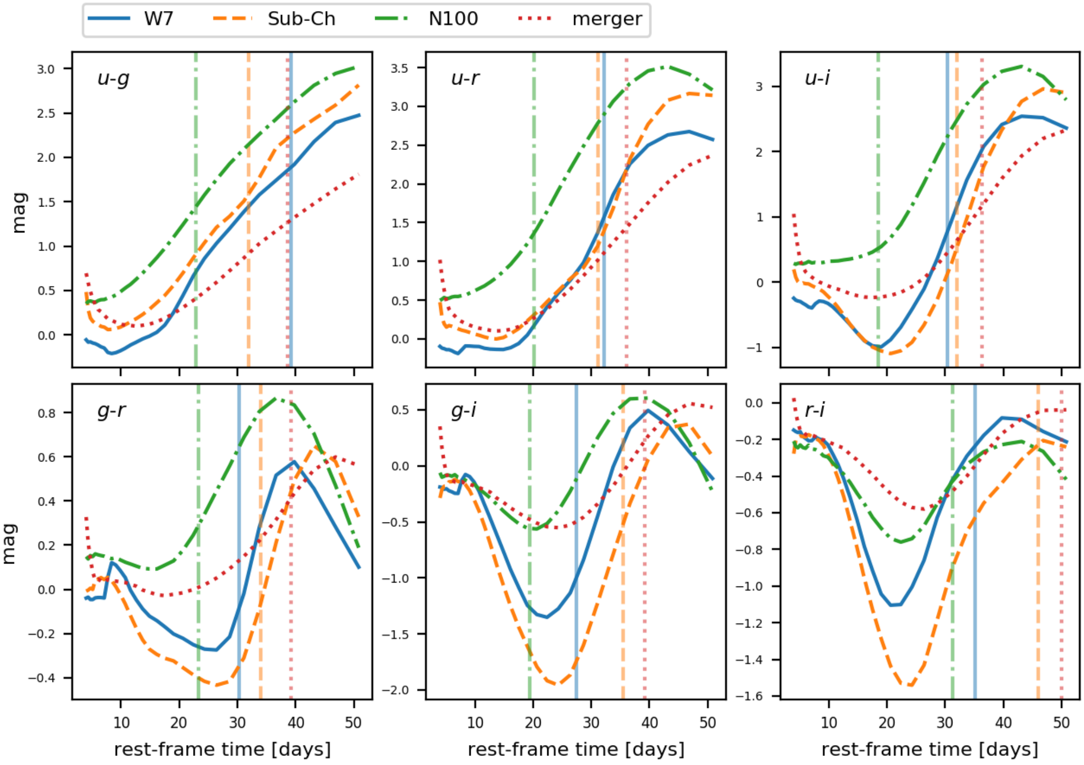


Fig. 3.6. Six rest-frame LSST color curves without microlensing for four different SN Ia models. The vertical lines mark the mean duration of the achromatic phase by averaging over 30 microlensing magnification maps (see Appendix 3.6.2). This figure contains only the color curves that are promising for time-delay measurements (i.e., color curves exhibiting features like extreme points or turning points that are located within the achromatic phase for at least three SN Ia models). For the remaining nine LSST color curves see Fig. 3.13.

bar et al. (2017)), rest-frame *u*-*g*, *u*-*r*, and *g*-*r* would be observed fully but colors containing rest-frame *i* band only partly. This suggests that follow-up observations should also be conducted in the infrared, but the results show the problem that even though we find on average an achromatic phase of around three rest-frame weeks for nearly all colors, more than half of them might not be useful for time-delay measurements due to the lack of features in the color curves. A case where redshifted color curves are investigated is discussed in Sect. 3.4.4.

We emphasize that these results are based on averaging over the investigated sample of magnification maps. In special cases, the achromatic phase might be much shorter. The results are summarized in Fig. 3.7, where the median and the 16th to 84th percentiles of the achromatic-phase duration are shown. From the left-hand panel, we find that the SN Ia models W7, sub-Ch and the merger yield on average a comparable achromatic-phase duration, and the N100 model has a shorter one. This might be related to the fact that the flux predicted by the N100 model around maximum light is too red in comparison to observations, which comes from an iron group element layer around a Ni-56 core (Sim et al., 2013), influencing the specific intensity profiles and therefore also the duration of the

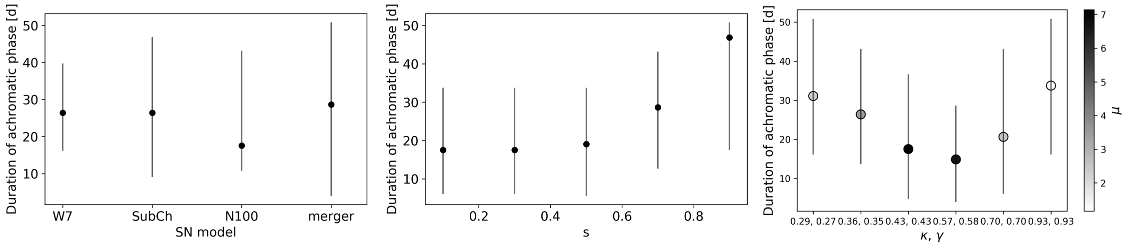


Fig. 3.7. Duration of the achromatic phase in rest-frame days for different SN Ia models (left-hand panel), different smooth matter fractions s (middle panel), and image configurations from strong lensing with their respective magnification factors μ shown in the color bar (right-hand panel). For these plots, a sample of four models, six different κ and γ values, five different s values, and 15 color curves has been investigated where we average over all parameters not shown on the x-axis. The dots correspond to the median, and the vertical bars indicate the range from the 16th to 84th percentiles. Due to limits in computing time, we only investigated the achromatic phase up to rest-frame day 50.8; therefore, for cases close to that limit, the presented result are a lower limit on the achromatic phase. While the median achromatic phase is typically $\gtrsim$20 rest-frame days, the spread due to different microlensing maps and color curves is quite large and the results can be seen in detail in Appendix 3.6.2.

achromatic phase. However, given the uncertainties on the duration (shown in the left-hand panel) and our tests of different criteria for computing the achromatic-phase duration, we conclude that we do not find a significant dependency of the average achromatic phase on the SN model. In addition, we find from the right-hand panel that strong lensing images with a high macro-magnification $\mu = ((1-\kappa)^2 - \gamma^2)^{-1}$ are influenced more by microlensing than low magnification cases. Concerning the s value, we find similar durations of the achromatic phase for $s \leq 0.5$ and an increased duration if we go to smoother maps (middle panel). Depending on the position of the lensed images, an appropriate s value can be chosen (Barnabè et al., 2011b; Oguri et al., 2014; Jiménez-Vicente et al., 2015). Overall we can say that combinations of κ, γ and s producing smoother microlensing maps yield a longer achromatic phase.

The spread of the 1σ range in Fig. 3.7 is large because there is quite some variation between different colors and microlensing maps, which makes it hard to give a general recipe for using color curves for time-delay measurements. Still, the median values in Fig. 3.7 are around three rest-frame weeks or longer and therefore follow-up resources for LSNe Ia should be allocated independent of the lensing parameters κ, γ and s.

3.4.2 Scale of magnification map

To probe the dependency on the scale of the microlensing map, namely R_{Ein}, we investigated for $\kappa, \gamma = (0.36, 0.35)$ and $\kappa, \gamma = (0.70, 0.70)$ a range of redshifts, $(z_{\mathrm{s}}, z_{\mathrm{d}}) = \{(0.77, 0.32)$, $(0.55, 0.16)$, $(0.99, 0.48)$, and $(0.99, 0.16)\}$. The first pair corresponds to the median values from the OM10 and the second and third pair are the 16th and 84th percentile separately taken for z_{s} and z_{d}. The fourth pair is the 84th percentile for z_{s} and 16th percentile of z_{d} to increase the variety of different R_{Ein}. The results are summarized in Fig. 3.8. We find a very

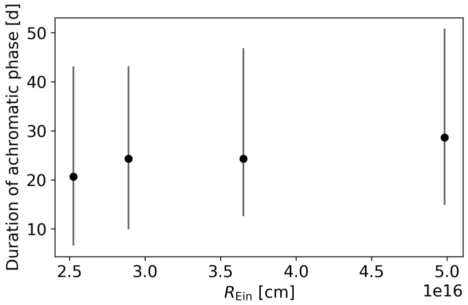


Fig. 3.8. Duration of achromatic phase as a function of R_{Ein}, the scale of variations in the microlensing map. The dots correspond to the median and the vertical bars indicate the range from the 16th to 84th percentile for the sample of four models, two different κ and γ pairs, five different s values, and 15 color curves that have been investigated, where we average over all parameters not shown on the x-axis. From left to right, R_{Ein} corresponds to $(z_{\mathrm{s}}, z_{\mathrm{d}}) = (0.99, 0.48), (0.77, 0.32), (0.55, 0.16), (0.99, 0.16)$.

slight trend that with larger R_{Ein} the duration of the achromatic phase becomes longer. Given the large uncertainties, we cannot report this trend to be significant, even though it would be plausible because with larger R_{Ein} the physical size of the magnification maps increases, which causes SNe Ia to appear smaller in these maps, and for example events where micro caustics are crossed will be less likely. Nevertheless, Fig. 3.8 suggests that, if present at all, this effect is only minor.

3.4.3 Asymmetric merger model

In this section, viewing angle effects are investigated for the asymmetric merger model. We compare the spherically symmetric approach, which takes into account photon packets leaving the SN Ia ejecta in any direction and averages over them to get the 1D dependency on the impact parameter p, to six cases where only photons from one half of the ejecta are taken into account, for example, just photons that leave the ejecta at a positive x-coordinate, which we label as $x > 0$. For this subset we also calculate the 1D impact parameter, but just averaging over photons leaving the ejecta at $x > 0$. The other cases that we investigate are $x < 0$, $y > 0$, $y < 0$, $z > 0$ and $z < 0$, and the results are shown in Fig. 3.9. We find that viewing angle effects influence shapes of color curves and also the duration of the achromatic phase, but only slightly and therefore useful color curves (marked by red frames) are the same as those pointed out in Section 3.4.1.

3.4.4 Redshifted color curves

Sect. 3.4.1 indicates that most useful color curves (curves with features for measuring time delays within the achromatic phase) in the SN rest-frame will be shifted to the infrared regime for typical redshifts expected for LSNe Ia. This section investigates if useful color curves can still be found in *ugrizy* coming from the rest-frame ultraviolet (UV), taking into account typical redshifts of SNe Ia. For this, we consider a set of representative redshifts: $z_{\mathrm{s}}, z_{\mathrm{d}} = \{(0.55, 0.16), (0.77, 0.32), \text{and } (0.99, 0.48)\}$.

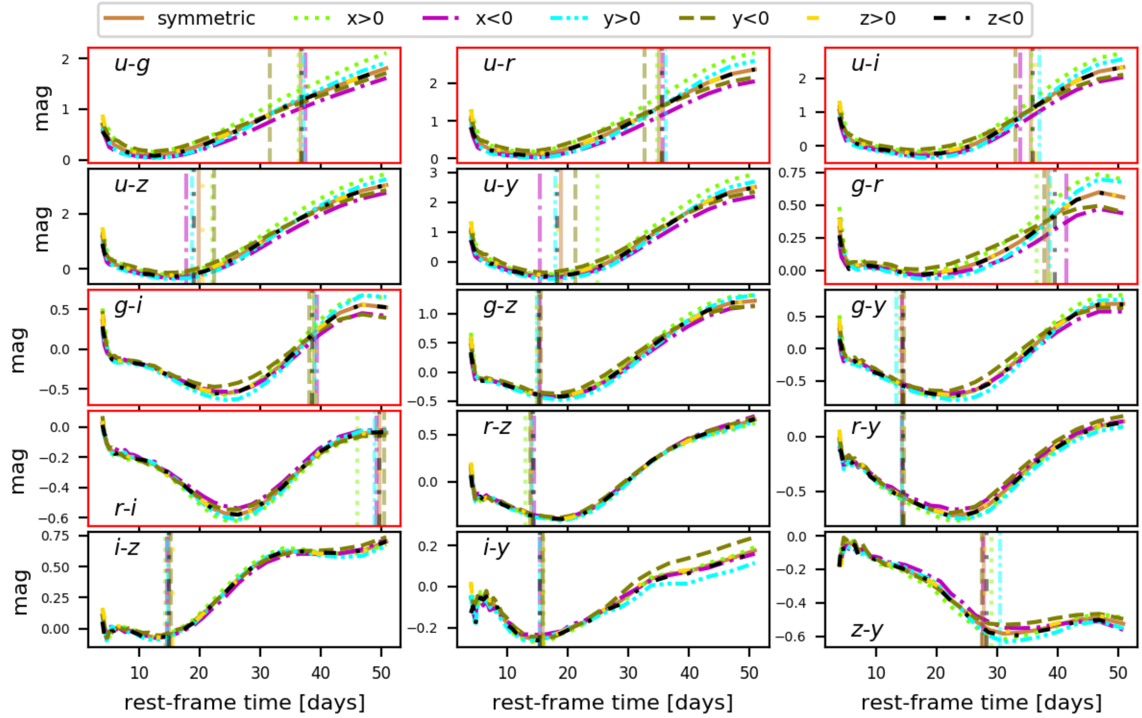


Fig. 3.9. All rest-frame LSST color curves without microlensing for the merger model assuming spherical symmetry and six asymmetric versions of the merger model. The vertical lines indicate the duration of the achromatic phase and red frames mark color curves that are promising for time-delay measurements. The similarities between the symmetric case and the asymmetric cases show that viewing angle effects do not influence our conclusions.

The results for $z_s = 0.55$ are shown in Fig. 3.10. In comparison to the rest-frame bands, we see from Fig. 3.1 that the *u* band will be observed in *r*, the *g* band in *i*, the *r* band in *y*, and the rest will be in the infrared regime not covered by LSST filters. Accordingly, we find three useful rest-frame color curves from Fig. 3.6 in Fig. 3.10, namely *u*-*g* is redshifted roughly to *r*-*i*, *u*-*r* to *r*-*y*, and *g*-*r* to *i*-*y*, although only two of the color curves are independent. Unfortunately none of the color curves coming from the rest-frame UV show strong features within the achromatic phase in at least three models that would be promising for time-delay measurements. The cases for $z_s = 0.77$ and $z_s = 0.99$ are presented in Appendix 3.6.3. For $z_s = 0.77$ we find two useful independent color curves for time-delay measurement, namely *i*-*z* and *z*-*y*, but at $z_s = 0.99$ there is only *z*-*y* remaining in the observed *u* through *y* bands. For some color curves like *g*-*i* for $z_s = 0.55$ or *r*-*z* for $z_s = 0.77$ one might argue that they are also useful if early features are captured but this is because they still contain a substantial amount of rest-frame *u*-*g*. Nevertheless we see the trend that with higher redshifts useful color curves are shifted to bands covering higher wavelengths but there are no useful color curves coming from the rest-frame UV. Therefore, the number of useful color curves observed in *u* through *y* bands decreases with higher redshift. As discussed by Suyu et al. (2020), the exact spectral shapes particularly of the rest-frame UV spectra depend on various approximations used in the radiative transfer calculations, such as metallicity of the progenitor or the number of ionization states (e.g., Lucy, 1999; Lentz et al., 2000; Kromer & Sim, 2009; Walker et al., 2012; Dessart et al., 2014; Kromer et al., 2016; Noebauer et al., 2017), and therefore the shapes of color curves in the rest-frame UV are more uncertain. With future more detailed radiative transfer calculations, one might find useful color curves also in the rest-frame UV but chances are low since only rest-frame UV colors of the merger model show features within the achromatic phase, and all other models do not (see Appendix 3.6.3).

Nevertheless these results suggest we need to follow up more in the infrared regime that corresponds to the promising rest-frame color curves shown in Fig. 3.6, especially for $z_s \gtrsim 0.6$. Apparent magnitudes for the six LSST filters as well as three infrared bands (*J*, *H* and *K*) are shown in Fig. 3.11. From this we find that in the *u* and *K* bands the light curves are too faint but follow-up in all other bands seems reasonable. Huber et al. (2019) have investigated light curves for time-delay measurement and found that the combination of *g, r* and *i* performs a few percent better than *r, i* and *z*. The reason for this is the assumed LSST-like 5σ depth, which is nearly two magnitudes shallower in the *z* band in comparison to the *g* band (in single visits). In the context of this work, *riz* might be chosen over *gri* for having more useful color curves. Nevertheless, three bands are not ideal since this would result in just two independent color curves. The more filters used for follow-up, the higher the chances are to get promising color curves. A set of six filters with *r, i, z, y, J,* and *H* seems most promising as this would include all useful rest-frame LSST color curves shown in Fig. 3.6 for typical source redshifts. If resources for a further band are available, then also the *g* band can be used. Covering this range of bands with a single follow-up telescope might be challenging. If only an optical or only an infrared facility is available, then we recommend to observe in the redder parts of the optical coverage (*rizy* bands) since

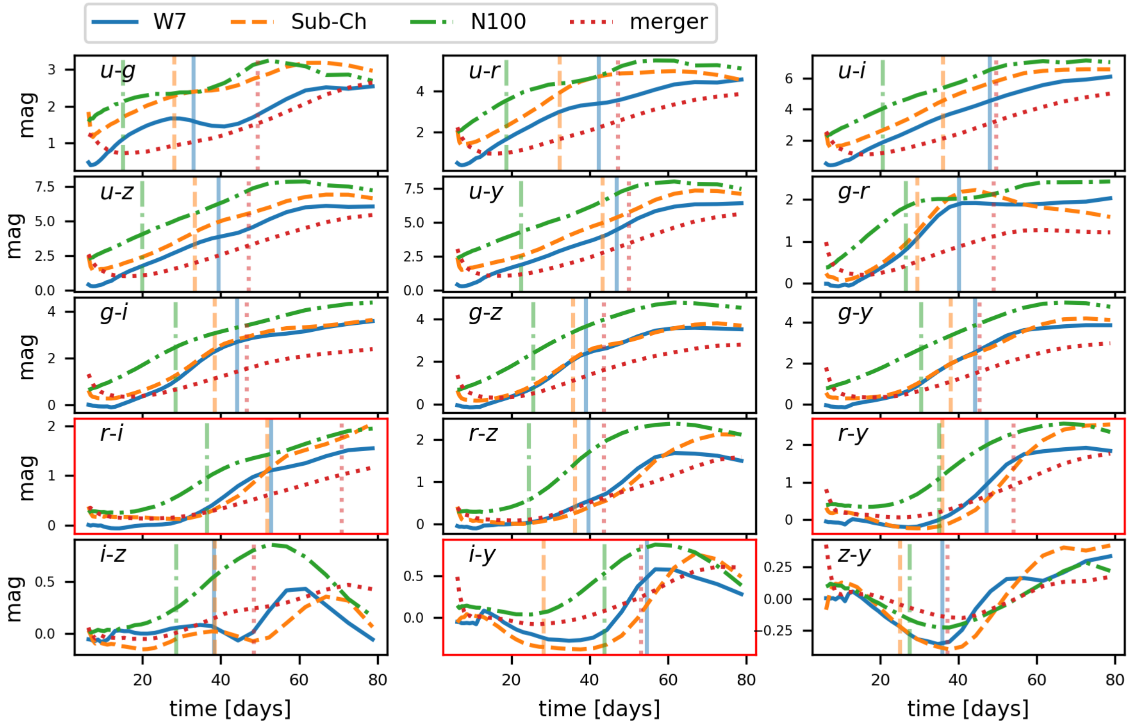


Fig. 3.10. All redshifted ($z_s = 0.55$) LSST color curves without microlensing for four different SN Ia models. The vertical lines mark the duration of the achromatic phase and red frames indicate color curves that are promising for time-delay measurements, in other words, color curves exhibiting features like extreme points or turning points that are located within the achromatic phase for at least three SNe Ia models.

these bands would yield better quality light curves given that LSNe Ia are bright there and observational uncertainties are lower than in the infrared regime. Only for high redshift cases ($z_s \gtrsim 0.9$) it would be worth to prefer the near-infrared over the optical range.

3.5 Discussion and summary

According to Goldstein et al. (2018), the achromatic phase lasts for three rest-frame weeks for the W7 model (Nomoto et al., 1984), which means that in this time frame, color curves are nearly independent of microlensing and therefore promising for time-delay measurements. In this work, we investigate in addition to the W7 model, a sub-Ch model (Sim et al., 2010), a merger model (Pakmor et al., 2012), and a 3D delayed detonation model (model N100 of Seitenzahl et al., 2013). Our results are in good agreement with Goldstein et al. (2018) leading on average to an achromatic phase around three rest-frame weeks or longer. Furthermore, we do not find a significant model dependency and also asymmetries in the merger model do not have a strong influence on the duration of the achromatic phase. While this sounds very promising for time-delay measurements, there are also downsides to report. From the 15 rest-frame LSST color curves, only six show promising features for time-delay measurements and just three of them are independent. These color curves contain combinations of the rest-frame filters *u, g, r,* and *i*. To observe these for typical LSN Ia redshifts, follow-up from bands *r* to *H* is necessary. In an ideal follow-up scenario, one

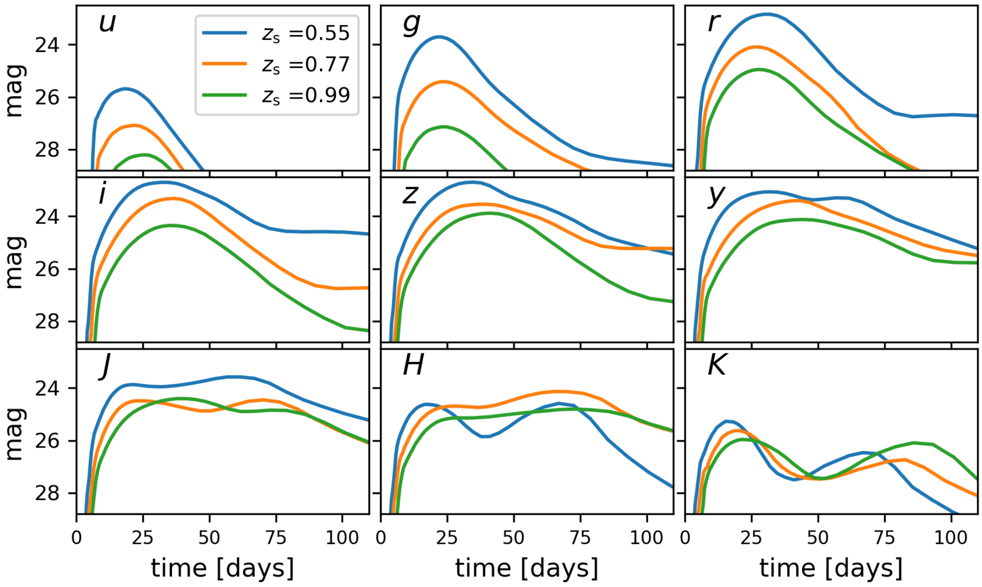


Fig. 3.11. Redshifted observed light curves for nine different filters and three different redshifts assuming the sub-Ch model. Light curves are too faint in the *u* and *K* bands but all other bands are potential candidates for follow-up observations.

would observe in bands *r, i, z, y, J, H,* and optionally also *g*. The bare minimum should cover *r, i,* and *z*. Observations just in three filters would make time-delay measurements from color curves hard, but in *riz* one can expect good quality light curves that can also be used for time-delay measurements (Huber et al., 2019) although microlensing uncertainties are larger.

Even though the duration of the average achromatic phase is around three rest-frame weeks or longer for most color curves, the spread of the duration is quite large. Depending on the configuration in the microlensing map, a very short achromatic phase of just a few days is also possible. Overall we find a longer achromatic phase for smoother microlensing maps (high s value) and image configurations with lower magnification factors. The trend that combinations of z_s, z_d, which yield a larger Einstein radius R_{Ein}, provide a longer achromatic phase is only very weak. Even though low s values and high magnification cases provide a shorter duration of the achromatic phase, the median for these images is still around three rest-frame weeks. Therefore, lens and image properties of a LSNe Ia (z_s, z_d, κ, γ and s) can be mostly neglected when allocating follow-up observations, except z_s, which sets the filters to target.

This study provides general guidance on the observing filters to follow up LSN Ia. In a real detection of a LSNe Ia where we have the SN redshift measurement, we can further refine and optimize the filters we have to employ to get promising color curves on a case-by-case basis.

Acknowledgements

We thank M. Oguri and P. Marshall for the useful lens catalog from Oguri & Marshall (2010), and W. Hillebrandt, S. Blondin, D. A. Goldstein for useful discussions. We also would like to thank the anonymous referee for helpful comments, which strengthened this work. SH and SHS thank the Max Planck Society for support through the Max Planck Research Group for SHS. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (LENSNOVA: grant agreement No 771776; COSMICLENS: grant agreement No 787886). This research is supported in part by the Excellence Cluster ORIGINS which is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC-2094 – 390783311. UMN has been supported by the Transregional Collaborative Research Center TRR33 'The Dark Universe' of the Deutsche Forschungsgemeinschaft. JHHC acknowledges support from the Swiss National Science Foundation and through European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (COSMICLENS: grant agreement No 787866). MK acknowledges support from the Klaus Tschira Foundation.

3.6 Appendix

3.6.1 Microlensing maps

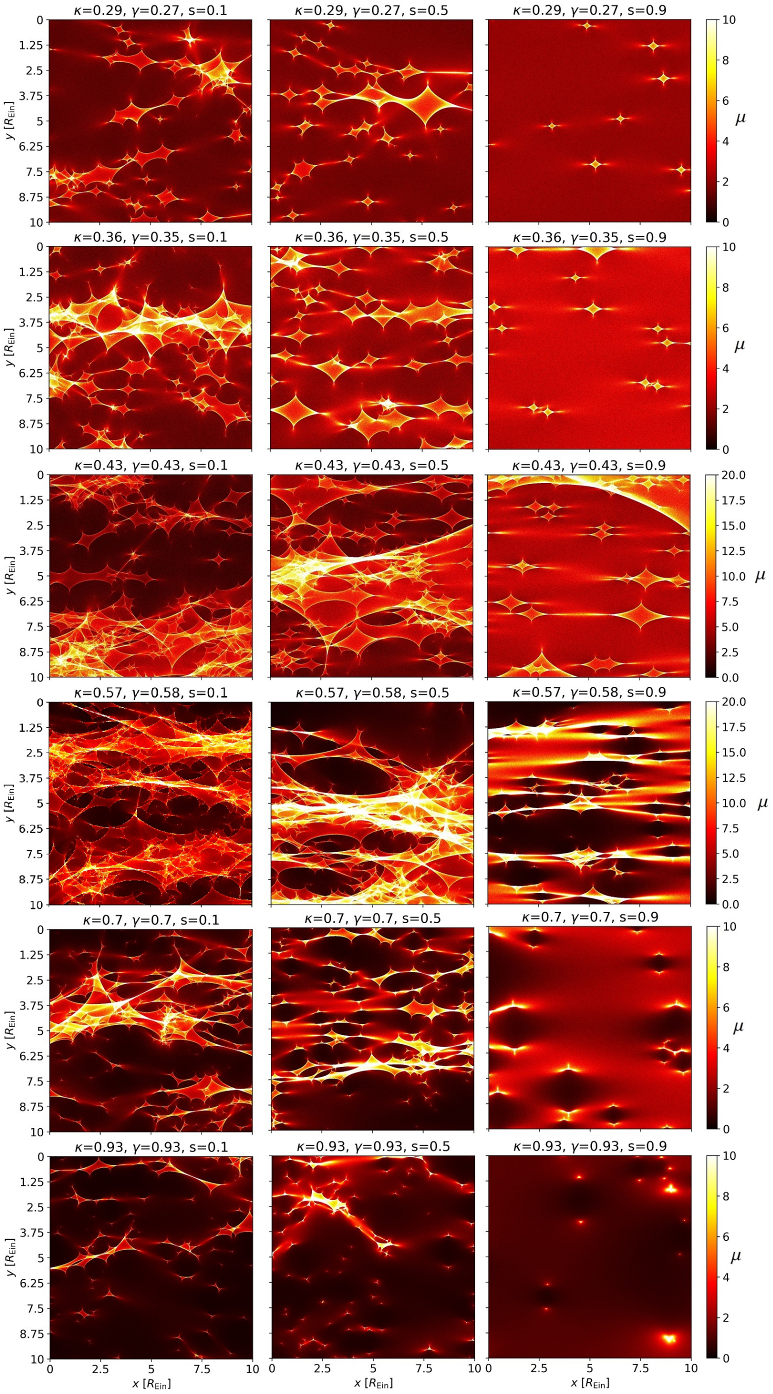


Fig. 3.12. Magnification maps for six different κ and γ values with smooth matter fraction $s = 0.1, 0.5$, and 0.9.

3.6.2 Achromatic phase in detail

	κ, γ	s	u-g	u-r	u-i	u-z	u-y	g-r	g-i	g-z	g-y	r-i	r-z	r-y	i-z	i-y	z-y
W7	0.36 , 0.35	0.1	39.8	31.1	28.7	19.0	19.0	31.1	24.3	16.2	16.2	31.1	16.2	14.9	17.5	17.5	31.1
		0.3	36.6	31.1	28.7	19.0	17.5	26.4	24.3	14.9	14.9	31.1	14.9	13.7	17.5	17.5	31.1
		0.5	36.6	31.1	26.4	17.5	17.5	26.4	24.3	14.9	14.9	28.7	14.9	13.7	17.5	16.2	31.1
		0.7	39.8	31.1	31.1	36.6	33.8	31.1	24.3	17.5	19.0	50.8	17.5	16.2	19.0	19.0	33.8
		0.9	50.8	39.8	39.8	46.8	39.8	36.6	36.6	50.8	39.8	50.8	43.2	50.8	43.2	50.8	43.2
	0.29 , 0.27	0.1	39.8	31.1	28.7	33.8	31.1	31.1	24.3	16.2	17.5	31.1	16.2	14.9	19.0	17.5	31.1
		0.3	39.8	31.1	28.7	33.8	31.1	31.1	24.3	16.2	17.5	46.8	16.2	16.2	17.5	17.5	33.8
		0.5	39.8	31.1	28.7	19.0	19.0	31.1	24.3	16.2	16.2	31.1	16.2	14.9	17.5	17.5	31.1
		0.7	46.8	33.8	33.8	39.8	36.6	33.8	31.1	39.8	36.6	50.8	43.2	22.4	24.3	22.4	39.8
		0.9	50.8	50.8	50.8	50.8	50.8	50.8	50.8	50.8	50.8	50.8	50.8	50.8	50.8	50.8	50.8
	0.43 , 0.43	0.1	39.8	31.1	28.7	19.0	19.0	31.1	24.3	16.2	16.2	31.1	16.2	14.9	17.5	17.5	31.1
		0.3	39.8	31.1	28.7	17.5	17.5	26.4	24.3	14.9	14.9	31.1	14.9	4.0	17.5	17.5	31.1
		0.5	36.6	31.1	26.4	17.5	17.5	26.4	24.3	14.9	14.9	28.7	14.9	13.7	17.5	16.2	31.1
		0.7	33.8	28.7	26.4	17.5	16.2	26.4	24.3	14.9	13.7	28.7	14.9	4.0	17.5	16.2	31.1
		0.9	39.8	31.1	31.1	33.8	33.8	31.1	24.3	16.2	17.5	50.8	16.2	16.2	19.0	17.5	33.8
	0.70 , 0.70	0.1	33.8	26.4	24.3	16.2	14.9	24.3	22.4	13.7	13.7	24.3	13.7	4.0	16.2	13.7	28.7
		0.3	33.8	26.4	24.3	16.2	16.2	24.3	22.4	14.9	13.7	26.4	14.9	13.7	17.5	14.9	28.7
		0.5	31.1	26.4	24.3	14.9	13.7	24.3	22.4	12.6	11.6	17.5	12.6	4.0	16.2	13.7	22.4
		0.7	33.8	31.1	28.7	33.8	31.1	28.7	24.3	17.5	17.5	28.7	17.5	16.2	19.0	19.0	31.1
		0.9	50.8	39.8	43.2	46.8	43.2	39.8	39.8	50.8	43.2	50.8	43.2	50.8	43.2	50.8	43.2
	0.57 , 0.58	0.1	36.6	28.7	26.4	17.5	17.5	26.4	24.3	14.9	14.9	28.7	14.9	4.0	17.5	16.2	31.1
		0.3	36.6	31.1	26.4	17.5	17.5	26.4	24.3	14.9	14.9	28.7	14.9	4.0	17.5	16.2	31.1
		0.5	33.8	26.4	24.3	16.2	16.2	24.3	22.4	13.7	13.7	24.3	13.7	4.0	16.2	14.9	28.7
		0.7	31.1	26.4	24.3	14.9	14.9	24.3	22.4	12.6	4.0	24.3	4.0	4.0	16.2	13.7	20.7
		0.9	31.1	26.4	24.3	14.9	13.7	22.4	22.4	12.6	11.6	16.2	12.6	5.6	16.2	13.7	20.7
	0.93 , 0.93	0.1	33.8	28.7	26.4	19.0	17.5	26.4	24.3	16.2	16.2	28.7	16.2	14.9	17.5	17.5	31.1
		0.3	33.8	28.7	26.4	19.0	17.5	26.4	24.3	16.2	16.2	28.7	16.2	14.9	17.5	17.5	31.1
		0.5	43.2	33.8	31.1	36.6	33.8	33.8	28.7	20.7	33.8	50.8	19.0	19.0	20.7	20.7	36.6
		0.7	50.8	39.8	39.8	46.8	39.8	36.6	36.6	50.8	39.8	50.8	43.2	50.8	43.2	50.8	43.2
		0.9	50.8	50.8	50.8	50.8	50.8	50.8	50.8	50.8	50.8	50.8	50.8	50.8	50.8	50.8	50.8
	mean		**39.2**	**32.2**	**30.4**	**26.8**	**25.3**	**30.3**	**27.4**	**22.1**	**21.2**	**35.1**	**21.1**	**18.1**	**22.6**	**22.5**	**33.1**
N100	0.36 , 0.35	0.1	20.7	17.5	14.9	31.1	17.5	19.0	12.6	17.5	19.0	26.4	13.7	4.0	9.1	43.2	14.9
		0.3	19.0	17.5	14.9	31.1	17.5	19.0	12.6	17.5	19.0	26.4	13.7	4.0	9.1	43.2	14.9
		0.5	19.0	16.2	14.9	31.1	17.5	19.0	13.7	16.2	19.0	26.4	13.7	26.4	9.1	43.2	14.9
		0.7	20.7	19.0	16.2	31.1	19.0	20.7	16.2	19.0	28.7	50.8	14.9	43.2	12.6	43.2	16.2
		0.9	43.2	31.1	33.8	50.8	33.8	39.8	39.8	50.8	36.6	50.8	24.3	50.8	43.2	50.8	39.8
	0.29 , 0.27	0.1	20.7	17.5	16.2	31.1	19.0	20.7	14.9	17.5	20.7	26.4	14.9	43.2	9.9	43.2	14.9
		0.3	20.7	17.5	16.2	31.1	19.0	20.7	16.2	19.0	26.4	28.7	14.9	43.2	11.6	43.2	16.2
		0.5	20.7	17.5	14.9	31.1	19.0	19.0	14.9	17.5	20.7	26.4	13.7	43.2	10.7	43.2	14.9
		0.7	31.1	26.4	28.7	36.6	31.1	33.8	36.6	43.2	36.6	50.8	20.7	46.8	20.7	50.8	20.7
		0.9	50.8	46.8	46.8	43.2	46.8	50.8	50.8	50.8	50.8	50.8	50.8	50.8	50.8	50.8	46.8
	0.43 , 0.43	0.1	20.7	17.5	14.9	31.1	19.0	19.0	13.7	17.5	20.7	26.4	4.0	4.0	9.1	43.2	14.9
		0.3	19.0	17.5	13.7	31.1	17.5	19.0	12.6	17.5	19.0	26.4	4.0	4.0	5.1	43.2	14.9
		0.5	19.0	16.2	13.7	31.1	17.5	17.5	11.6	16.2	19.0	26.4	4.0	4.0	7.1	9.1	13.7
		0.7	17.5	16.2	12.6	31.1	16.2	17.5	11.6	4.0	4.0	26.4	4.0	4.0	4.0	4.0	13.7
		0.9	22.4	19.0	17.5	31.1	20.7	22.4	16.2	19.0	28.7	50.8	14.9	43.2	11.6	43.2	16.2
	0.70 , 0.70	0.1	14.9	13.7	10.7	28.7	14.9	16.2	10.7	4.0	4.0	13.7	4.0	4.0	4.0	4.0	12.6
		0.3	14.9	13.7	11.6	26.4	14.9	16.2	11.6	13.7	17.5	13.7	5.1	24.3	7.1	7.7	12.6
		0.5	13.7	12.6	10.7	24.3	14.9	16.2	10.7	4.0	4.0	12.6	4.0	4.0	4.0	4.0	12.6
		0.7	19.0	17.5	14.9	28.7	17.5	19.0	16.2	17.5	20.7	26.4	14.9	43.2	11.6	43.2	14.9
		0.9	36.6	31.1	33.8	43.2	33.8	39.8	39.8	50.8	39.8	50.8	39.8	50.8	39.8	50.8	39.8
	0.57 , 0.58	0.1	17.5	16.2	12.6	28.7	16.2	17.5	11.6	16.2	17.5	24.3	4.0	4.0	5.1	7.7	13.7
		0.3	19.0	16.2	12.6	31.1	17.5	17.5	11.6	16.2	17.5	26.4	4.0	4.0	5.1	9.1	13.7
		0.5	16.2	14.9	11.6	28.7	14.9	16.2	10.7	4.0	4.0	24.3	4.0	4.0	4.0	4.0	12.6
		0.7	14.9	12.6	9.9	26.4	14.9	16.2	9.9	4.0	4.0	12.6	4.0	4.0	4.0	4.0	12.6
		0.9	13.7	12.6	9.9	20.7	13.7	14.9	10.7	5.1	4.0	11.6	4.0	4.0	4.0	4.0	12.6
	0.93 , 0.93	0.1	19.0	16.2	13.7	28.7	17.5	19.0	13.7	16.2	19.0	26.4	13.7	26.4	9.9	43.2	14.9
		0.3	17.5	16.2	13.7	28.7	16.2	19.0	13.7	16.2	19.0	24.3	13.7	26.4	9.9	39.8	14.9
		0.5	22.4	20.7	20.7	31.1	22.4	24.3	31.1	20.7	31.1	50.8	17.5	43.2	14.9	43.2	17.5
		0.7	31.1	26.4	28.7	36.6	31.1	36.6	36.6	50.8	36.6	50.8	24.3	50.8	36.6	50.8	24.3
		0.9	50.8	50.8	50.8	50.8	50.8	50.8	50.8	50.8	50.8	50.8	50.8	50.8	50.8	50.8	50.8
	mean		**22.9**	**20.2**	**18.5**	**32.2**	**21.4**	**23.2**	**19.5**	**21.1**	**22.0**	**31.3**	**14.5**	**25.3**	**14.5**	**32.1**	**18.9**

Table 3.2. Duration of the achromatic phase t_{achro} in rest-frame days for all 15 LSST color curves for the W7 and N100 model and 30 different microlensing magnification maps (κ, γ, and s) as in Sect. 3.4.1, where for each map 10000 random positions are drawn. The mean values are plotted in Fig. 3.6.

	κ, γ	s	u-g	u-r	u-i	u-z	u-y	g-r	g-i	g-z	g-y	r-i	r-z	r-y	i-z	i-y	z-y
sub-Ch	0.36 , 0.35	0.1	31.1	31.1	31.1	17.5	14.9	36.6	36.6	12.6	10.7	50.8	14.9	6.0	14.9	6.6	26.4
		0.3	31.1	28.7	31.1	16.2	13.7	36.6	36.6	11.6	9.9	50.8	14.9	6.0	14.9	6.0	24.3
		0.5	28.7	28.7	28.7	16.2	12.6	36.6	33.8	11.6	9.9	50.8	14.9	6.6	14.9	6.6	24.3
		0.7	31.1	31.1	31.1	36.6	31.1	36.6	36.6	14.9	13.7	50.8	16.2	14.9	17.5	12.6	26.4
		0.9	46.8	39.8	43.2	50.8	43.2	46.8	46.8	50.8	46.8	50.8	46.8	50.8	46.8	50.8	46.8
	0.29 , 0.27	0.1	31.1	31.1	31.1	19.0	17.5	36.6	36.6	12.6	11.6	50.8	16.2	7.7	16.2	7.7	26.4
		0.3	31.1	31.1	31.1	36.6	31.1	36.6	36.6	13.7	12.6	50.8	16.2	11.6	17.5	10.7	26.4
		0.5	31.1	31.1	31.1	20.7	17.5	36.6	36.6	12.6	11.6	50.8	16.2	9.9	17.5	9.1	26.4
		0.7	33.8	33.8	36.6	43.2	36.6	43.2	43.2	50.8	43.2	50.8	43.2	50.8	39.8	50.8	39.8
		0.9	50.8	50.8	50.8	50.8	50.8	50.8	50.8	50.8	50.8	50.8	50.8	50.8	50.8	50.8	50.8
	0.43 , 0.43	0.1	31.1	31.1	31.1	17.5	14.9	36.6	36.6	11.6	9.9	50.8	14.9	6.0	14.9	6.0	26.4
		0.3	31.1	31.1	31.1	16.2	12.6	36.6	36.6	11.6	9.9	50.8	4.0	5.6	6.0	6.0	24.3
		0.5	28.7	28.7	28.7	16.2	12.6	36.6	33.8	11.6	9.1	50.8	4.0	5.6	6.0	6.0	24.3
		0.7	28.7	28.7	28.7	14.9	9.9	36.6	33.8	9.1	8.4	46.8	4.0	4.4	4.0	5.6	20.7
		0.9	31.1	31.1	31.1	36.6	31.1	36.6	36.6	13.7	12.6	50.8	16.2	12.6	17.5	11.6	26.4
	0.70 , 0.70	0.1	26.4	26.4	26.4	12.6	8.4	12.6	28.7	6.0	6.6	28.7	4.0	4.4	4.0	4.4	19.0
		0.3	26.4	28.7	28.7	14.9	10.7	13.7	31.1	9.9	9.1	31.1	6.0	6.0	6.6	6.0	20.7
		0.5	26.4	26.4	26.4	11.6	8.4	12.6	16.2	6.6	7.7	26.4	4.0	4.4	4.0	5.6	17.5
		0.7	28.7	28.7	28.7	33.8	31.1	36.6	36.6	14.9	14.9	46.8	17.5	14.9	17.5	14.9	26.4
		0.9	39.8	36.6	39.8	50.8	39.8	46.8	46.8	50.8	46.8	50.8	46.8	50.8	46.8	50.8	43.2
	0.57 , 0.58	0.1	28.7	28.7	28.7	16.2	11.6	36.6	33.8	9.9	8.4	46.8	4.0	4.4	6.0	5.6	22.4
		0.3	28.7	28.7	28.7	16.2	11.6	36.6	33.8	11.6	9.1	46.8	4.0	5.6	6.0	6.0	24.3
		0.5	28.7	28.7	28.7	12.6	9.1	13.7	31.1	8.4	7.7	46.8	4.0	4.4	4.0	5.6	19.0
		0.7	26.4	26.4	26.4	11.6	7.7	12.6	16.2	5.6	6.0	26.4	4.0	4.4	4.0	4.4	17.5
		0.9	24.3	13.7	26.4	11.6	9.1	12.6	16.2	7.7	7.7	26.4	4.0	5.6	5.6	5.6	17.5
	0.93 , 0.93	0.1	28.7	28.7	28.7	17.5	14.9	36.6	33.8	12.6	11.6	46.8	16.2	10.7	14.9	9.9	24.3
		0.3	28.7	28.7	28.7	16.2	13.7	36.6	33.8	11.6	9.9	46.8	14.9	7.7	14.9	7.7	24.3
		0.5	31.1	31.1	33.8	36.6	33.8	39.8	39.8	19.0	39.8	50.8	20.7	19.0	20.7	19.0	31.1
		0.7	36.6	36.6	36.6	46.8	36.6	43.2	43.2	50.8	43.2	50.8	43.2	50.8	43.2	50.8	39.8
		0.9	50.8	50.8	50.8	50.8	50.8	50.8	50.8	50.8	50.8	50.8	50.8	50.8	50.8	50.8	50.8
	mean		**31.9**	**31.2**	**32.1**	**25.6**	**21.6**	**34.1**	**35.4**	**19.2**	**18.3**	**46.1**	**17.9**	**16.4**	**18.3**	**16.5**	**27.9**
merger	0.36 , 0.35	0.1	43.2	36.6	36.6	5.1	4.0	39.8	39.8	4.0	4.0	50.8	4.0	4.0	5.6	4.7	24.3
		0.3	43.2	36.6	36.6	17.5	4.7	39.8	39.8	4.0	4.0	50.8	4.4	4.0	5.6	4.7	24.3
		0.5	39.8	33.8	33.8	14.9	4.7	36.6	36.6	4.7	4.7	50.8	4.7	4.7	5.6	4.7	24.3
		0.7	43.2	36.6	39.8	22.4	43.2	39.8	39.8	13.7	8.4	50.8	14.9	8.4	17.5	17.5	28.7
		0.9	50.8	46.8	46.8	50.8	50.8	50.8	50.8	50.8	50.8	50.8	43.2	50.8	43.2	50.8	43.2
	0.29 , 0.27	0.1	43.2	36.6	36.6	17.5	8.4	39.8	39.8	5.1	4.7	50.8	5.1	4.7	13.7	5.6	28.7
		0.3	43.2	36.6	36.6	19.0	43.2	39.8	39.8	6.0	5.6	50.8	6.0	5.6	14.9	7.7	28.7
		0.5	43.2	36.6	36.6	19.0	14.9	39.8	39.8	6.0	5.6	50.8	6.0	5.6	13.7	8.4	24.3
		0.7	50.8	43.2	43.2	28.7	46.8	46.8	43.2	24.3	46.8	50.8	22.4	50.8	24.3	50.8	39.8
		0.9	50.8	50.8	50.8	50.8	50.8	50.8	50.8	50.8	50.8	50.8	50.8	50.8	50.8	50.8	50.8
	0.43 , 0.43	0.1	43.2	39.8	39.8	4.4	4.0	39.8	39.8	4.0	4.0	50.8	4.0	4.0	4.7	4.7	28.7
		0.3	43.2	36.6	36.6	4.0	4.0	39.8	39.8	4.0	4.0	50.8	4.0	4.0	4.7	4.7	24.3
		0.5	43.2	36.6	36.6	4.4	4.0	39.8	36.6	4.0	4.0	50.8	4.0	4.0	4.7	4.7	24.3
		0.7	39.8	33.8	33.8	4.0	4.0	36.6	36.6	4.0	4.0	50.8	4.0	4.0	4.7	4.0	24.3
		0.9	46.8	39.8	39.8	22.4	43.2	43.2	39.8	13.7	9.9	50.8	14.9	9.1	17.5	19.0	33.8
	0.70 , 0.70	0.1	22.4	28.7	28.7	4.0	4.0	31.1	33.8	4.0	4.0	46.8	4.0	4.0	4.7	4.0	22.4
		0.3	22.4	28.7	28.7	10.7	5.1	31.1	33.8	4.7	4.7	46.8	5.1	4.7	5.6	4.7	22.4
		0.5	20.7	22.4	24.3	4.4	4.0	28.7	31.1	4.0	4.0	46.8	4.0	4.0	4.7	4.7	20.7
		0.7	31.1	33.8	33.8	19.0	20.7	36.6	36.6	11.6	8.4	50.8	11.6	8.4	14.9	11.6	24.3
		0.9	50.8	46.8	46.8	50.8	46.8	50.8	46.8	50.8	50.8	50.8	43.2	50.8	43.2	50.8	43.2
	0.57 , 0.58	0.1	36.6	33.8	33.8	4.0	4.0	36.6	36.6	4.0	4.0	50.8	4.0	4.0	4.7	4.0	24.3
		0.3	39.8	33.8	33.8	4.0	4.0	39.8	36.6	4.0	4.0	50.8	4.0	4.0	4.7	4.7	24.3
		0.5	24.3	31.1	31.1	4.0	4.0	33.8	33.8	4.0	4.0	50.8	4.0	4.0	4.7	4.0	22.4
		0.7	20.7	24.3	26.4	4.0	4.0	28.7	33.8	4.0	4.0	46.8	4.0	4.0	4.7	4.0	20.7
		0.9	19.0	20.7	20.7	4.4	4.0	26.4	31.1	4.0	4.0	43.2	4.0	4.0	4.7	4.7	20.7
	0.93 , 0.93	0.1	31.1	33.8	33.8	17.5	13.7	36.6	36.6	7.7	7.1	50.8	7.7	6.6	13.7	8.4	24.3
		0.3	26.4	31.1	31.1	16.2	9.9	33.8	36.6	6.0	5.6	50.8	6.0	5.6	12.6	7.1	22.4
		0.5	46.8	39.8	39.8	33.8	43.2	43.2	39.8	19.0	43.2	50.8	20.7	22.4	20.7	26.4	33.8
		0.7	46.8	43.2	43.2	50.8	46.8	46.8	46.8	50.8	46.8	50.8	39.8	50.8	39.8	50.8	43.2
		0.9	50.8	50.8	50.8	50.8	50.8	50.8	50.8	50.8	50.8	50.8	50.8	50.8	50.8	50.8	50.8
	mean		**38.6**	**36.1**	**36.4**	**18.8**	**19.9**	**39.2**	**39.2**	**14.3**	**15.2**	**50.0**	**13.5**	**14.8**	**15.5**	**16.1**	**29.1**

Table 3.3. Duration of the achromatic phase t_{achro} in rest-frame days for all 15 LSST color curves for the sub-Ch and the merger model and 30 different microlensing magnification maps (κ, γ, and s) as in Sect. 3.4.1, where for each map 10000 random positions are drawn. The mean values are plotted in Fig. 3.6.

3.6.3 Additional color curves

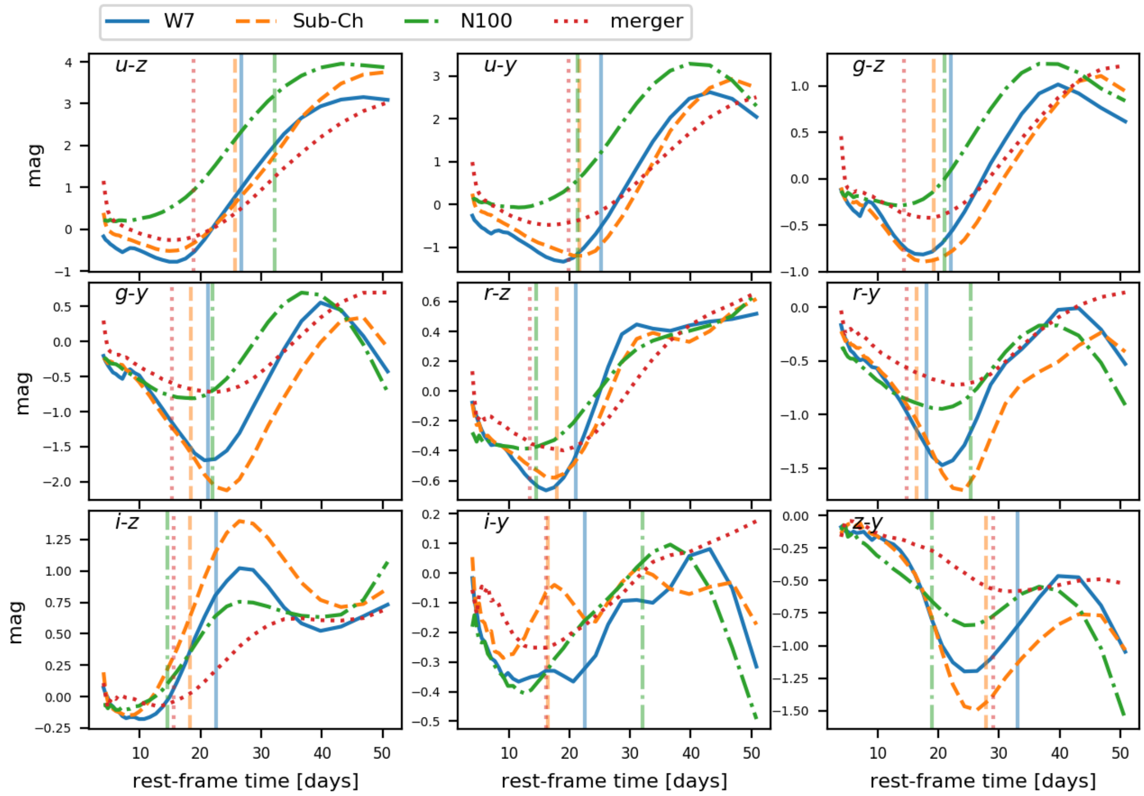


Fig. 3.13. Rest-frame LSST color curves without microlensing for four different SN Ia models similar as in Fig. 3.6, but this time showing the remaining nine color curves that are not as useful for time-delay measurements.

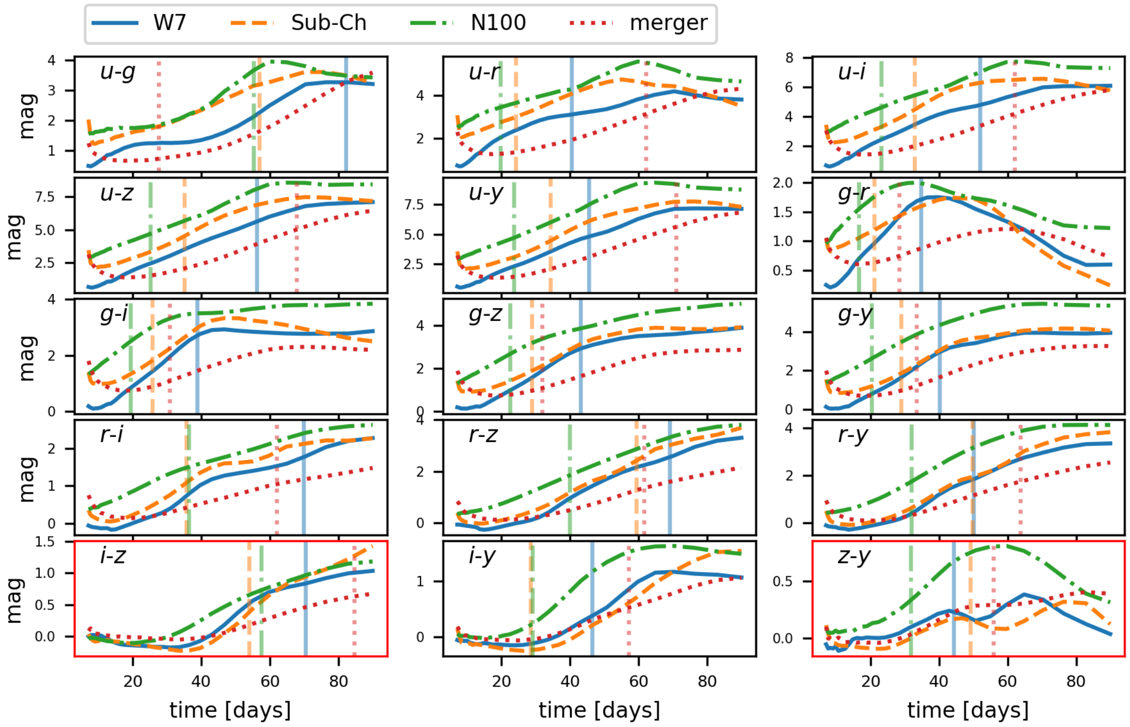


Fig. 3.14. All redshifted LSST color curves without microlensing for four different SN Ia models similar to Fig. 3.10 but at $z_s = 0.77$.

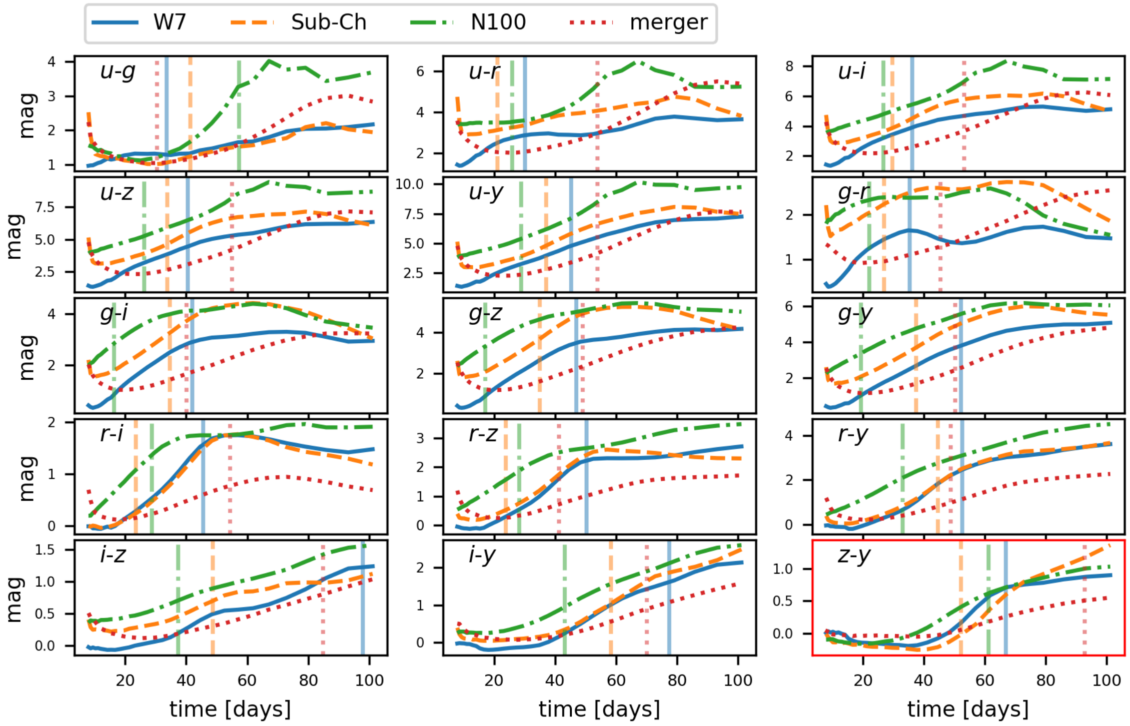


Fig. 3.15. All redshifted LSST color curves without microlensing for four different SN Ia models similar to Fig. 3.10 but at $z_s = 0.99$.

4 Time-delay measurement of strongly lensed Type Ia supernovae using fully connected neural networks and random forests

In Sect. 1.2, we mentioned already that we expect from the 10-year LSST survey, 10 to 25 LSNe Ia[23], with accurate and precise time-delays (Huber et al., 2019). This estimate assumes follow-up observation with an LSST-like 5σ depth and a two-day cadence. Going one magnitude deeper in follow-up monitoring improves the number of LSNe Ia with well-measured time delays by roughly a factor of 1.5 (Huber et al., 2019). The large spread in the prediction is related to different potential LSST observing strategies under consideration. However, many other science cases depend on the LSST observing strategy (Lochner et al., 2018; Lochner et al., 2022), leaving only limited room for specific optimization for LSNe Ia. Therefore, it is necessary to improve the time-delay measurement methods, which is the scope of this and the next chapter.

To measure the time delays in Huber et al. (2019), we used the free-knot spline estimator from Python Curve Shifting (PyCS; Tewes et al., 2013; Bonvin et al., 2016), and therefore, the characteristic light-curve shape of a SN Ia is not taken into account. Another time-delay measurement tool was developed by Pierel & Rodney (2019), which uses SN templates from empirical models. However, in contrast to this thesis, microlensing on LSNe Ia is approximated as wavelength-independent by Pierel & Rodney (2019) and, thus, per definition, achromatic. Goldstein et al. (2018) suggested *color* curves to reduce microlensing uncertainties in the first three rest-frame weeks after the explosion. Although we confirmed the result in Chapter 3, in practice, one faces the issue that color curves with robust features are limited for typical SN Ia redshifts. Therefore, we investigate in the following SN Ia *light* curves to measure time delays in LSNe Ia.

We use two simple machine learning approaches; a fully connected neural network (FCNN) and a random forest (RF, Breiman, 2001). The disadvantage of these two techniques is that they always require the same amount of input data points. Therefore, it is necessary to train a machine learning model for each observed LSNe Ia separately. However, the advantage of this approach is that our models are trained very specifically for a certain observation pattern, noise, and microlensing uncertainties, so accurate and precise time-delay measurements are expected. In this work, we test the machine learning models on a few example LSNe Ia, where our main mock system is a double LSN Ia at a source redshift of 0.76 close to the median source redshift of the OM10 catalog. We assume follow-up observations with a cadence of two days with a few random gaps and a 5σ point-source depth of 25.7, 25.3, 24.7, and 23.8 for the bands g, r, i, and z (which is one magnitude deeper than that of the LSST single-epoch depth). For our training process of the two machine learning models, we use four theoretical SN Ia models to include the effect of microlensing, as described in Sect. 2.3. While both approaches yield unbiased results on a test set based on the

---

[23] This approach considered only bright and spatially resolved LSNe Ia as in the OM10 catalog (Oguri & Marshall, 2010).

theoretical models used in the training process, the FCNN shows biases on a data set based on the empirical `SNEMO15` SN Ia light curves (Saunders et al., 2018). However, the RF is still bias-free and can be used for real applications.

We achieve a ~1.5-day uncertainty in the i band with the RF for our specific mock observation. If light curves from three filters are available, where a RF is trained for each observation separately, we achieve an uncertainty of ~1.0 day. From the investigation of different source redshifts z_s, we find that for $z_\mathrm{s} \lesssim 0.6$, the *gri* bands are most promising, and for higher source redshifts, the *riz* bands are preferred. Further investigations show that there is no gain in training a single RF for a quad LSNe Ia instead of a separate RF per pair of images, which means treating it like a double system. Additionally, we investigate different detection scenarios of LSNe Ia where we find that if observations are available 8 to 10 days before the peak of the first image, a time delay longer than 15 days is enough to ensure accuracy of the delay measurement better than 1%. The precision is around ~1.5 days and drops to ~1.8 days for a later detection around four days before the peak. We also investigated these cases using `PyCS`, where the precision drops from ~1.7 days to ~2.6 days, and therefore, the RF outperforms `PyCS`. However, if no data points are available before the peak, `PyCS` has a lower bias than the RF, although taking the precision into account, both measurements are insufficient. Therefore, we aim for a LSNe detection before the peak. Further, our results show that even though microlensing is responsible for roughly a factor of two in the uncertainty of the measurement, the most dominant error source is the observational noise, which makes the depth of the follow-up observations an essential factor.

The rest of this chapter contains our publication Huber et al. (2022) with all the details related to this summary. I am the main author of this paper. I simulated all data sets for the training and testing process of the machine-learning approaches. Further, I created the setup of both machine learning models, carried out the training process, and did all the analysis described in the paper. I made all figures and tables. All co-authors provided feedback for the final manuscript.

HOLISMOKES - VII. Time-delay measurement of strongly lensed Type Ia supernovae using machine learning

S. Huber, S. H. Suyu, D. Ghoshdastidar, S. Taubenberger, V. Bonvin, J. H. H. Chan, M. Kromer, U. M. Noebauer, S. A. Sim, and L. Leal-Taixe

ABSTRACT

The Hubble constant (H_0) is one of the fundamental parameters in cosmology, but there is a heated debate around the >4σ tension between the local Cepheid distance ladder and the early Universe measurements. Strongly lensed Type Ia supernovae (LSNe Ia) are an independent and direct way to measure H_0, where a time-delay measurement between the multiple supernova (SN) images is required. In this work, we present two machine learning approaches for measuring time delays in LSNe Ia, namely, a fully connected neural network (FCNN) and a random forest (RF). For the training of the FCNN and the RF, we simulate mock LSNe Ia from theoretical SN Ia models that include observational noise and microlensing. We test the generalizability of the machine learning models by using a final test set based on empirical LSN Ia light curves not used in the training process, and we find that only the RF provides a low enough bias to achieve precision cosmology; as such, RF is therefore preferred over our FCNN approach for applications to real systems. For the RF with single-band photometry in the *i* band, we obtain an accuracy better than 1% in all investigated cases for time delays longer than 15 days, assuming follow-up observations with a 5σ point-source depth of 24.7, a two day cadence with a few random gaps, and a detection of the LSNe Ia 8 to 10 days before peak in the observer frame. In terms of precision, we can achieve an approximately 1.5-day uncertainty for a typical source redshift of ~0.8 on the *i* band under the same assumptions. To improve the measurement, we find that using three bands, where we train a RF for each band separately and combine them afterward, helps to reduce the uncertainty to ~1.0 day. The dominant source of uncertainty is the observational noise, and therefore the depth is an especially important factor when follow-up observations are triggered. We have publicly released the microlensed spectra and light curves used in this work.

Credit: Huber et al., A&A 658, A157, 2022, published in A&A with Open Access MPI agreement ©Huber.

4.1 Introduction

The Hubble constant, H_0, is one of the fundamental parameters in cosmology, but there is a tension[24] of $>4\sigma$ (Verde et al., 2019) between the early Universe measurements inferred from the cosmic microwave background (CMB; Planck Collaboration, 2020) and late Universe measurements from the Supernova H_0 for the Equation of State (SH0ES) project (e.g., Riess et al., 2016, 2018, 2019, 2021). However, results from Freedman et al. (2019, 2020) using the tip of the red giant branch (TRGB) or from Khetan et al. (2021) using surface brightness fluctuations (SBFs) are consistent with both. An independent analysis using the TRGBs by Anand et al. (2022) has derived a slightly higher H_0 value, bringing it closer to the results of the SH0ES project, which is based on Cepheids. Moreover, recent results from Blakeslee et al. (2021) using SBFs that are calibrated through both Cepheids and TRGBs are in good agreement with the SH0ES project and $\sim 2\sigma$ higher than the CMB values. As an alternative to the distance ladder approach, Pesce et al. (2020) measured H_0 from the Megamaser Cosmology Project, which also agrees well with the SH0ES results and is about $\sim 2\sigma$ higher than the Planck value. Gravitational wave sources acting as standard sirens also provide direct luminosity distances and thus H_0 measurements (e.g., Abbott et al., 2017). While this is a promising approach, current uncertainties on H_0 from standard sirens preclude them from being used to discern between the SH0ES and the CMB results.

Lensing time-delay cosmography, as an independent probe, can address this tension by measuring H_0 in a single step. This method, first envisaged by Refsdal (1964), combines the measured time delay from the multiple images of a variable source with lens mass modeling and line-of-sight mass structure to infer H_0. The COSmological MOnitoring of GRAvItational Lenses (COSMOGRAIL; Courbin et al., 2018) and H0 Lenses in COSMOGRAIL's Wellspring (H0LiCOW; Suyu et al., 2017) collaborations, together with the Strong lensing at High Angular Resolution Program (SHARP) (Chen et al., 2019), have successfully applied this method to lensed quasar systems (e.g., Bonvin et al., 2018; Birrer et al., 2019; Sluse et al., 2019; Rusu et al., 2019; Chen et al., 2019). The latest H_0 measurement from H0LiCOW using physically motivated mass models is consistent with measurements from SH0ES but is in $>3\sigma$ tension with results from the CMB (Wong et al., 2020). The STRong-lensing Insights into the Dark Energy Survey (STRIDES) collaboration has further analyzed a new lensed quasar system (Shajib et al., 2020). The newly formed Time-Delay COSMOgraph (TDCOSMO) organization (Millon et al., 2020), consisting of H0LiCOW, COSMOGRAIL, SHARP and STRIDES, has recently considered a one-parameter extension to the mass model to allow for the mass-sheet transformation (e.g., Falco et al., 1985; Schneider & Sluse, 2013; Kochanek, 2020). Birrer et al. (2020) used the stellar kinematics to constrain this single parameter, resulting in an H_0 value with a larger uncertainty, which is statistically consistent with the previous results using physically motivated mass models. In addition to placing constraints on H_0, strongly lensed quasars also provide tests of the cosmological principle, especially of spatial isotropy, given the independent sight line and

---

[24] `https://github.com/shsuyu/H0LiCOW-public/tree/master/H0_tension_plots` (Bonvin & Millon, 2020)

distance measurement that each lensed quasar yields (e.g., Krishnan et al., 2021; Krishnan et al., 2022).

In addition to strongly lensed quasars, supernovae (SNe) lensed into multiple images are promising as a cosmological probe and are in fact the sources envisioned by Refsdal (1964). Even though these systems are much rarer in comparison to quasars, they have the advantage that SNe fade away over time, facilitating measurements of stellar kinematics of the lens galaxy (Barnabè et al., 2011a; Yıldırım et al., 2017; Shajib et al., 2018; Yıldırım et al., 2020) and surface brightness distributions of the lensed-SN host galaxy (Ding et al., 2021) to break model degeneracies, for example, the mass-sheet transformation (Falco et al., 1985; Schneider & Sluse, 2014). Furthermore, strongly lensed type Ia supernovae (LSNe Ia) are promising given that they are standardizable candles and therefore provide an additional way to break model degeneracies for lens systems where lensing magnifications are well characterized (Oguri & Kawano, 2003; Foxley-Marrable et al., 2018).

So far, only three LSNe with resolved multiple images have been observed, namely SN "Refsdal" (Kelly et al., 2016a,b), a core-collapse SN at a redshift of $z = 1.49$, the LSN Ia iPTF16geu (Goobar et al., 2017) at $z = 0.409$, and AT2016jka (Rodney et al., 2021) at $z = 1.95$, which is most likely a LSN Ia. Nonetheless, with the upcoming Rubin Observatory Legacy Survey of Space and Time (LSST; Ivezić et al., 2019), we expect to find $\sim 10^3$ LSNe, of which 500 to 900 are expected to be type Ia SNe (Quimby et al., 2014; Goldstein & Nugent, 2017; Goldstein et al., 2018; Wojtak et al., 2019). Considering only LSNe Ia with spatially resolved images and peak brightnesses[25] brighter than 22.6 in the i band, as in the Oguri & Marshall (2010, hereafter OM10) lens catalog, leads to 40 to 100 LSNe Ia, depending on the LSST observing strategy, of which 10 to 25 systems yield accurate time-delay measurements (Huber et al., 2019).

To measure time delays between multiple images of LSNe Ia, Huber et al. (2019) used the free-knot spline estimator from Python Curve Shifting (`PyCS`; Tewes et al., 2013; Bonvin et al., 2016), and therefore the characteristic light-curve shape of a SN Ia is not taken into account. Furthermore, they do not explicitly model the variability due to microlensing (Chang & Refsdal, 1979; Irwin et al., 1989; Wambsganss, 2006; Mediavilla et al., 2016), an effect where each SN image is separately influenced by lensing effects from stars in the lens, leading to the additional magnification and distortion of light curves (Yahalomi et al., 2017; Goldstein et al., 2018; Foxley-Marrable et al., 2018; Huber et al., 2019; Pierel & Rodney, 2019; Huber et al., 2021). While PyCS has the advantage of being flexible without making assumptions on the light-curve forms, model-based methods are complementary in providing additional information to measure the time delays more precisely.

One such model-based time-delay measurement method was implemented by Pierel & Rodney (2019), where template SN light curves are used. Even though microlensing is taken into account in this work, it is done in the same way for each filter. A more realistic microlensing treatment for SNe Ia, with variations in the SN intensity distribution across wavelengths, was first introduced by Goldstein et al. (2018) using specific intensity profiles

---

[25] of the fainter image for a double system; for a quad system, the peak brightness of the third brightest image is considered.

from the theoretical W7 model (Nomoto et al., 1984) calculated via the radiative transfer code `SEDONA` (Kasen et al., 2006). Huber et al. (2019, 2021) have built upon this study, but using the radiative transfer code `ARTIS` (Kromer & Sim, 2009) to calculate synthetic observables for up to four theoretical SN explosion models. In this work, we follow the approach of Huber et al. (2019, 2021) to calculate realistic microlensed light curves for LSNe Ia to train a fully connected neural network (FCNN) and a random forest (RF) model for measuring time delays. In addition, this method also allows us to identify dominant sources of uncertainties and quantify different follow-up strategies.

This paper is organized as follows. In Sect. 4.2 we present our calculation of mock light curves, which includes microlensing and observational uncertainties. The creation of our training, validation, and test sets is explained with an example mock observation in Sect. 4.3, followed by an introduction to the machine learning (ML) techniques used in this work in Sect. 4.4. We apply these methods to the example mock observation in Sect. 4.5, where we also test the generalizability by using empirical LSN Ia light curves not used in the training process. In Sect. 4.6 we investigate, based on our example mock observation, potential filters for follow-up observations and the impact of microlensing and noise on the uncertainty, before we investigate more mock observations in Sect. 4.7. We discuss our results in Sect. 4.8 before summarizing in Sect. 4.9. Magnitudes in this paper are in the AB system.

We have publicly released the microlensed spectra and light curves used in this work at `https://github.com/shsuyu/HOLISMOKES-public/tree/main/HOLISMOKES_VII`.

4.2 Simulated light curves for LSNe Ia

The goal is to develop a software that takes photometric light-curve observations of a LSN Ia as input and predicts as an output the time delay between the different images. For a ML approach, we need to simulate a realistic data set where we account for different sources of uncertainties. We therefore specify in Sect. 4.2.1 our calculation of microlensing, and we explain in Sect. 4.2.2 our determination of observational uncertainties including estimates of the moon phase.

4.2.1 Microlensing and SN Ia models

To calculate light curves for a LSN Ia with microlensing, we combine magnifications maps from `GERLUMPH` (Vernardos & Fluke, 2014; Vernardos et al., 2014, 2015) with theoretical SN Ia models, where synthetic observables have been calculated with `ARTIS` (Kromer & Sim, 2009). The basic idea is to place a SN in a magnification map and solve for the observed flux:

$$F_{\lambda,\mathrm{o}}(t) = \frac{1}{{D_\mathrm{lum}}^2(1+z_\mathrm{s})} \int \mathrm{d}x \int \mathrm{d}y \, I_{\lambda,\mathrm{e}}(t, p(x,y))\,\mu(x,y), \tag{4.1}$$

where D_lum is the luminosity distance to the source, z_s is the redshift of the source, $\mu(x,y)$ is the magnification factor depending on the positions (x,y) in the magnification map, and

$I_{\lambda,\mathrm{e}}(t,p)$ is the emitted specific intensity at the source plane as a function of wavelength, λ, time since explosion, t, and impact parameter, p (i.e., the projected distance from the ejecta center, where we assume spherical symmetry similar to Huber et al. 2019, 2021). Lensing magnification maps depend on three main parameters, namely the convergence κ, the shear γ and the smooth matter fraction $s = 1 - \kappa_*/\kappa$, where κ_* is the convergence of the stellar component. Further, our maps have a resolution of 20000×20000 pixels with a total size of $20\ R_{\mathrm{Ein}} \times 20\ R_{\mathrm{Ein}}$, where the Einstein radius, R_{Ein}, is a characteristic size of the map that depends on the source redshift z_{s}, lens redshift z_{d}, and masses of the microlenses. As in Huber et al. (2021), we follow Chan et al. (2021) for generating the microlensing magnification maps and assume a Salpeter initial mass function (IMF) with a mean mass of the microlenses of $0.35 M_\odot$; the specifics of the assumed IMF have negligible impact on our studies. From the flux we obtain the AB magnitudes via

$$m_{\mathrm{AB,X}}(t) = -2.5 \log_{10}\left(\frac{\int \mathrm{d}\lambda\, \lambda S_{\mathrm{X}}(\lambda)\, F_{\lambda,\mathrm{o}}(t)}{\int \mathrm{d}\lambda\, S_{\mathrm{X}}(\lambda)\, c\, \lambda^{-1}} \times \mathrm{cm}^2\, \mathrm{erg}^{-1} \right) - 48.6 \tag{4.2}$$

(Bessell & Murphy, 2012), where c is the speed of light and $S_{\mathrm{X}}(\lambda)$ is the transmission function for the filter X (that can be u, g, r, i, z, y, J, or H in this work). This calculation is discussed in much greater detail by Huber et al. (2019), which was initially motivated by the work of Goldstein et al. (2018).

The calculation of microlensing of LSNe Ia requires a theoretical model for the SN Ia that predicts the specific intensity. To increase the variety of different light-curve shapes we use four SNe Ia models computed with `ARTIS` (Kromer & Sim, 2009). These models have also been used in Suyu et al. (2020) and Huber et al. (2021), and are briefly summarized in the following: i) the parameterized 1D deflagration model W7 (Nomoto et al., 1984) with a Chandrasekhar mass (M_{Ch}) carbon-oxygen (CO) white dwarf (WD), ii) the delayed detonation model N100 (Seitenzahl et al., 2013) of a M_{Ch} CO WD, iii) a sub-Chandrasekhar (sub-Ch) detonation model of a CO WD with $1.06 M_\odot$ (Sim et al., 2010), and iv) a merger model of two CO WDs of $0.9 M_\odot$ and $1.1 M_\odot$ (Pakmor et al., 2012).

Fig. 4.1 shows the light curves for the four SN Ia models in comparison to the empirical `SNEMO15` model (Saunders et al., 2018). The light curves are normalized by the peak. Magnitude differences between SN Ia models are within 1 magnitude. To produce the median and 2σ (97.5th percentile − 2.5th percentile) light curves of `SNEMO15`, we consider all 171 SNe Ia from Saunders et al. (2018). Data of the empirical models cover only 3305 Å to 8586 Å and therefore the u band, starting at 3200 Å, is only an approximation, but an accurate one since the filter transmission in the missing region is low. The rest-frame u and g cover approximately the observed r, i, and z bands for a system with redshift of 0.76, which we investigate in Sects. 4.3 and 4.5. Light curves from theoretical and empirical models show the same evolution, although there are quite some differences in the shapes. The variety of different theoretical models is helpful to encapsulate the intrinsic variation of real SNe Ia. In building our training, validation and test sets for our ML methods, we also normalize the light curves after the calculation of the observational noise, which we describe next.

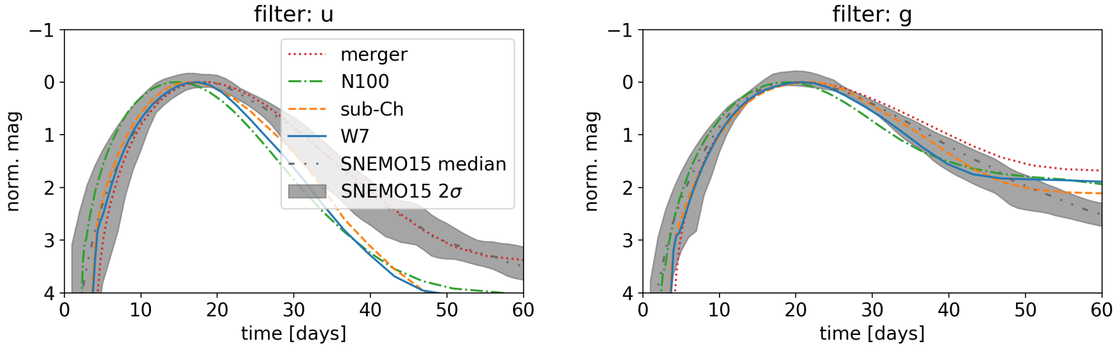


Fig. 4.1. Normalized LSST *u*- and *g*-band rest-frame light curves for four theoretical SN Ia models (merger, N100, sub-Ch, and W7) in comparison to the empirical model `SNEMO15`.

4.2.2 Observational uncertainty and the moon phase

Magnitudes for filter X including observational uncertainties can be calculated via

$$m_{\mathrm{data,X}} = m_{\mathrm{AB,X}} + r_{\mathrm{norm}}\sigma_{1,\mathrm{X}}, \tag{4.3}$$

where $m_{\mathrm{AB,X}}$ is the intrinsic magnitude without observational noise, r_{norm} is a random Gaussian number with a standard deviation of unity, and $\sigma_{1,\mathrm{X}}$ is a quantity that depends mainly on $m_{\mathrm{AB,X}}$ relative to the 5σ depth m_5. This calculation is based on LSST Science Collaboration (2009) (for more details, see also Appendix 4.10.1).

In order to calculate $m_{\mathrm{data,X}}$, the 5σ depth of the corresponding filter X is needed. In this work we consider eight filters, namely the six LSST filters, *u, g, r, i, z,* and *y*, as well as two infrared bands, *J* and *H*. To estimate the moon phase dependence of filter X, we used the exposure time calculator (ETC) of the European Southern Observatory (ESO) with a flat template spectrum. For *ugriz* we used the ETC of OmegaCAM[26], and for *yJH* we used the ETC of VIRCAM[27], where we assume an airmass of 1.2. Further, we used the typical fixed sky model parameters with seeing $\leq 1''$ as provided by the ETC, which we found to be a conservative estimate of the 5σ depth by testing other sky positions. We investigated one cycle phase (25 August 2020 to 24 September 2020) to obtain relative changes of the 5σ depth with time and matched these relative changes to the typical mean of the single-epoch LSST-like 5σ depth plus one magnitude, given by (23.3+1, 24.7+1, 24.3+1, 23.7+1, 22.8+1, and 22.0+1) for (*u, g, r, i, z,* and *y*), respectively, assuming a fixed exposure time. These mean values take into account that in typical LSST observing strategies, redder bands are preferred around full moon, while bluer bands are used more around new moon. Going one magnitude deeper than the LSST 5σ depth provides a better quality of photometric measurements for time-delay measurements, and is feasible even for a 2 m telescope (Huber et al., 2019). The absolute values for *J* and *H* bands are set by the ETC of VIRCAM in comparison to the *y* band.

---

[26] `https://www.eso.org/observing/etc/bin/gen/form?INS.NAME=OMEGACAM+INS.MODE=imaging`

[27] `https://www.eso.org/observing/etc/bin/gen/form?INS.NAME=VIRCAM+INS.MODE=imaging`

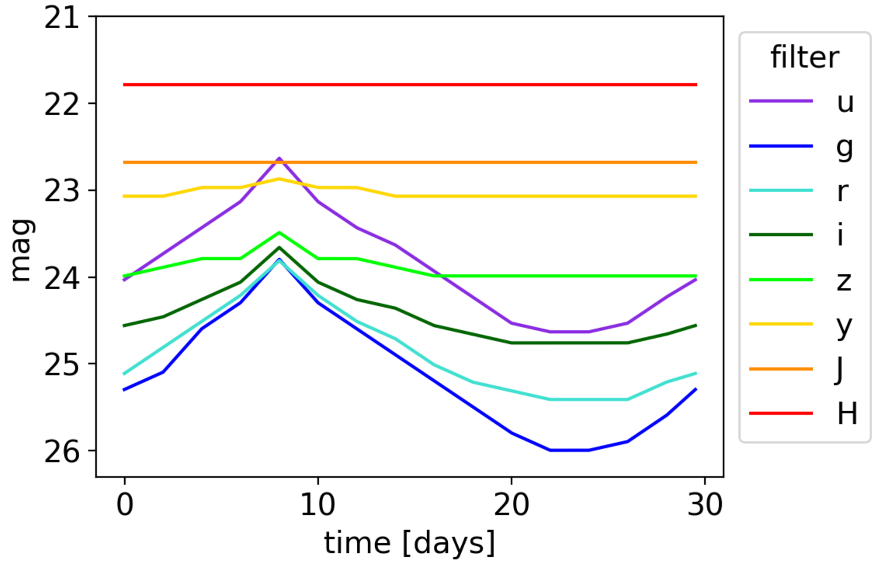


Fig. 4.2. Estimated 5σ depth for eight different filters, *u, g, r, i, z, y, J,* and *H,* accounting for the moon phase. Day 0 corresponds to the first quarter in the moon phase. Full moon is around day 8, and new moon is on day 23.

The results for one cycle phase are shown in Fig. 4.2, where we find full moon around day 8 and new moon around day 23. As expected, bluer bands are much more influenced by the moon phase in comparison to redder bands. As we are typically interested in getting LSNe Ia with time delays greater than 20 days (Huber et al., 2019), it is important to take the moon phase into account. Furthermore, we note that our approach on the 5σ depth assumes an isolated point source, where in reality we also have contributions from the host and lens light, which are the lowest for faint hosts and large image separations. Even though these are the systems we are interested in targeting, our uncertainties are on the optimistic side. The construction of light curves in the presence of the lens and host is deferred to future work, although LSNe have the advantage that the SNe fade away and afterward an observation of the lensing system without the SN can be taken and used as a template for subtraction.

4.3 Example mock observation and data set for machine learning

In this section we present a specific mock observation as an example, to explain the data structure required for our ML approaches.

4.3.1 Mock observation

As an example, we take a LSN Ia double system of the OM10 catalog (Oguri & Marshall, 2010), which is a mock lens catalog for strongly lensed quasars and SNe. The parameters of the mock LSN Ia are given in Table 4.1, where we have picked a system with a source redshift close to the median source redshift $z_s = 0.77$ of LSNe Ia in OM10 (Huber et al., 2021). The corresponding mock light curves are produced assuming the W7 model, where the *i* band is shown in Fig. 4.3 and all bands (*ugrizyJH*) together are shown in Appendix

Table 4.1. Mock system of the OM10 catalog for generating mock light curves to train our ML techniques.

z_s	z_d	image 1 (κ,γ)	image 2 (κ,γ)	time delay [days]
0.76	0.252	(0.251, 0.275)	(0.825, 0.815)	32.3

Notes. We assume $s = 0.6$, similar to Huber et al. (2021). The image separation for this double system is 1.7 arcsec and therefore typically resolvable with certain ground-based telescopes under most seeing conditions.

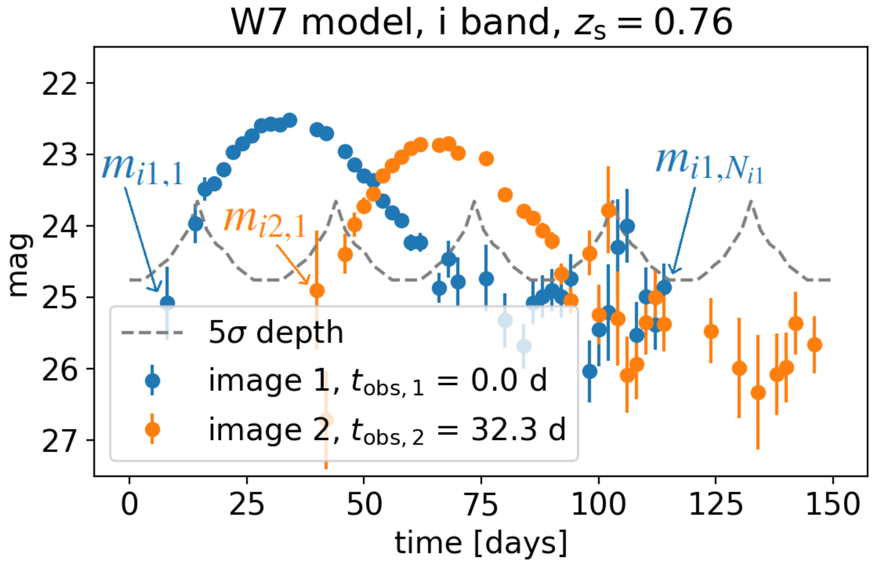


Fig. 4.3. Simulated observation for which ML models will be trained to measure the time delay. The gray dashed curve marks the 5σ point-source depth, that accounts for the moon phase. The marked data points are also listed in Eq. (4.4).

4.10.2. To calculate magnitudes with observational noise we use Eq. (2.106). For the moon phase we assume a configuration where the i-band light curve peaks around new moon. Further configurations in the moon phases will be discussed in Sect. 4.7.1. To avoid unrealistic noisy data points $m_{\mathrm{data,X}}$ for our mock system in Fig. 4.3, we only take points $m_{\mathrm{AB,X}}$ brighter than $m_5 + 2\mathrm{mag}$ into account, before we add noise on top. Furthermore, we assume a two day cadence with a few random gaps.

4.3.2 Data set for machine learning

Our data of the mock LSN Ia contain measurements of light curves in one or more filters of the two SN images. The input data for our ML approaches are ordered, such that for a given filter, all magnitude values from image 1 are listed (first observed to last observed), followed by all magnitude values from image 2. This structure is illustrated in the following definition and will be referred to as a single sample,

$$m_{i1,1} \dots m_{i1,N_{i1}} \, m_{i2,1} \dots m_{i2,N_{i2}} \equiv d_1 \, d_2 \dots d_{N_d}, \tag{4.4}$$

for an example of a double LSN Ia with observations in the i band. There are N_{i1} photometric measurements in the light curve for SN image 1, and N_{i2} photometric measurements for SN image 2. The magnitude value of the first data point in the i band from the first image in Fig. 4.3 is denoted as $m_{i1,1}$, and the last data point is $m_{i1,N_{i1}}$. The first data point of the second image in the i band is denoted as $m_{i2,1}$. For simplification, we define $N_\mathrm{d} = N_{i1} + N_{i2}$, and d_j as the j-th magnitude value in Eq. (4.4). If multiple filters are available, then a ML model can be trained per band, or multiple bands can be used for a single ML model, which will be explored in Sect. 4.6.3.

We introduce our FCNN and RF methods in detail in Sect. 4.4; we describe here the data set required for these two approaches in the remainder of this section. Important to note is that both methods always require the same input structure as defined in Eq. (4.4), with exactly the same number of data points[28]. From this input, we can then build a FCNN or a RF that predicts the time delay. As additional information, the 5σ depth is required for each data point, to create noise in a similar way as in our mock observation. Furthermore, microlensing uncertainties are taken into account by using the κ, γ, and s values of each LSN Ia image. The weakness of this approach is that we need to train a ML model for each observed LSNe Ia individually, but the advantage is that we can train our model very specifically for the observation pattern, noise and microlensing uncertainties such that we expect an accurate result with a realistic account of the uncertainties. Given that the data production and training of such a system take less than a week and multiple systems can be trained in parallel, this approach is easily able to measure the delays of the expected 40 to 100 potentially promising LSNe in the 10 year LSST survey (Huber et al., 2019).

Our ML approaches require the same number of data points in each sample. We therefore produce our data set, for training, validation and testing of the ML models, such that the number of data points is always the same as in our mock observation in Fig. 4.3. We calculate the light curves for the SN images via Eqs. (4.1) and (4.2) where we use random microlensing map positions. We then shift the light curves for each SN image randomly in time around a first estimate of the delay. In our example, we use the true observed time values of the mock observation $t_\mathrm{obs,1} = 0.0\,\mathrm{d}$ and $t_\mathrm{obs,2} = 32.3\,\mathrm{d}$ as the first estimate for the SN images 1 and 2, respectively. For a real system, we do not know these time values exactly and therefore probe a range of values around these first estimates in our training, validation and test sets. In particular, for each sample in the training set, we pick random values between $t_\mathrm{obs} - 10\,\mathrm{d}$ and $t_\mathrm{obs} + 12\,\mathrm{d}$ as the "ground truth" (input true time value) for that specific sample. Different samples in the training set have different ground truth values. We also tested more asymmetric ranges with $t_\mathrm{obs} - 10\,\mathrm{d}$ and $t_\mathrm{obs} + t_\mathrm{est}$, where $t_\mathrm{est} = 16, 18, 22, 30\,\mathrm{d}$, and find results in very good agreement, with no dependence on asymmetries in the initial estimate.

Data points are then created at the same epochs as the initial observation. Using the 5σ depth of each data point of our observation, we calculate for each random microlensing position 10 random noise realizations following Eq. (2.106). Since we are not interested in the overall magnitude values we normalize the resulting light curve by its maximum. Our

---

[28] To avoid unrealistically noisy data points, we limit the maximum amount of noise allowed, as described in Appendix 4.10.1.

total data set used for training has a size of 400000 samples coming from 4 theoretical SN Ia models, 10000 microlensing map positions and 10 noise realizations. Each sample has the data structure of Eq. (4.4). For the validation and test sets, we calculate two additional microlensing maps with the same κ, γ and s values as the training set, but with different microlensing patterns from random realizations of the stars. This provides "clean" validation and test sets that the ML methods have not encountered during training in order to fairly assess the performance of the methods. Our validation and test set have each a size of 40000 samples, from 4 models, 1000 microlensing map positions and 10 noise realizations.

Two examples of our training data are shown in Fig. 4.4 in open circles. The first panel (sample 5) shows for the first SN image a good match to the initial mock observation (in solid circle). The simulated training data are therefore almost the same as the mock observation. Differences in fainter regions (higher normalized magnitudes) come from observational noise. For the second image, the time value t_2 of the simulated training data is larger than the true value $t_{\mathrm{obs},2}$ and therefore we find the peak a few data points later. The general idea of providing data in such a way is that the ML model learns to translate the location of the peak region into the time value t. The difference between the two time values from the first and second image is then the time delay we are interested in. The second panel (sample 33) in Fig. 4.4 is a nice example illustrating why going directly for the time delay is not working that well in this approach. We see that both simulated images for training are offset to the right by almost the same amount. This would in the end lead to a very similar time delay as the initial mock observations, even though the input values are very different from those of the initial observations.

Our described approach can be seen as a fitting process that has the weakness that if the models for training are very different in comparison to a real observation, our approach will fail. From Fig. 4.1 we see that the four SN Ia models predict different shapes of the light curves and locations of peaks. Therefore, to compensate for different peak locations, we randomly shift the four SN Ia models in time by -5 to 5 days. Furthermore, to make the noise level more random and compensate for different peak brightness, we vary also the overall magnitude values by -0.4 to 0.4. The random shifts in time and magnitude are the same for a single sample, and therefore this approach creates basically a new model with the same light-curve shape, but slightly different peak location and brightness. Since the ML models do not know the actual values of the random shifts in time or magnitude the location of the peak for a certain SN Ia model is smeared out. Therefore, this approach introduces a much larger variety in the SN Ia models and Appendix 4.10.3 shows that this helps to generalize to light curves from sources that were not used in training the ML model. We also tested random multiplication factors to stretch or squeeze the light curves in time (instead of the random constant shift in time as just described), but our approach with the random shifts works slightly better as discussed in Appendix 4.10.3. We therefore use the random shifts for the rest of this paper.

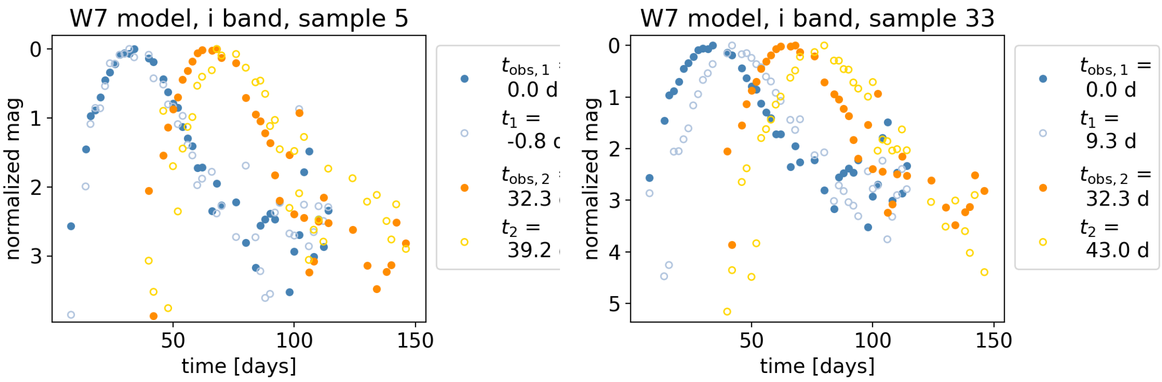


Fig. 4.4. Simulated data to train a ML model. The filled dots correspond to the mock observation shown in Fig. 4.3. The open dots represent the simulated training samples, where two out of the 400000 are shown for the *i* band in the left and right-hand panels.

4.4 Machine learning techniques

In this section we explain the two different ML models used in this work, namely a deep learning network using fully connected layers and a RF. We use these simple ML approaches to get started, because if they work well, then more complicated models might not be necessary. Results from these simple approaches would also serve as a guide for the development of more complex ML models. The techniques all use the input data structure as described in Sect. 4.3, and provide for each image of the LSN Ia a time value t as shown in Fig. 4.3. For the first appearing image, the (ground truth) time $t = 0$ is the time of explosion and for the next appearing image it is the time of explosion plus the time delay Δt. Given our creation of the data set, which is done like a fitting process for each light curve, we do not train the system to predict only the time delay, but instead we have as output one time value per image as described in Sect. 4.3.2.

4.4.1 Deep learning - Fully connected neural network

Neural networks are a powerful tool with a broad range of applications. To solve our regression problem, we used a FCNN, consisting of an input layer, two hidden layers, and one output layer, as shown in Fig. 4.5. Although universal approximation results (Cybenko, 1989; Hornik et al., 1989) suggest that a FCNN with only one hidden layer of arbitrarily large width can approximate any continuous function, FCNNs with finite widths but more layers have shown to be more useful in practice. We therefore used two hidden layers instead of one and tested different widths of the networks by introducing the scaling factor f for a variable number of nodes in the hidden layers in order to optimize the number of hidden nodes.

In our FCNN, each node of the input layer corresponds to a magnitude value of a single observation for a given filter and image, sorted as in the example of Eq. (4.4). Each node of the input layer (d_j) is connected by a weight ($w_{1,jk}$) to each node of the first hidden layer ($h_{1,k}$). In addition, a bias ($b_{1,k}$) is assumed and we introduce non linearities, by using a rectified linear units (ReLU) activation function (e.g., Glorot et al., 2010; Maas et al., 2013), which is 0 for all negative values and the identity function for all positive values. Therefore, the nodes of the first hidden layer can be calculated via

$$h_{1,k} = \mathrm{ReLU}(\sum_{j=1}^{N_\mathrm{d}} w_{1,jk} d_j + b_{1,k}), \qquad k = 1, 2, ...10f. \tag{4.5}$$

Further, all nodes in the first hidden layer are connected to all nodes in the second hidden layer in a similar manner:

$$h_{2,k} = \mathrm{ReLU}(\sum_{j=1}^{10f} w_{2,jk} h_{1,j} + b_{2,k}), \qquad k = 1, 2, ...5f. \tag{4.6}$$

The nodes from the second hidden layer are then finally connected to the output layer to produce the time values

$$t_k = \sum_{j=1}^{5f} w_{3,jk} h_{2,j} + b_{3,k}, \qquad k = \begin{cases} 1, 2 \text{ (double system)} \\ 1, 2, 3, 4 \text{ (quad system).} \end{cases} \tag{4.7}$$

The output layer consists of two nodes for a double LSN Ia and four nodes for a quad LSN Ia. We tested also other FC network structures such as using a different network for each image, using three hidden layers, or using a linear or leaky ReLU activation function, but our default approach described above works best.

We train our system for a certain number of epochs N_epoch, where we use the ML library `PyTorch` (Paszke et al., 2019). At each epoch, we subdivide our training data randomly into mini batches with size N_batch. Each mini batch is propagated through our network to predict the output that we compare to the ground-truth values by using the mean squared error (MSE) loss. To optimize the loss function, we use the Adaptive Moment Estimation (Adam) algorithm (Kingma & Ba, 2014) with a learning rate α on the MSE loss to update the weights in order to improve the performance of the network[29]. Per epoch, we calculate the MSE loss of the validation set from our FCNN, and store in the end the network at the epoch with the lowest validation loss. By selecting the epoch with the lowest validation loss, we minimize the chance of overfitting to the training data. Typically we reach the lowest validation loss around epoch 200 and an example for the training and validation curve for our FCNN is shown in Appendix 4.10.4.

The test data set is used in the end to compare different FCNNs, which have been trained with different learning rates α, sizes f and mini-batch sizes N_batch.

4.4.2 Random forest

The RF (Breiman, 2001) is a method used for classification and regression problems, by constructing many random decision trees. In this section we give a brief introduction on the idea of a RF and explain the setup we are using.

To build a RF, we construct many random regression trees, which are a type of decision

---

[29] For the other hyperparameters of the Adam optimizer, we used the PyTorch default values.

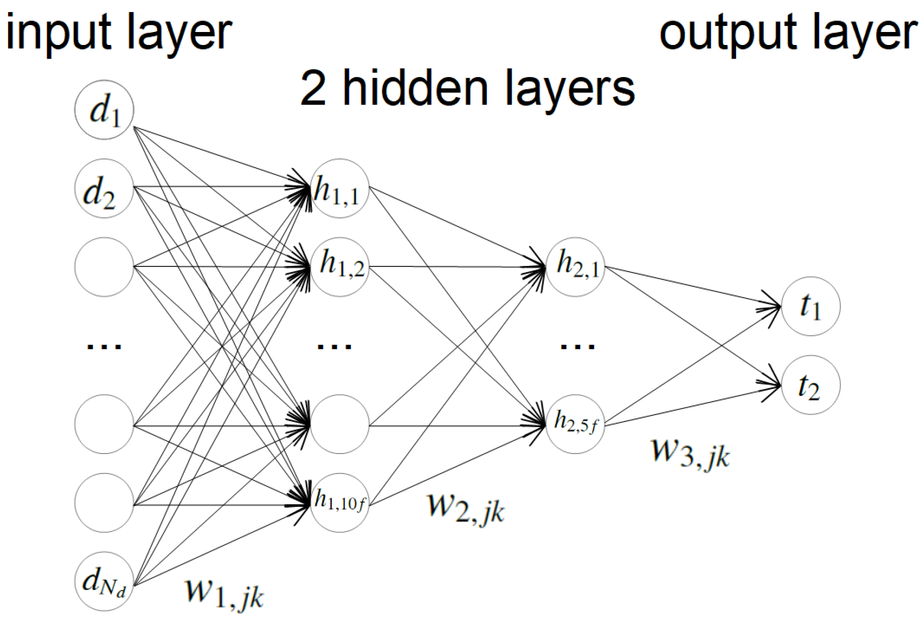


Fig. 4.5. FCNN, where the input layer has N_d data points and d_j stands for the magnitude value of the j-th data point in Eq. (4.4). The size of the two hidden layers scales by a factor f, and the outputs are two (four) time values for a double (quad) LSN Ia.

trees, where each leaf represents numeric values (for the outputs). For our case, we create a total of N_trees random regression trees where a schematic example for a single regression tree is shown in Fig. 4.6. The root node is shown in magenta, the internal nodes in gray and the leaf nodes in green. The root node splits our whole data set containing samples as defined by Eq. (4.4), into two groups based on a certain criterion (e.g., $m_{i1,2} < 1.2$): first where the criterion is true, and second where it is not. The internal nodes split the data in the same manner, until no further splitting is possible and we end up at a leaf node to predict the two time values t_1 and t_2 as output.

To create random regression trees we use a bootstrapped data set, which draws randomly samples from the whole training data (400000 samples) until it reaches a given size $N_\mathrm{max\,samples}$. Importantly, an individual sample of the original training data can be drawn multiple times and each random regression tree is built from an individual bootstrapped data set, which is used to create the root, internal and leaf nodes. However, only a random subset of the features (e.g., just $m_{i1,2}$, $m_{i2,5}$, and $m_{i2,9}$) is considered to construct the root node or a single internal node, where the splitting criterion (e.g., $m_{i1,2} < 0.5$) of a single feature is defined based on the mean value (e.g., $\bar{m}_{i1,2} = 0.5$) from all samples under investigation ($N_\mathrm{max\,samples}$ for the root node and fewer samples for the internal nodes depending on how the data set was split before). The number of available features we pick randomly from all features for the creation of a node is $N_\mathrm{max\,features}$[30].

In the following, we demonstrate the construction of the root node for a regression tree, as shown in Fig. 4.6, for the example of $N_\mathrm{max\,features} = 3$. Therefore, we randomly picked three features from Eq. (4.4), which we assumed to be $m_{i1,2}$, $m_{i2,5}$, and $m_{i2,9}$. From a bootstrapped data set with $N_\mathrm{max\,samples}$ samples of our training data set, we assumed to find

---

[30] A single feature such as $m_{i1,2}$ can be picked multiple times.

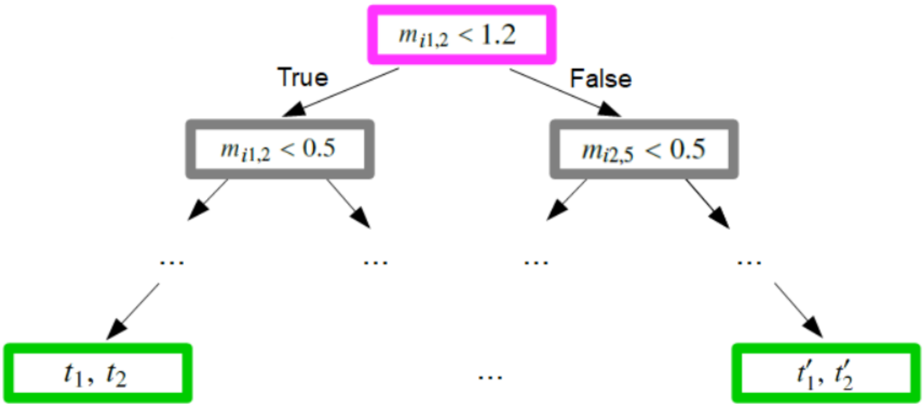


Fig. 4.6. Schematic example of a regression tree for a double system that predicts two time values for certain input data as in Eq. (4.4). The root node is represented by the magenta box, the internal nodes by gray boxes, and the leaf nodes by green boxes.

the mean values $\bar{m}_{i1,2} = 1.2$, $\bar{m}_{i2,5} = 1.0$, and $\bar{m}_{i2,9} = 0.6$. Therefore, we investigated the three criteria $m_{i1,2} < 1.2$, $m_{i2,5} < 1.0$, and $m_{i2,9} < 0.6$ as potential candidates for the root node, where each of the criteria splits the $N_{\text{max samples}}$ training samples into two groups. We selected the best splitting criterion as the one that results in the lowest variance in the predictions within each of the groups created by the split. In other words, we can compute through this comparison a residual for t_1 and t_2 for each sample. From this, we can calculate the sum of squared residuals for each candidate criterion, and the criterion that predicts the lowest sum of squared residuals will be picked as our root node, which would be $m_{i1,2} < 1.2$ in our schematic example. For each of the resulting two groups, we followed exactly the same procedure to construct internal nodes that split the data further and further until no further splitting is possible or useful[31] and we end up at a leaf node to predict the output. To avoid a leaf node containing just a single training sample, we used two parameters, namely, N_{msl}, the **m**inimum number of **s**amples required to be in a **l**eaf node, and N_{mss}, the **m**inimum number of **s**amples required to **s**plit an internal node. From the multiple training samples in a leaf node, the t_1 and t_2 values of a leaf node are the average of all samples in the leaf node.

Following the above procedure, many random regression trees are built; to create an output for a single (test) sample, all regression trees are considered and the final output is created from averaging over all trees.

For this approach we used the object `sklearn.ensemble.RandomForestRegressor` of the software `scikit-learn` (Pedregosa et al., 2011; Buitinck et al., 2013), where we assume the default parameters except for the previously mentioned N_{msl}, N_{mss}, N_{trees}, $N_{\text{max samples}}$ and $N_{\text{max features}}$.

4.5 Machine learning on example mock observation

In this section we apply the ML techniques from Sect. 4.4 to our example mock observation of a double LSNe Ia described in Sect. 4.3. In Sect. 4.5.1 we find the best FCNN and RF

---

[31] Further splitting is not useful if none of the investigated splitting criteria would lead to further improvements of the sum of squared residuals in comparison to not splitting the remaining samples.

and compare results from the corresponding test sets based on the four theoretical models also used in the training process. In Sect. 4.5.2 we use the best FCNN and RF and apply it to an empirical data set not used in the training process to test the generalizability of both models. This final test is very important since in reality we can never assure that our assumed light-curve shapes in the training process will fully match a real observation.

4.5.1 Best fit: Fully connected neural network versus random forest

To find a FCNN and a RF that provide the best fit to our mock observation from Fig. 4.3, we explore a set of hyperparameters as listed in Table 4.2 for the FCNN and Table 4.3 for the RF.

To find the best ML model for our mock observation, we used the test set to evaluate each set of hyperparameters. This is just to find an appropriate set of hyperparameters, which we will use for the sake of simplicity from here on throughout the paper[32]. Our final judgment of the performances of the ML models will be based on the "SNEMO15 data set" where light curves will be calculated using an empirical model (see Sect. 4.5.2). The distinctions between the various data sets for our ML approaches are summarized in Table 4.4. For each sample i of the test set, we get two time values, $t_{1,i}$ and $t_{2,i}$, from which we can calculate the time delay $\Delta t_i = t_{1,i} - t_{2,i}$, which we compare to the true time delay $\Delta t_{\mathrm{true},i}$ to calculate the "time-delay deviation" of the sample as

$$\tau_i = \Delta t_i - \Delta t_{\mathrm{true},i}. \tag{4.8}$$

We investigate here the absolute time-delay deviation instead of the relative one ($\tau_i/\Delta t_{\mathrm{true},i}$), because this allows us to draw conclusions about the minimum time delay required to achieve certain goals in precision and accuracy. From our results, we do not find a dependence on the absolute time delay (e.g., 32.3 d for Fig. 4.4) used in the training process, which is what we expect from the setup of the FCNN and the RF and is demonstrated in Sect. 4.7.4

For the FCNN, we find that $(\alpha, f, N_{\mathrm{batch}}, N_{\mathrm{epoch}}) = (0.0001, 40, 256, 400)$ provides the best result, meaning that the median of τ_i of the whole test set is lower than 0.05 days (to reduce the bias) and the 84th−16th percentile (1σ credible interval) of the test set is the lowest of all networks considered. For the RF the hyperparameters (N_{trees}, N_{mss}, N_{msl}, $N_{\mathrm{max\,samples}}$, $N_{\mathrm{max\,features}}$) = (800, 4, 1, 200000, $\sqrt{N_{\mathrm{all\,features}}}$) provide the best result. In the following we always use these two sets of hyperparameters for the FCNN or the RF, unless specified otherwise. We note that $N_{\mathrm{trees}} = 800$ is on the upper side of what we investigated, but increasing the number of trees further makes the computation even more costly. Nevertheless, we tested also N_{trees} = 1000, 1200, 1600, 2000 and $N_{\mathrm{trees}} = 3000$ with (N_{mss}, N_{msl}, $N_{\mathrm{max\,samples}}$, $N_{\mathrm{max\,features}}$) from the best fit as listed above. We find results that are basically the same as for $N_{\mathrm{trees}} = 800$ or slightly worse (0.02 d at most) and therefore we stick with $N_{\mathrm{trees}} = 800$, which is sufficient. The comparison between the FCNN and the RF is shown

---

[32] Such hyperparameter optimizations are usually performed using the validation set, whereas we are using the test set because we have the SNEMO15 data set for ultimate performance test.

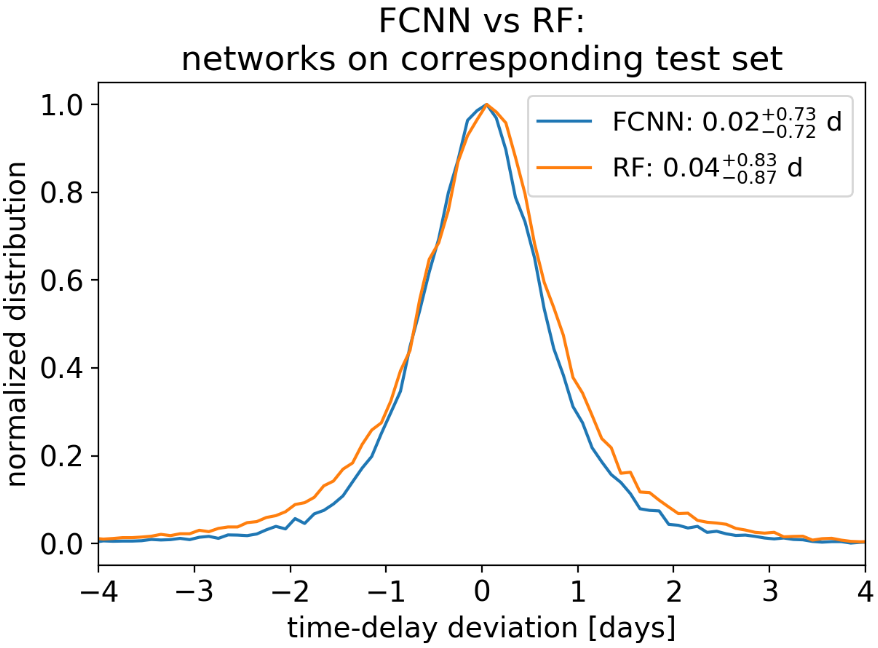


Fig. 4.7. FCNN and RF on the whole sample of light curves from the specified test set for the mock observation in Fig. 4.3. The ML models' hyperparameters are set to the values at which the test set yields a bias below 0.05 days and the smallest 68% credible interval of the time-delay deviation (in Eq. (4.8)).

in Fig. 4.7, where we quote the median (50th percentile), with the 84th-50th percentile (superscript) and 16th-50th percentile (subscript) of the whole sample of light curves from the corresponding test set. The results include microlensing and observational uncertainties as described in Sect. 4.2. For the training and testing, we considered the four SN Ia models, merger, N100, sub-Ch and W7 (therefore we use the description "corresponding test set" in the title of Fig. 4.7). Further, the results are based on using just the i band, assuming the data structure as defined in Eq. (4.4).

Instead of looking at the whole sample of light curves from the test set at once, we show in Appendix 4.10.5 how the time-delay deviation τ_i depends on the time delay of the test samples. We find for both networks a slight trend that time delays far away from the true time delay of the mock observation yield larger deviations, where the effect is stronger for the RF in comparison to the FCNN. However, this is not surprising as very long time delays come from rare scenarios where the t_1 value of the first image is highly underestimated and the t_2 value of the second image is highly overestimated. Similarly, very short time delays tend to have t_1 that is highly overestimated and t_2 that is underestimated. Given that these scenarios are rare in the training set, it is more difficult to learn these cases. Still, the FCNN compensates for these edge effects better, which explains the better performance of the FCNN in comparison to the RF on the corresponding test set as shown in Fig. 4.7.

However, we see that both ML models provide accurate measurements of the time delay with the 1σ uncertainty for the FCNN around 0.7 days and the RF around 0.8 days, where both have low bias ($\leq$0.04 days). Nevertheless, the training and test set is produced by using the same SN Ia models. If light curves in the test sets are different from the ones used for

Table 4.2. Investigated parameters for the training process of the FCNN (see Fig. 4.5) for the system listed in Table 4.1.

α	0.01, 0.001, 0.0001, 0.00001
f	5, 10, 20, 40, 80, 160
N_{batch}	64, 128, 256, 512
N_{epoch}	400

Notes. We vary the learning rate, α, the size of the hidden layers by a factor f, and the size of the mini batches, N_{batch}. Furthermore, 400 training epochs (N_{epoch}) are sufficient given that the minimum loss of the validation set is typically reached around 200 training iterations.

Table 4.3. Investigated parameters for the training process of the RF (see Fig. 4.6 for a single regression tree) for the system listed in Table 4.1.

N_{trees}	200, 400, 800
N_{mss}	2, 4, 8
N_{msl}	1, 2, 4
$N_{\mathrm{max\,samples}}$	50000, 100000, 200000, 300000, 400000
$N_{\mathrm{max\,features}}$	1, $\sqrt{N_{\mathrm{all\,features}}}$, $N_{\mathrm{all\,features}}$

Notes. We vary the number of trees, N_{trees}, the minimum number of samples required to split an internal node, N_{mss}, the minimum number of samples required to be in a leaf node, N_{msl}, the size of the bootstrapped data set, $N_{\mathrm{max\,samples}}$, and the maximum number of features, $N_{\mathrm{max\,features}}$, considered to create a root or internal node.

training, this can lead to broadened uncertainties, and more critically, also to biases (see Appendix 4.10.3). Further, we learn from Appendix 4.10.3 that, as soon as the different light curves used for training cover a broad range, the trained ML model can be used for light-curve shapes it has never seen. Therefore, in Sect. 4.5.2, we evaluate the RF and the FCNN trained on four theoretical models on a data set based on the empirical `SNEMO15` model.

4.5.2 Generalizability of ML models: Evaluation on *SNEMO15* data set

To test if the ML models trained on four SN Ia models with the random shifts in time and magnitude as introduced in Sect. 4.3.2 can generalize well enough to real SN Ia data, we created a data set based on the empirical `SNEMO15` model, which is shown in Fig. 4.1. The empirical model covers only a wavelength range from 3305 Å to 8586 Å, and with $z_{\mathrm{s}} = 0.76$ (Table 4.1), the *i* band is the bluest band we can calculate.

To account for macrolensing and brightness deviations for the `SNEMO15` model in comparison to the theoretical SN models, we set the median `SNEMO15` light curve equal to the mean value of the four macrolensed SN Ia models. Since the light curves are normalized before the training process, this is only important to avoid over- or underestimations of the observational noise. Furthermore, to include microlensing, we use microlensed light curves

Table 4.4. Explanation of the different types of data sets used for our ML approaches.

Data set type	Description and purpose
training set	To train the ML models. Light curves of four theoretical SNe Ia models are used.
validation set	To find the training epoch for the FCNN that has lowest validation loss (four SNe Ia models as in training process).
(corresponding) test set	To evaluate the performance of a ML model using four theoretical models as in the training process. The term "corresponding" is used if all parameters (e.g., κ, γ, ...) for the production of the test set are the same as for the training set. This data set does not test the generalizability to different SN Ia light-curve shapes.
SNEMO15 data set	Final test set using light curves from the empirical `SNEMO15` model not used in the training process, which most importantly tests the generalizability of the trained ML models.

from the four theoretical models, initially created for the corresponding test set, and subtract the macrolensed light curve, assuming $\mu_{\text{macro}} = 1/((1-\kappa)^2 - \gamma^2)$. Therefore, we get from our 4 models 4000 microlensing contributions for the light curves, which the FCNN or the RF have not seen in its training process. For each of the microlensing contributions, we then draw randomly one of the 171 `SNEMO15` light curves to create a microlensed `SNEMO15` light curve. From the 4000 microlensing contributions, we have a sample of 4000 microlensed light curves. For each light curve, we then draw 10 random noise and time-delay realizations to create a data set, as described in Sect. 4.3.2. We call this the `SNEMO15` data set.

Fig. 4.8 shows the results where we evaluate the FCNN and the RF from Fig. 4.7, trained on four theoretical SN Ia models, on the corresponding test set (built from the same four theoretical SN Ia models) and on the `SNEMO15` data set. The first important thing we note is that the RF shows almost no bias, whereas the FCNN has a higher bias when evaluated on the `SNEMO15` data set. To investigate this further, we look at results from the RF and the FCNN for the set of hyperparameters as listed in Tables 4.2 and 4.3 for three different cases using the i band, z band, or y band.

We find that the absolute bias of the FCNN for the different hyperparameters and bands (i, z, and y) is mostly below 0.4 days but higher values are also possible. The problems are that these variations in the bias in the `SNEMO15` data set are not related to biases we see in the corresponding test sets or due to a specific set of hyperparameters. As a result, we cannot identify the underlying source of the bias, apart from that it is due to suboptimal generalization of the theoretical SN Ia models to `SNEMO15` in the FCNN framework.

The RF works much better in this context, as the absolute bias is always lower than

0.12 days for the i, z, and y bands. Only the hyperparameter $N_{\max\,\mathrm{features}} = 1$ can lead to a higher bias up to 0.22 days, but this hyperparameter is excluded because of its much worse performance in precision on the corresponding test set in comparison to $N_{\max\,\mathrm{features}} = \sqrt{N_{\mathrm{all\,features}}}$ or $N_{\max\,\mathrm{features}} = N_{\mathrm{all\,features}}$. Therefore, as long as we restrict ourselves to LSNe Ia with delays longer than 12 days we can achieve a bias below 1 percent, which allows accurate measurements of H_0. Furthermore, the bias is not the same in all filters. While the absolute bias in the y band goes up to 0.12 days, we have a maximum of 0.08 days in the z band and 0.03 days in the i band. The comparison of multiple bands therefore helps to identify some outliers.

The bias investigation of the FCNN and the RF is summarized in Fig. 4.9 using all hyperparameters (except $N_{\max\,\mathrm{features}} = 1$, which is excluded because of its bad performance on the corresponding test set) and the i, z, and y bands. From the left-hand panel we see that the large biases of our FCNN on the `SNEMO15` data set are not related to biases we see in the corresponding test set and therefore identifying a set of hyperparameters just from the corresponding test set which works also well on the `SNEMO15` data set is not possible. From the right-hand panel of Fig. 4.9 we see that also the biases in the RF from the corresponding test set and the `SNEMO15` data set are not directly related with each other but this is not a problem as the biases on the `SNEMO15` data set are low enough for precision cosmology. From this example we see that the RF is able to generalize to a new kind of data not used in the training process, which does not work well for our FCNN. In principle this was already suggested by the investigation done in Appendix 4.10.3, but with the random shifts in time we introduced, it seemed to significantly improve the generalizability, but it was still not enough for the final test on the `SNEMO15` data set. Investigating the importance of all the input features as listed in Eq. (4.4), we find that the FCNN focuses mostly on the peak directly whereas for the RF the features before and after the peak are the most important ones. More about this is discussed in Appendix 4.10.6.

In the remainder of the paper, we proceed to present results based on the RF, because the significant bias in our FCNN makes accurate cosmology difficult to achieve especially for LSN Ia systems with short delays. Using deeper networks would not be enough to improve our FCNN, as this would just allow a better fit to the training data but does not ensure any improvement on the generalizability of the network. Therefore, it would be necessary to provide more realistic input light curves for the training process, as it has problems to generalize to light-curve shapes it has not seen. Such an improvement could be achieved by using the `SNEMO15` light curves as well in the training process, but then a test set with light-curve shapes it has never seen would be missing. Another approach would be to incorporate regularization or dropout into our FCNN or by constructing a network that outputs in addition to the time values the associated uncertainties, but given that this was not necessary for the corresponding test set to perform well, it would be some kind of fine tuning to our `SNEMO15` data set, because all tests before were encouraging to proceed to the final test. Therefore, we postpone further investigations of FCNNs to future studies, especially since other network architecture, such as recurrent neural networks, long short-term memory networks (Sherstinsky, 2020), or Gaussian processes, could potentially reduce

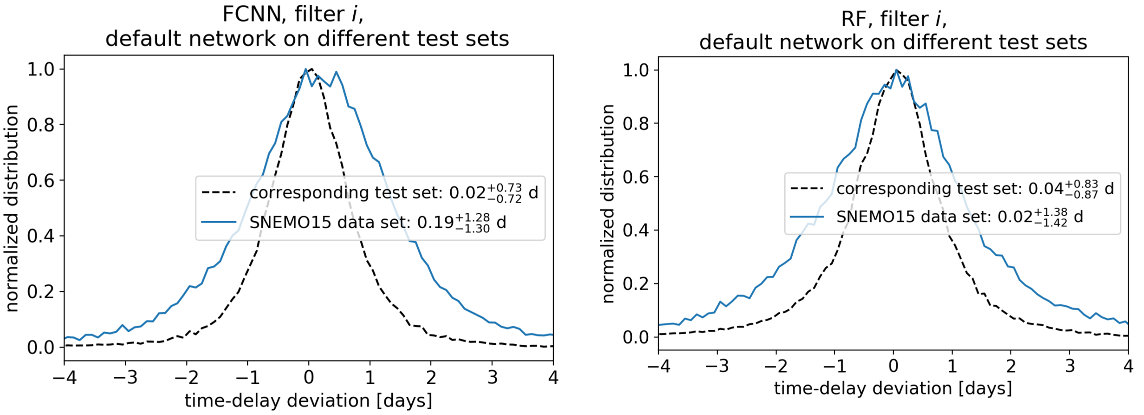


Fig. 4.8. FCNN and RF trained on four theoretical models for the i band evaluated on the whole sample of light curves from the two specified data sets. The dashed black line represents the corresponding test set based on the four theoretical models, and the data set of the blue line is based on the empirical `SNEMO15` model.

model complexity while having lower inductive bias[33] (Wilson & Izmailov, 2020).

Another thing we learn is that the distribution of the recovered time delays from the `SNEMO15` data set is ∼0.5 days broader than that of the corresponding test sets. This is not surprising as the RF and the FCNN have never seen such light curves in the training process. A ∼1.4 day precision on a single LSN Ia is still a very good measurement and allows us to conduct precision cosmology from a larger sample of LSNe Ia. Nevertheless, we see in this section that even though the uncertainties for the RF are larger than that of the FCNN, the RF provides low bias when used on empirical data and is therefore preferred.

4.6 Microlensing, noise, and choice of filters

In this section we use the RF from Sect. 4.5.1 and apply it to the mock observation from Sect. 4.3, for hypothetical assumptions about microlensing and noise to find sources of uncertainties (Sect. 4.6.1 and 4.6.2). We further investigate potential bands to target for follow-up observations (Sect. 4.6.3). In this section all results presented are based on the RF on test sets from the four theoretical models. The conclusions drawn in this section would be the same if the results from the FCNN would be presented.

4.6.1 Microlensing map parameters κ, γ, s

To investigate uncertainties in the microlensing characterization, we use the RF from Sect. 4.5.1, but evaluate it on different test sets with varying κ, γ, and s values, which deviate from the original training data.

Fig. 4.10 shows the RF evaluated on different test sets. The black dashed line represents the evaluation of the RF on the corresponding test set, which is calculated according to Sect. 4.3.2. The blue and orange lines represent very similar test sets, but calculated on a different

---

[33] Inductive bias refers to the bias coming from assumptions that a ML model has to make to generalize based on training samples.

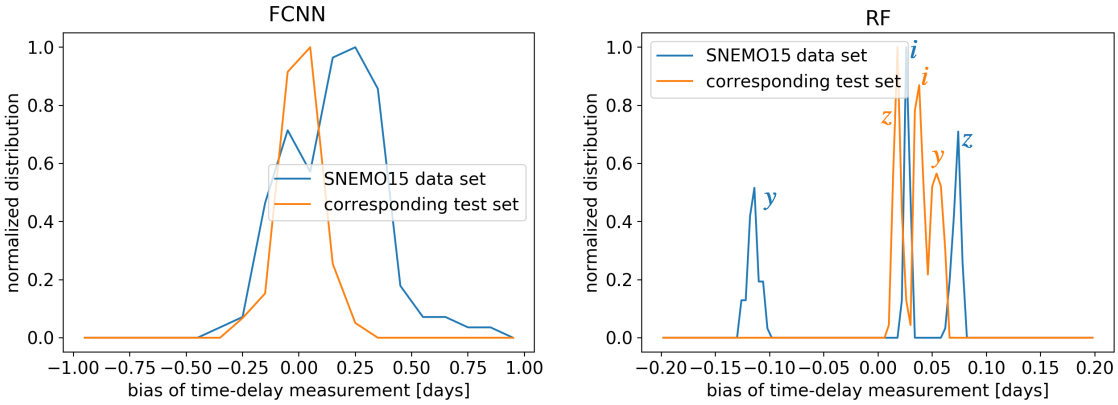


Fig. 4.9. Bias of FCNN and RF on the corresponding test set, composed of four theoretical SN Ia models used for training, and a data set based on the empirical `SNEMO15` model, not used during training, for a variety of different hyperparameters and filters (i, z, and y), i.e., from model averaging. The large biases on the `SNEMO15` data set up to 1 day in our FCNN approach come from the different hyperparameters even though the corresponding test set provides biases within 0.25 days. The RF provides much lower biases in all cases; it depends only weakly on the hyperparameters and is instead mostly set by the filters under consideration.

microlensing map. Instead of the κ and γ values listed in Table 4.1, we assume for the first image $(\kappa,\gamma) = (0.201, 0.225)$ and for the second image $(\kappa,\gamma) = (0.775, 0.765)$ to calculate the test set corresponding to the blue line. The orange line represents a LSNe Ia where we have for the first image $(\kappa,\gamma) = (0.301, 0.325)$ and for the second image $(\kappa,\gamma) = (0.875, 0.865)$. Even though the RF has never seen (κ,γ) configurations as represented by the orange and blue line in the training process, the results are very similar to the corresponding test set of the RF and given that typical model uncertainties are around 0.05 (e.g., More et al., 2017), uncertainties in κ and γ are not critical for our procedure.

In Fig. 4.11 we do a similar investigation, but this time we vary the s value of the microlensing maps. From the comparison of the black dashed line to the orange line, which represents almost the same s value, we see that the uncertainties are almost comparable. Therefore, the much wider uncertainty for $s = 0.3$ (blue line) is not due to variations from different microlensing maps for the same parameter set, but from the fact that lower s values provide more micro caustics in the map, which leads to more events where these caustics are crossed and therefore to more microlensing events and higher uncertainties. This also explains the much tighter uncertainties of $s = 0.9$, which corresponds to a much smoother microlensing map. These results are in good agreement with those of Huber et al. (2021), who also showed that higher s values lead to lower microlensing uncertainties.

For a real observation, the s value is often not known very precisely, which is no problem as the RF still works very well. The only thing one has to be careful about is that an underestimation of the s value leads to an overestimation of the overall uncertainties. Therefore, going for a slightly lower s value as one might expect is a good way to obtain a conservative estimate of the uncertainties.

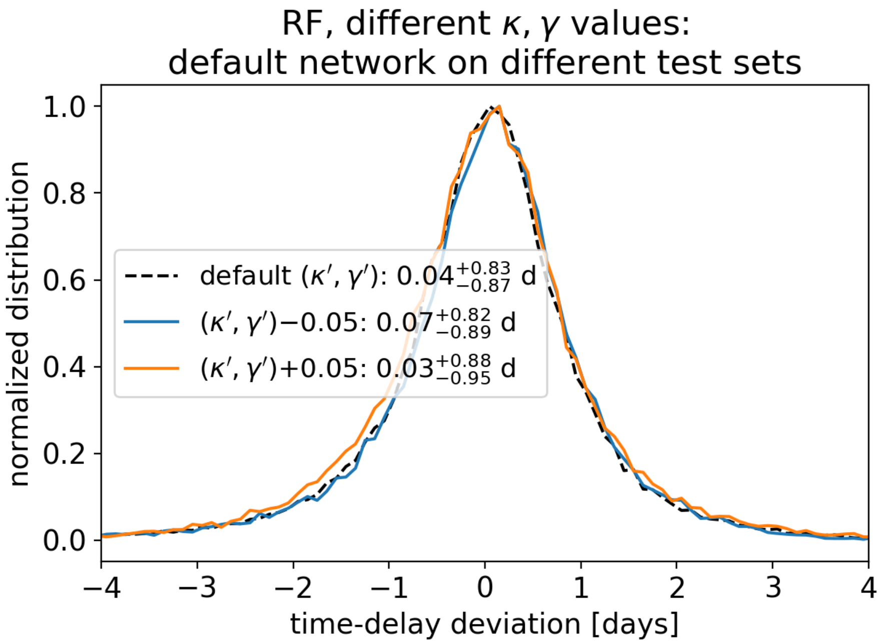


Fig. 4.10. RF evaluated on all samples from its corresponding test set (black dashed line, where training and test sets have the same κ and γ values) and on all samples from two other test sets (blue and orange), with slightly different κ and γ values of the microlensing map in comparison to that of the training data.

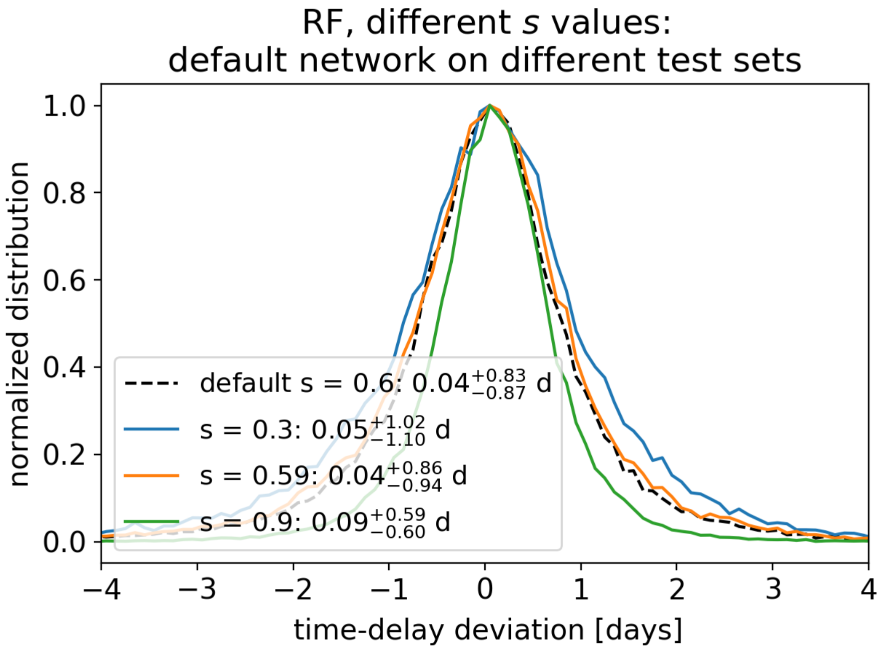


Fig. 4.11. RF evaluated on all samples from its corresponding test set (black dashed line, where training and test sets have the same s value) and on all samples from three other test sets (blue, orange, and green), with different s values of the microlensing map in comparison to that of the training data.

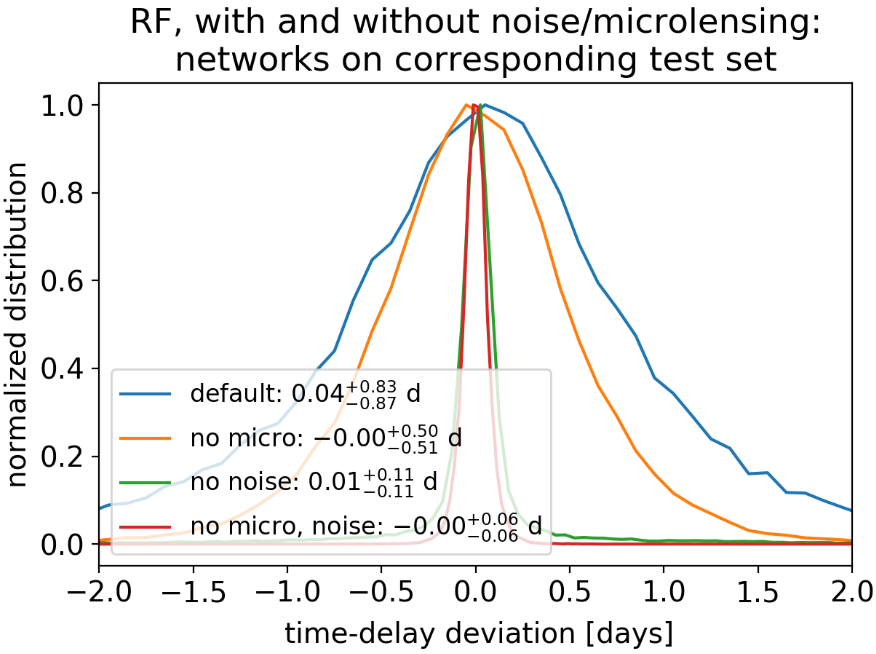


Fig. 4.12. Comparison of the RF model from Sect. 4.5.1 to three other RF models with hypothetical assumptions about noise and microlensing. Each histogram is based on the whole sample of light curves from their corresponding test set. For our realistic mock observation, the noise in the light curves dominates over microlensing as the main source of uncertainty for measuring the time delays.

4.6.2 Uncertainties due to microlensing and noise

In this section we compare the RF from Sect. 4.5.1 to other RF models with various assumptions about microlensing and noise as shown in Fig. 4.12.

From the two cases containing microlensing in comparison to the two cases without microlensing, we find that microlensing increases the uncertainties almost by a factor of two. Although this is quite substantial, we see that the contribution of the observational noise is much higher and is the dominant source of uncertainty in the time-delay measurement. Therefore, to achieve lower uncertainties, deeper observations with smaller photometric uncertainties are required. This is in agreement with Huber et al. (2019), who found that a substantial increase in the number of LSNe Ia with well measured time delays can be achieved with greater imaging depth.

4.6.3 Filters used for training

In this section we investigate eight different filters (*ugrizyJH*) and possible combinations of them to get more precise measurements. Fig. 4.13 shows eight RF models where each is trained and evaluated on a single band. The *i* band, presented first in Sect. 4.5.1 provides the most precise measurement. The next promising filters are *r*, *z*, *g*, and *y* in that order. For the bands *u*, *J*, and *H*, the precision of the measurement is poor and therefore almost not usable. The reason for the strong variation between different bands is the quality of the light curve, which becomes clear from Fig. 4.20, where only the *g* to *y* bands provide observations where the peak of the light curves can be identified. Light curves with the best quality are the r and i bands, which therefore work best for our RF.

There are different ways to combine multiple filters to measure the time delay. The first possibility would be to construct color curves to reduce the effect of microlensing in the so-called achromatic phase (Goldstein et al., 2018; Huber et al., 2021). However, as pointed out by Huber et al. (2021) our best quality color curve $r-i$ would be not ideal as there are no features for a delay measurement within the achromatic phase. Further, we saw in Sect. 4.6.2 that our dominant source of uncertainty is the observational noise instead of microlensing. Therefore, using color curves for this mock example is not practical. We further see that even though color curves are in theory a good way to reduce microlensing uncertainties, in a real detection it might fail because not enough bands with high quality data are available.

Another way of combining multiple filters is to train a single RF model for multiple filters. Generalizing Eq. (4.4) for the r and i bands, we used as input structure

$$m_{r1,1}\, m_{r1,2} \,..\, m_{r1,N_{r1}}\, m_{r2,1} \,..\, m_{r2,N_{r2}}\, m_{i1,1} \,..\, m_{i1,N_{i1}}\, m_{i2,1} \,..\, m_{i2,N_{i2}}, \tag{4.9}$$

and more bands will be attached in the same way. The results are summarized in Fig. 4.14, where we see that combining the two most promising bands improves the uncertainty by about 0.1 days, but adding more bands does not help. Comparing these results to Fig. 4.15, where different distributions from Fig. 4.13 are multiplied with each other[34], we see that a single RF model for multiple filters does not profit much from multiple bands. Therefore, it is preferable to use a single RF model per band and combine them afterward. Using three or more filters can also help to identify potential biases in a single band as pointed out in Sect. 4.5.2. Combining the r, i, and z bands via multiplication helps to reduce the uncertainty by more than a factor of two in comparison to using just the i band for our system with $z_s = 0.76$. Further bands that might be considered for follow-up observations are the g and y bands.

The choice of the ideal filters depends on the source redshift and therefore we show in Fig. 4.26 a similar plot as in Fig. 4.13 but for $z_s = 0.55$ and $z_s = 0.99$, which corresponds to the 16th and 84th percentile of the source redshift from LSNe Ia in the OM10 catalog. From this we learn that the three most promising filters are the g, r, and i bands for $z_s \lesssim 0.6$, whereas for $z_s \gtrsim 0.6$ the r, i, and z bands are preferred. The main reason for this behavior is the low rest-frame UV flux of SNe Ia due to line blanketing, which gets shifted more and more into the g band for higher z_s. If four filters could be used, then we have g, r, i, and z for $z_s \lesssim 0.8$ and r, i, z, and y for $z_s \gtrsim 0.8$. If resources for five filters are available, we recommend g, r, i, z, and y; the J band might be preferred over the g band for high source redshifts ($z_s > 1.0$). However, given the poor precision in the g and J bands at such high redshifts, it is questionable how useful the fifth band is in these cases.

---

[34] We assume that different filters have independent detector noise.

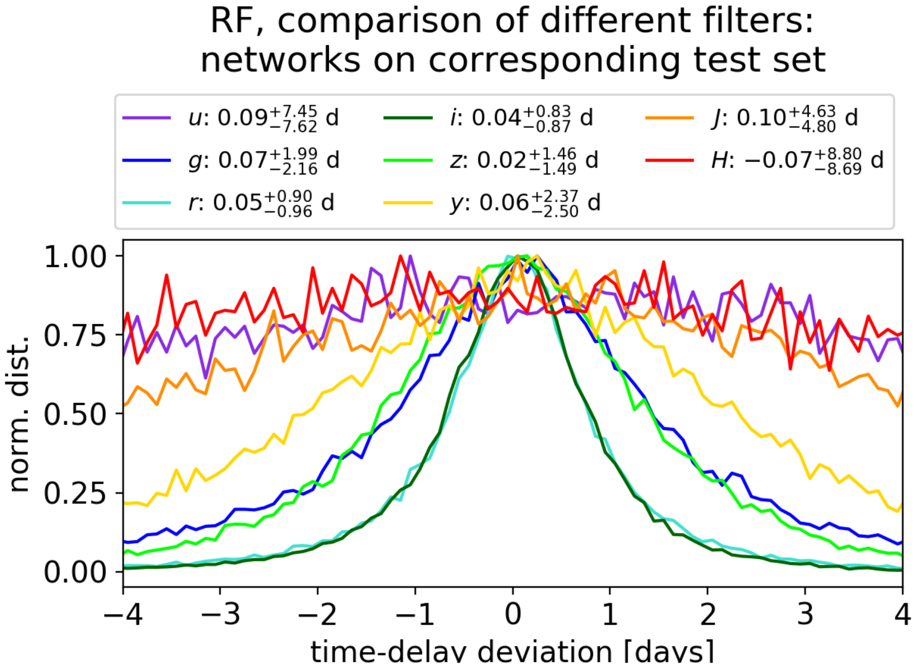


Fig. 4.13. Eight different RF models, each trained on a data set from a single band (as indicated in the legend) and evaluated on the whole sample from the corresponding test set, similar in procedure to Sect. 4.5.1.

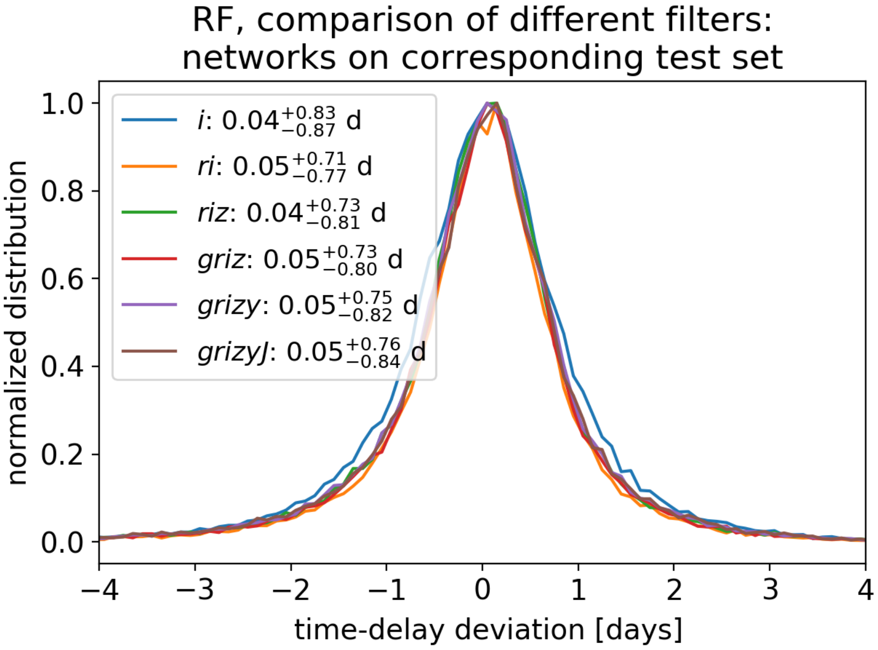


Fig. 4.14. Multiple filters used to train a single RF following the data structure as defined in Eq. (4.9) (example for *ri*). Each histogram is based on the whole sample of light curves from their corresponding test set. Using more than two filters does not improve the results further.

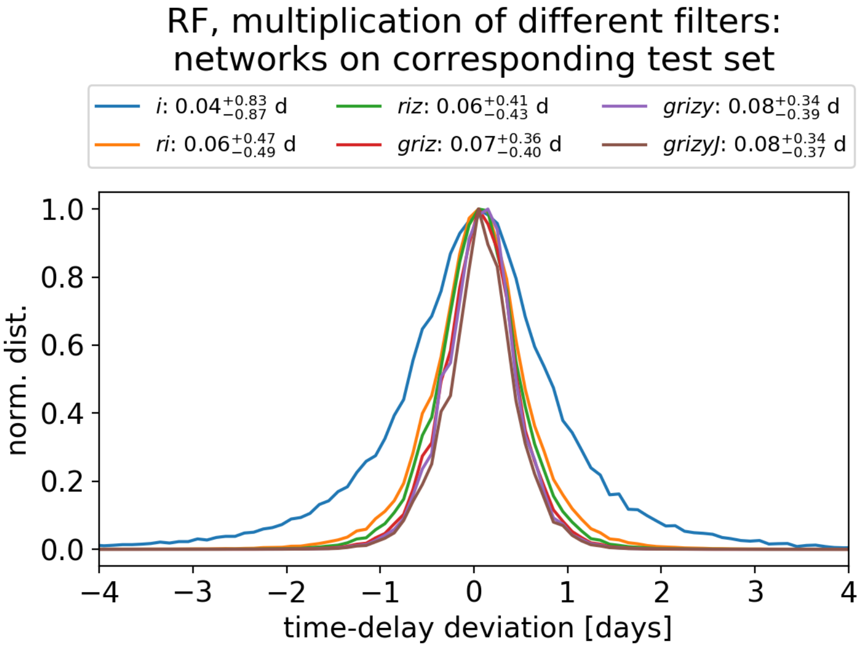


Fig. 4.15. Single RF trained per filter using a similar data structure as in Eq. (4.4) (example for the i band), leading to six RF models for the six filters g, r, i, z, y, and J. The combination of the filters is done by multiplying the corresponding distributions shown in Fig. 4.13. We see that multiple filters help to drastically reduce the uncertainties. Therefore, observing three to four bands would be ideal.

4.7 Machine learning on further mock observations

In this section we investigate further mock systems. We test systems with different moon phases (Sect. 4.7.1) and source, respectively lens redshifts (Sect. 4.7.2) to investigate the change of the uncertainties in comparison to our mock system from Sects. 4.3, 4.5, and 4.6. Furthermore, we test the number of data points required before peak to achieve good time-delay measurements (Sect. 4.7.3) and a quad system with various different properties in comparison to our previous studies (Sect. 4.7.4).

4.7.1 Different moon phases

In this section we address the effect of different moon phases. We assume the same LSN Ia as in Sects. 4.3, 4.5, and 4.6, but place it differently in time. From Fig. 4.20, we can already estimate that if we ignore the u band, which has too low signal-to-noise anyway, mostly the g band will be influenced as other bands are significantly brighter than the 5σ point-source depth or there is only a minor dependence on the moon phase.

For the LSN Ia presented in Sects. 4.3, 4.5, and 4.6, we see from Fig. 4.20 that for the g band, the observations before the peak are significantly affected by moon light, which according to Fig. 4.13 leads to an uncertainty around 2.1 d. For a case where the peak in the g band overlaps with the full moon we find a similar uncertainty, whereas a case where the peak in the g band matches the new moon has an uncertainty around 1.7 d. For cases where the peak is not significantly brighter than the 5σ point-source depth, the moon phase is important, but given that our ML models work with a variable 5σ point-source depth, the effect of the moon phase is taken into account in our uncertainties. In terms of follow-up observations, one might consider to observe longer at full moon especially in the bluer bands to reach a greater depth or resort to redder bands if the moon will likely affect the observations in the bluer bands adversely, but apart from that, we recommend in general to follow-up all LSNe Ia independently of the moon phase.

4.7.2 Source and lens redshifts

The mock system we investigated in Sects. 4.3 and 4.5 has $z_\mathrm{s} = 0.76$, which roughly corresponds to the median source redshift of the OM10 catalog. Furthermore, we have learned from Sect. 4.6.2 that the observational noise is the dominant source of uncertainty and we therefore expect a large dependence of the time-delay measurement on z_s (assuming a fixed exposure time during observations).

We therefore investigate in this section $z_\mathrm{s} = 0.55$ and $z_\mathrm{s} = 0.99$, which correspond to the 16th and 84th percentiles, respectively, of the source redshift from LSNe Ia in the OM10 catalog. To probe just the dependence on z_s, we leave all other parameters as defined in Table 4.1. We do not scale the absolute time delay with the source redshift, since this is just a hypothetical experiment to demonstrate how different brightnesses, related to the source redshift, influence the time-delay measurement.

The two cases are shown in Fig. 4.16, where we see the much better quality of the light curve for $z_\mathrm{s} = 0.55$ (left-hand panel) in comparison to $z_\mathrm{s} = 0.99$ (right-hand panel). Further, we also probe the lens redshift by investigating $z_\mathrm{d} = 0.16$ and $z_\mathrm{d} = 0.48$, which also corresponds to the 16th and 84th percentile of the OM10 catalog and where we also leave other parameters unchanged.

The results are summarized in Table 4.5. We see that in comparison to $z_\mathrm{s} = 0.76$, the case with $z_\mathrm{s} = 0.55$ has an improved uncertainty by ~0.2 d, where the case $z_\mathrm{s} = 0.99$ has a reduced uncertainty by ~0.7 d. This trend is expected, and means that especially for the case of $z_\mathrm{s} = 0.99$, a greater depth would improve the results significantly. Comparing the results of varying lens redshifts, we see a much smaller impact on the uncertainty. Still there is a slight trend that higher lens redshifts correspond to larger time-delay uncertainties, which is in good agreement with Huber et al. (2021), who find the tendency that microlensing uncertainties increase with higher lens redshift if everything else is fixed. The reason for this is that the physical size of the microlensing map decreases with higher lens redshift, which makes a SN Ia appear larger in the microlensing map and therefore events where micro caustics are crossed are more likely. More details are available in Huber et al. (2021).

The impact of the source redshift on the best filters to target is discussed previously in Sect. 4.6.3.

4.7.3 Data points before peak

In this section we discuss the number of data points required before peak to achieve a good time-delay measurement. The case presented in Sect. 4.3 has a large number of data points

Table 4.5. Time-delay measurement of different LSNe Ia with varying source and lens redshifts.

z_s, z_d	corresponding test set	`SNEMO15` data set
0.76, 0.252 (Fig. 4.3)	$0.04^{+0.83}_{-0.87}$ d	$0.02^{+1.38}_{-1.42}$ d
0.55, 0.252 (Fig. 4.16)	$0.04^{+0.59}_{-0.67}$ d	$0.01^{+1.29}_{-1.25}$ d
0.99, 0.252 (Fig. 4.16)	$0.01^{+1.64}_{-1.66}$ d	$0.02^{+2.1}_{-2.16}$ d
0.76, 0.16	$0.04^{+0.83}_{-0.89}$ d	$-0.09^{+1.26}_{-1.30}$ d
0.76, 0.48	$0.06^{+0.97}_{-1.06}$ d	$-0.09^{+1.45}_{-1.51}$ d

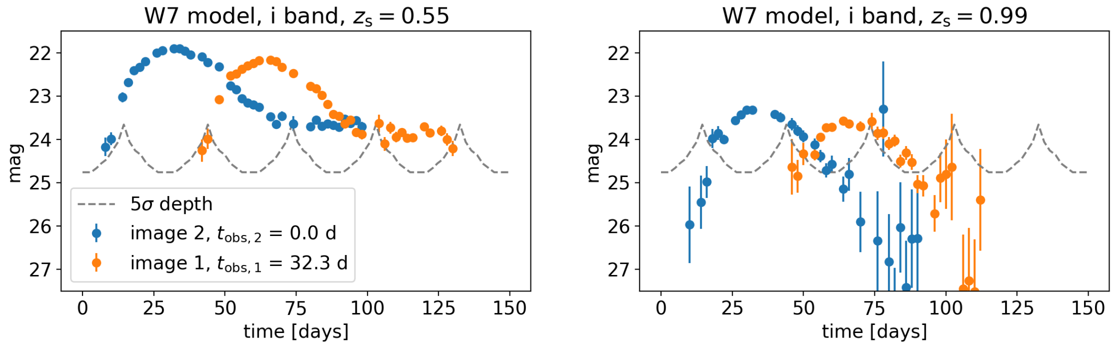


Fig. 4.16. Two LSNe Ia similar to Fig. 4.3 but with different source redshifts. The LSN Ia in the left-hand panel has $z_s = 0.55$, and the one in the right-hand panel has $z_s = 0.99$.

before peak, which is not always achievable in practice, especially since vetting of transient candidates and triggering of light-curve observations often require additional time. Therefore, we investigate a similar mock system as in Fig. 4.3, but with a later detection in the first-appearing SN image. In Fig. 4.17, we show a case where we have the first data point at the peak in the *i* band in comparison to three other cases where we have four, three, or two data points before the peak. The case for the at-peak detection provides as expected the worst precision but more worrying is the large bias of 0.83 days. Already two data points before peak improve the results significantly and allow precision cosmology for LSNe Ia with a time delay greater than 22 days. Nevertheless, we aim for four data points before peak as we could achieve a bias below 1 percent already for a delay greater than 10 days; furthermore, the precision is also improved substantially and almost at the level of the mock observation in Fig. 4.3 and corresponding results in Fig. 4.8. This would correspond in the observer frame to a detection about eight to ten days before the peak in the *i* band. Given that a SN Ia typically peaks ~18 rest-frame days after explosion and the typical lensed SN redshift is ~0.7, we would need to detect and start follow-up observations of the first-appearing SN image within ~15 days (observer frame) in order to measure accurate time delays. The results presented here are in good agreement with the feature importance investigations shown in Fig. 4.25, where we find that especially the rise slightly before the peak is very important for the RF.

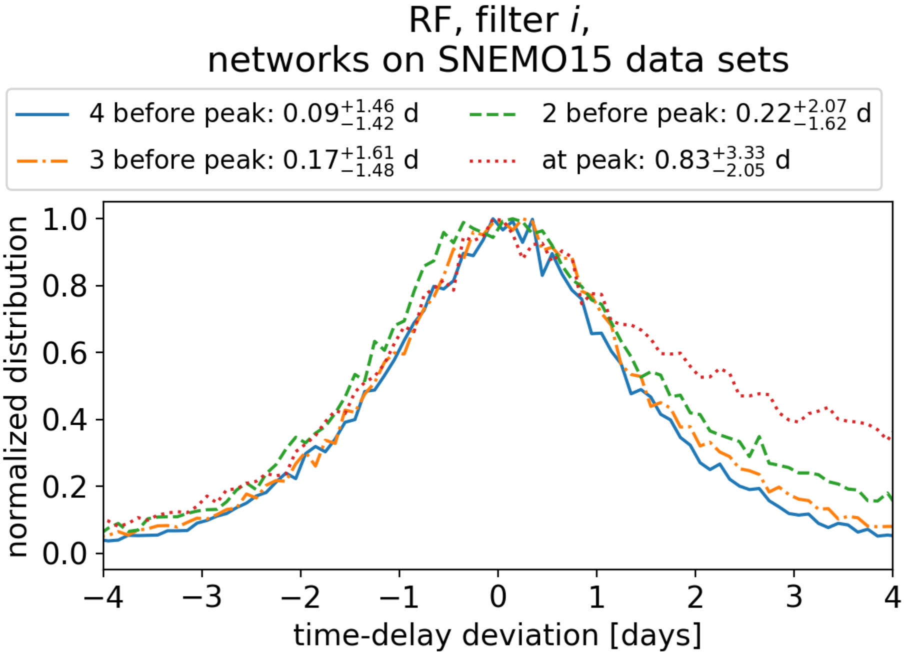


Fig. 4.17. Time-delay deviations of mock observations similar to Fig. 4.3 but with a later detection, meaning fewer data points before the peak in the *i* band of the first-appearing SN image. Each histogram is based on the whole sample of light curves from the related `SNEMO15` data set. We compare the cases where we have four, three, or two data points before the peak in comparison to an at-peak detection.

Table 4.6. Source redshift, z_s, lens redshift, z_d, convergence, κ, shear, γ, and the time values for the four images of a mock quad LSN Ia.

	z_s	z_d	(κ, γ)	t [d]
image 1	0.76	0.252	(0.435, 0.415)	≡ 0.00
image 2	0.76	0.252	(0.431, 0.424)	0.01
image 3	0.76	0.252	(0.567, 0.537)	0.34
image 4	0.76	0.252	(1.28, 1.253)	20.76

Notes. The image separation varies between 0.6 and 1.6 arcsec, and therefore it might be challenging to resolve all images with ground-based telescopes given limits due to seeing.

4.7.4 Quad LSNe Ia and higher microlensing uncertainties

So far we have only discussed double LSNe Ia, but in this section we present a LSN Ia with four images. Our mock quad LSN Ia is similar to the one presented in Sect. 4.3, but we varied the source position for the double system in the same lensing environment using the `GLEE` software (Suyu & Halkola, 2010; Suyu et al., 2012) such that we get a quad system, where the parameters are listed in Table 4.6 and the light curves from the system are shown in Fig. 4.18. For images one to three, the κ and γ values are closer to 0.5 in comparison to the double system from Table 4.1, which means that the macro magnification is higher but microlensing uncertainties are increased as shown in Huber et al. (2021). For image four, we have κ and γ values far from 0.5; this leads to lower microlensing uncertainties but therefore also to a much fainter image, which can be seen in Fig. 4.18.

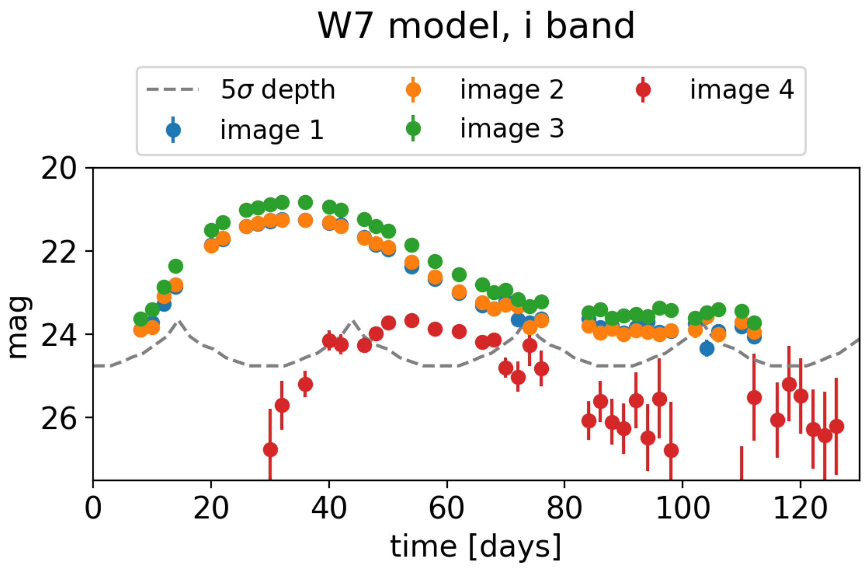


Fig. 4.18. Light curves of the mock quad LSN Ia from Table 4.6 for the *i* band.

In principle such a quad system can be investigated in two ways. The first approach is to train a separate RF per pair of images, leading to six RF models in total. The other way is to train a single RF for the whole quad system that takes as input magnitude values of four images instead of two images, similar to Eq. (4.4). The outputs as shown in Figs. 4.5 and 4.6 are then four instead of two time values.

The results for both approaches are summarized in Table 4.7 and the correlation plots are shown in Appendix 4.10.8. We find fewer correlations for the approach "separate RF per pair of images" than for the approach "single RF for all images," especially for the cases where the noisy fourth image is included in the time-delay measurement. This is because in the first case, six RF models are trained independently from each other, whereas the second case only uses a single RF that predicts four time values for the four images. Still the case "separate RF per pair of images" is preferred because it provides lower biases and tighter constraints. This is not surprising, as providing all the data from the four images at once is a much more complex problem to handle in comparison to training a RF for just two images. While the time-delay deviations between both approaches are almost comparable for pairs of images among the first, second and third images, for the cases where the fourth image is included, the single RF for the whole quad system performs much worse. This suggests that especially handling noisy data can be treated better in the approach of a separate RF for each pair of images and therefore it is always preferred to train a separate RF per pair of images.

In the following we analyze the different uncertainties of the time-delay measurements from different pairs of images as shown in Table 4.7. The most precise time delay is the one between the first and second image, but if we compare this uncertainty to the uncertainty of the right-hand panel of Fig. 4.8 for the double LSNe Ia from Fig. 4.3, we see that the precision is 0.2 days worse. This can be easily explained by the higher microlensing uncertainties coming from the κ, and γ values much closer to 0.5 as shown in Table 4.6 in comparison to

Table 4.7. Deviations of the time-delay measurements ($\tau_{ij} = \Delta t_{ij} - \Delta t_{\mathrm{true},ij}$) for the LSN Ia quad system shown in Fig. 4.18.

	separate RF per pair of images	single RF for all images
Time-delay dev. of Δt_{21}	$0.01^{+1.63}_{-1.63}$ d	$-0.01^{+1.65}_{-1.64}$ d
Time-delay dev. of Δt_{31}	$-0.05^{+1.85}_{-1.85}$ d	$0.01^{+1.89}_{-1.87}$ d
Time-delay dev. of Δt_{41}	$0.15^{+1.84}_{-1.96}$ d	$0.26^{+2.24}_{-2.36}$ d
Time-delay dev. of Δt_{32}	$-0.03^{+1.86}_{-1.89}$ d	$0.01^{+1.89}_{-1.88}$ d
Time-delay dev. of Δt_{42}	$0.14^{+1.81}_{-1.93}$ d	$0.26^{+2.25}_{-2.33}$ d
Time-delay dev. of Δt_{43}	$0.15^{+2.07}_{-2.19}$ d	$0.25^{+2.40}_{-2.55}$ d

Notes. The second column shows the case where a separate RF is trained per pair of images, leading to six RF models in total, in comparison to a single RF (third column) for the whole quad system.

Table 4.1. Higher microlensing uncertainties are also the reason why uncertainties of Δt_{31} and Δt_{32} are larger than Δt_{21}, even though the third image is the brightest one and therefore has the lowest amount of observational noise. The precision and also accuracy of the time-delay measurement where image four is involved are the worst in Table 4.7, which is explained by the very poor quality of the light curve from the fourth image. We further see that Δt_{31} and Δt_{32} as well as Δt_{41} and Δt_{42} have very similar uncertainties, which is expected since light curves from image one and two are almost identical and therefore this is a good check of consistency.

Even though the time-delay measurements between the first three images have the lowest time-delay deviation in days, the absolute time delay is very short, which leads to a very high relative deviation. For this specific mock quad LSNe Ia, it would only make sense to measure time delays with respect to the fourth image, where we would achieve a precision around 10 percent and an accuracy of 0.7 percent.

4.8 Discussion

We train a FCNN with two hidden layers and a RF using four theoretical SN Ia models, to measure time delays in LSNe Ia. We find that both ML models work very well on a test set based on the same four theoretical models used in the training process, providing uncertainties around 0.7 to 0.9 days for the *i* band almost without any bias. Applying the trained ML models to the `SNEMO15` data set, which is composed of empirical SN Ia light curves not used in the training process, we find that the uncertainties increase by about 0.5 days, but this is not surprising as such light curves have never been used in the training process and a measurement with a 1.5-day uncertainty on a single band is still a very good measurement.

However, when applied to the `SNEMO15` data set, the FCNN yields biased results. The biases are mostly within 0.4 days, but larger ones are also possible, making our FCNN approach not suitable for precision cosmology. Furthermore, this shows that the generalizability to light-curve shapes not used in the training process is not working for our FCNN approach, since biases on the corresponding test set composed of four theoretical models as used in the training process are negligible. This was already suggested by results presented in Fig. 4.21, where the training on three theoretical models was not general enough to perform well on the fourth model not used in the training process. However, we introduced random shifts in time of the light curves, which reduced the bias significantly and motivated us to apply our FCNN trained on four theoretical models (with random shifts in time to reduce the bias), to the `SNEMO15` data set as a final test, where we find unfortunately significant biases. Deeper and larger fully connected networks will not solve this problem as they will just fit the training data better and do not guarantee the generalizability. To overcome this, regularization, dropout or uncertainty estimates as additional output to the time values might help. However, this would be some kind of fine tuning to our `SNEMO15` data set, because our investigations up to that stage (which shows that our FCNN, trained using three theoretical models, generalizes well to a test set composed of four theoretical models, with negligible resulting biases), were very encouraging to apply our FCNN to the final test (`SNEMO15` data set), which it failed. However, we defer further investigations of FCNN to future work, especially since more complex ML approaches such as recurrent neural networks or long short-term memory networks (Sherstinsky, 2020) might fit the problem even better.

The RF provides significantly lower biases on the `SNEMO15` data set – with 4 or more data points before peak, which means a detection of the first LSNe Ia image about eight to ten days before peak, the bias can be kept within 0.10 days. If one of the images is very faint as shown in Fig. 4.18, we still can reach an accuracy of 0.15 days and therefore a delay longer than 15 days provides already a time-delay measurement better than 1 percent. Given the low bias in the RF especially in comparison to the FCNN, the RF is the one to use for a real application.

Huber et al. (2019) used the free-knot spline estimator from `PyCS` (Tewes et al., 2013; Bonvin et al., 2016) to measure time delays for LSNe Ia. To compare this approach to our results, we apply `PyCS` as used in Huber et al. (2019) to the `SNEMO15` data set. For the system shown in Fig. 4.3 with a very well sampled light curve, we achieve similar uncertainties as the RF shown in Fig. 4.8. However, as soon as we look at cases, where we have a reduced number of data points before peak as shown in Fig. 4.19 (in comparison to the RF results in Fig. 4.17), we see that the RF approach achieves a much higher precision. In terms of the bias, as long as we provide two data points or more before peak the RF and `PyCS` provide sufficient results. For the case where the first data point is at the peak of the *i* band, even though `PyCS` provides a much better bias than the RF, the measurement has substantially poorer precision. Overall the RF works better to measure time delays in LSNe Ia in most cases in comparison to `PyCS`. However in a real application, both approaches could be used to cross-check the time-delay measurements.

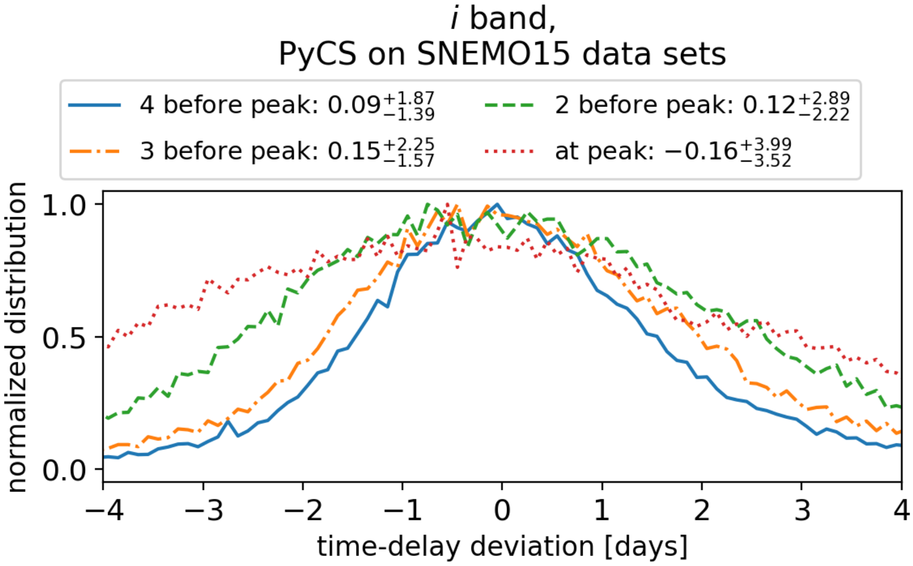


Fig. 4.19. Same as Fig. 4.17 but this time using PyCS on all samples from the `SNEMO15` data set.

4.9 Summary

In this work we have introduced two ML techniques, namely a deep learning network using a FCNN with two hidden layers and a RF to measure time delays of LSNe Ia. We simulated LSN Ia light curves for the training process, including observational noise and microlensing uncertainties using four different theoretical models. Our training set is composed of 400000 LSNe Ia coming from 4 theoretical models, 10000 microlensing map positions, and 10 noise realizations. Our test set has a size of 40000 LSNe Ia, and we drew 1000 microlensing map positions instead of 10000 as for the training set. We constructed a further data set based on the empirical `SNEMO15` model to create realistic LSN Ia light curves not used in the training process to check if our approach is general enough to handle real observations of LSNe Ia. To add microlensing to the `SNEMO15` model, we used the microlensed light curves from the theoretical models, subtracting the macrolensed light curve to get the microlensing contribution.

To summarize our results, we looked at the more realistic results from the empirical `SNEMO15` data set. From the investigation of the RF and the FCNN, we find that only the RF provides sufficiently low bias and is therefore the approach to use in a real application. From all investigated systems where we assumed a two-day cadence with a few random gaps, we found that we can achieve an accuracy better than 1% for the RF if we restrict ourselves to LSN Ia systems with a delay longer than 15 days, where we obtain the first data point around eight to ten days before the peak in the light curve of the first-appearing SN image. In terms of precision, we can achieve an uncertainty of 1.5 days from the *i* band alone for the median source redshift ~0.76 of LSNe Ia in OM10. Using three bands where the time delay is measured separately for each RF and combined afterward, we can reach an approximately 1.0-day uncertainty. The three most promising filters to target are g, r, and i for $z_s \lesssim 0.6$ and r, i, and z for higher source redshifts. As a fourth and fifth band, the z and y for $z_s \lesssim 0.6$ and the g and y for $z_\mathrm{s} \gtrsim 0.6$ might be considered. We find that the gain from multiple filters is the best if a ML model is trained individually per band. The other bands investigated in this work (u, J, and H) provide very poor-quality light curves and are therefore not useful.

From our investigations, we find mainly that the observational noise is the dominant source of uncertainty in measuring time delays, and to improve the results presented here, a greater depth would be required. The depth we assumed for follow-up observations is one magnitude deeper than the single-epoch LSST-like 5σ depth, meaning 25.7, 25.3, 24.7, 23.8, and 23.0 for g, r, i, z, and y, respectively. From the investigation of the source redshifts, we find that in comparison to the median source redshift of ~0.76 of LSNe Ia in OM10, $z_\mathrm{s} = 0.55$ can improve the precision in the i band by 0.2 days, but $z_\mathrm{s} = 0.99$ might lower the uncertainty by 0.7 days, which suggests that, especially for higher source redshifts, a greater depth might be required. Although a greater depth could also compensate for the moon phase, the impact on the uncertainty is weaker (an at most 0.4 days worse uncertainty in our investigation) and becomes even less relevant the redder the bands are. We further find that typical uncertainties in the microlensing parameters (κ, γ, and s) are not relevant for our training process. Only a significantly overestimated s value would lead to an underestimation of the uncertainties. Furthermore, we find that our approach works best if an individual RF is trained per pair of images.

In comparison to the free-knot spline estimator from `PyCS` (Tewes et al., 2013; Bonvin et al., 2016) as used in Huber et al. (2019), our approach works better overall, providing an improved precision of up to ~0.8 days. We therefore can expect slightly more LSNe Ia with well-measured time delays than the number predicted by Huber et al. (2019).

In this work we have developed a new method to measure time delays of LSNe Ia. The RF provides accurate and precise time-delay measurements that are comparable with or better than current methods and is therefore an important tool to pave the way for LSNe Ia as cosmological probes. The downsides of our approach are: that a RF needs to be trained separately for each individual system's observing pattern; the dependence on the SN Ia models used in the training process; and that our approach cannot for the moment be applied to other types of LSNe. A highly promising approach to overcoming this and building a ML network that is more general is the use of recurrent neural networks or long short-term memory networks (Sherstinsky, 2020), which will be investigated in a future study.

Acknowledgements

We thank F. Courbin, S. Schuldt and R. Cañameras for useful discussions. We also would like to thank the anonymous referee for helpful feedback, which strengthened this work. SH and SHS thank the Max Planck Society for support through the Max Planck Research Group for SHS. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 771776). This research is supported in part by the Excellence Cluster ORIGINS which is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC-2094 – 390783311. DG acknowledges support from the Baden-Württemberg Foundation through the Baden-Württemberg Eliteprogramm for Postdocs. UMN has been supported by the Transregional Collaborative Research Center TRR33 'The Dark Universe' of the Deutsche Forschungsgemeinschaft. JHHC acknowledges support from the Swiss National Science Foundation and through European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (COSMICLENS: grant agreement No 787866). MK acknowledges support from the Klaus Tschira Foundation.

4.10 Appendix

4.10.1 Photometric uncertainties

The photometric uncertainty σ_1 from Eq. (4.3) is defined as:

$$\sigma_1^2 = \sigma_{\mathrm{sys}}^2 + \sigma_{\mathrm{rand}}^2, \quad \text{where} \quad \sigma_{\mathrm{sys}} = 0.005\,\mathrm{mag} \tag{4.10}$$

and

$$\sigma_{\mathrm{rand}}^2 = (0.04 - \gamma^c)x + \gamma^c x^2 (\mathrm{mag}^2). \tag{4.11}$$

The parameter γ^c varies from 0.037 to 0.040 for different filters and $x = 10^{0.4(m_{\mathrm{AB,X}} - m_5)}$, where $m_{\mathrm{AB,X}}$ is the AB magnitude in filter X from Eq. (4.2) of the SN data point and m_5 is the 5σ point-source depth (for more details, see LSST Science Collaboration (2009), Sect. 3.5, p. 67).

A magnitude value $m_{\mathrm{AB,X}}$, which is much higher (fainter) than the 5σ point-source depth m_5, can lead to unrealistic $m_{\mathrm{data,X}}$ in Eq. (4.3). Normally one would just delete such a data point, but for our ML model this is not possible, given that we always need the same number of data points as input. Therefore, to avoid such outliers in our data set used for training, validation and testing, we ensure that all magnitude values lower than m_5 cannot exceed m_5 (bright data point just due to large error) or be much fainter than this value (data point not observable). Specifically, if the SN image brightness $m_{\mathrm{AB,X}}$ (see Eq. (4.3)) is fainter than m_5 and the calculated uncertainties would lead to a $m_{\mathrm{data,X}}$ that is smaller (brighter) than m_5 or larger (fainter) than $\max_t(m_{\mathrm{AB,X}}(t)) + (\max_{\mathrm{t}}(m_{\mathrm{AB,X}}(t)) - m_5)$, we replace that data point with a uniform random number between m_5 and $m_{\mathrm{AB,X}} + (m_{\mathrm{AB,X}} - m_5)$, where the term $(m_{\mathrm{AB,X}} - m_5)$ ensures that the uniform random number can be fainter than $m_{\mathrm{AB,X}}$, but not by more than the magnitude difference between $m_{\mathrm{AB,X}}$ and m_5.

4.10.2 Light curves of mock observation in multiple bands

Fig. 4.20 shows all the bands from the mock observation discussed in Sects. 4.3, 4.5, and 4.6.

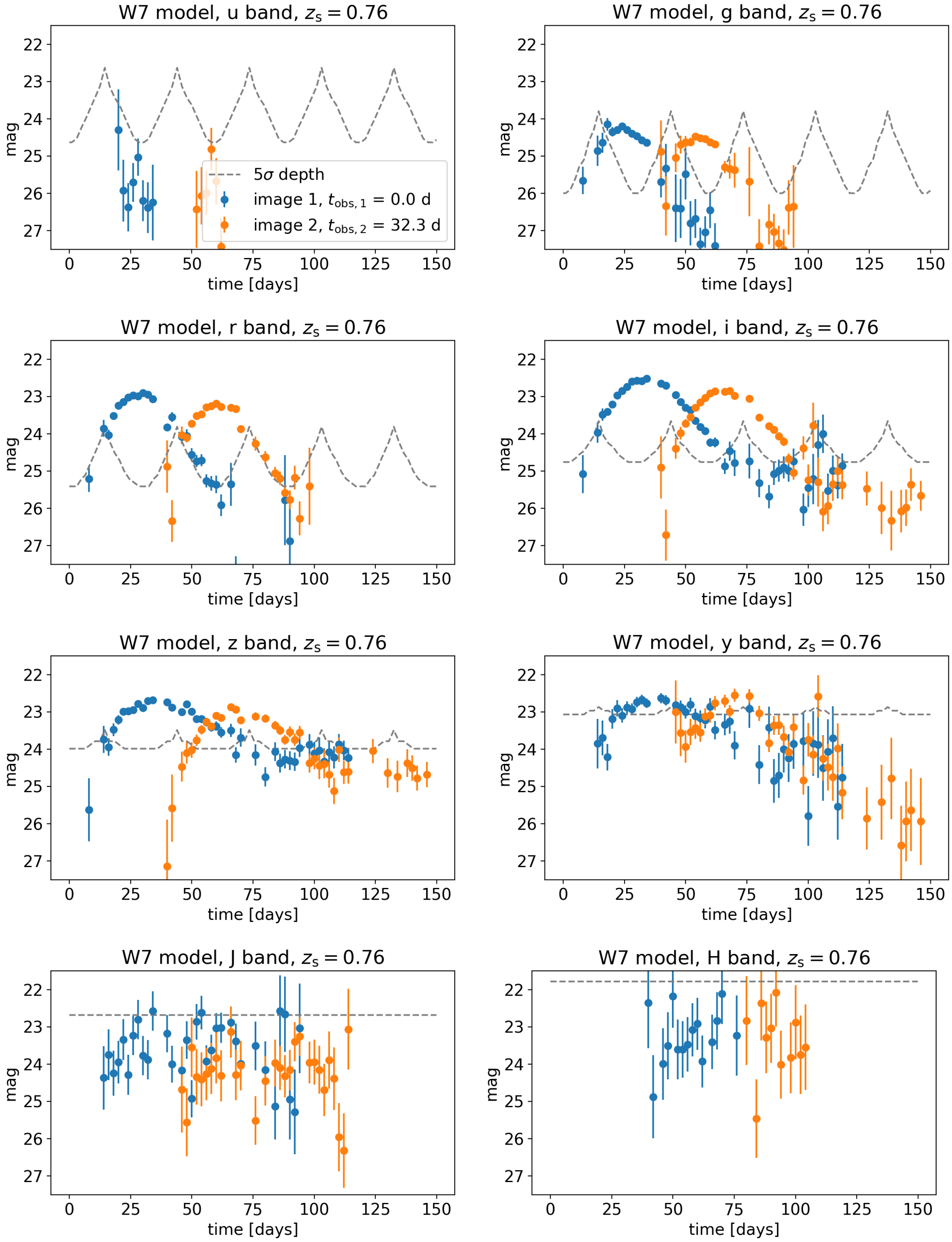


Fig. 4.20. Further bands for the mock observation of Sect. 4.3. The dashed gray line marks the 5σ point-source depth that accounts for the moon phase.

4.10.3 Bias reduction: Training on three models

In this section we discuss some hypothetical cases where we assume that only three of the four theoretical models are available for the training of the ML method.

We start with the FCNN as described in Sect. 4.4.1 and investigate four different cases namely: (merger, N100, sub-Ch), (merger, N100, W7), (merger, sub-Ch, W7), and (N100, sub-Ch, W7) for the training process. The black solid line in each panel of Fig. 4.21 shows the case where the trained model is evaluated on the corresponding test set, meaning that the test set contains light curves from all three models in the same fraction as in the training data. The other four distributions, shown in each panel, correspond to the evaluation of the FCNN trained on 3 models, on a test set that contains just light curves from a single SN Ia model. The shown results are the median (50th percentile), with the 84th-50th percentile (superscript) and 16th-50th percentile (subscript). The left column contains the normalized light curves as displayed in Fig. 4.1 and the right column contains the same models but during the training process we allow for a random shift in time from $[-5, 5]$ days and a random shift in magnitude from $[-0.4, 0.4]$. We only apply the shifts in the training set and not in the test set although the results look similar. While the shift in magnitude is only applied to make the noise level more flexible, the shifts in time help to increase the variety of the light curves, especially so that the locations of the peak from different models do overlap.

From the left column of Fig. 4.21, we learn that as soon as a model was included in the training process the network performs very well on the test set of the model. For the case of (merger, N100, sub-Ch) for the training process, the network works also very well on the test set from W7, even though it has never seen light curves from that model. The reason for that seems to be that light curves from the W7 model are close to the sub-Ch model for early times and close to the N100 model in later times (especially for the rest-frame g band) as shown in Fig. 4.1. The other three cases still work on the model it was not trained on, but we detect biases on the order of almost a day, especially for the sub-Ch and merger model (left column, rows two and four). From the right column of Fig. 4.21 we find that our applied random shifts in time and magnitude significantly help to overcome these biases. Even though this broadens the distributions of the models it was trained on slightly, it tightens the distribution of the models it was not trained on and therefore helps to generalize to real observations.

In Fig. 4.22 we do the same experiment but this time using the RF. For the cases (merger, N100, sub-Ch), (merger, N100, W7), and (merger, sub-Ch, W7) we find that applying the random shifts in the training process improves the bias and precision on the test set based on the model not used in the training process. Only for the case of using (N100, sub-Ch, W7) in the training process the merger results get worse. This suggests that especially the merger model, which deviates most from the other three models as shown in Fig. 4.1 and therefore helps to increase the variaty, is required in the training process. However, the comparison between Figs. 4.21 and 4.22 shows that the RF does already a pretty good job without the random shifts, and therefore the RF generalizes better to slightly different light-curve shapes in comparison to our FCNN approach. Nevertheless, the results from the right column of Fig. 4.21 are an encouraging sign that with the random shifts in time our FCNN approach can also be generalized better to SN Ia light curves, which were not used in the training process.

Instead of random shifts in time, one might argue that different random multiplication factors that stretch or squeeze the light curves in time might work even better. Therefore, we tested different ranges for the random factors, namely, (0.95 to 1.05), (0.9 to 1.1), (0.85 to 1.15), (0.8 to 1.2), (0.7 to 1.3), (0.6 to 1.4), (0.5 to 1.5), and (0.2 to 1.8), where the factor 1 provides the light curves as shown in Fig. 4.1. We find that (0.7 to 1.3) works best to reduce biases in a similar way as shown in Fig. 4.21. Nevertheless applying a FCNN and RF model trained on four theoretical models with the random multiplication factors to an empirical data set based on `SNEMO15` similar as in Sect. 4.5.2, we find that even though the precision for the FCNN is roughly 0.05 days better as in Fig. 4.8 the FCNN still predicts substantial biases for the `SNEMO15` data set up to 0.5 days and is therefore not useful. Furthermore, the precision of the RF on the `SNEMO15` data set for the random shifts in time between [−5, 5] days (as shown in Fig. 4.8) is roughly 0.2 days better than results from using random multiplication factors and therefore using the random shifts in time was a valuable choice.

4.10.4 Train and validation loss

In Fig. 4.23 we see for the FCNN the training loss in comparison to the validation loss for 400 training epochs. The network that provides the lowest validation loss is the one that will be stored to reduce the chance of overfitting the training data.

4.10.5 Time-delay deviation as a function of time delay

Fig. 4.24 shows the time-delay deviation as a function of time for the LSN Ia from Fig. 4.4.

4.10.6 Feature importance of FCNN and RF

In this section we investigate which of the features in Eq. (4.4) are the most important ones for the FCNN and the RF. For the FCNN the estimate is difficult and therefore we consider as an approximation just the input layer, where we calculate for each feature (input node) the mean of all the weights connected to that feature (negative weights are removed because a ReLU activation function is used). The results are summarized in the left-hand panel of Fig. 4.25, where we see that the FCNN focuses mainly on the region of the peak.

To estimate the importance of the features for the RF we use from the software `scikit-learn` the `feature_importances_` tool[35] (Pedregosa et al., 2011; Buitinck et al., 2013). This tool basically measures the decrease in performance of a RF if a specific feature would be removed. The results are shown in the right-hand panel of Fig. 4.25. Comparing these results to the FCNN, we find that the RF is mainly focusing on the rise before the peak but also the decline afterward is important. The peak itself does not matter much. We

---

[35] https://scikit-learn.org/stable/auto_examples/ensemble/plot_forest_importances.html

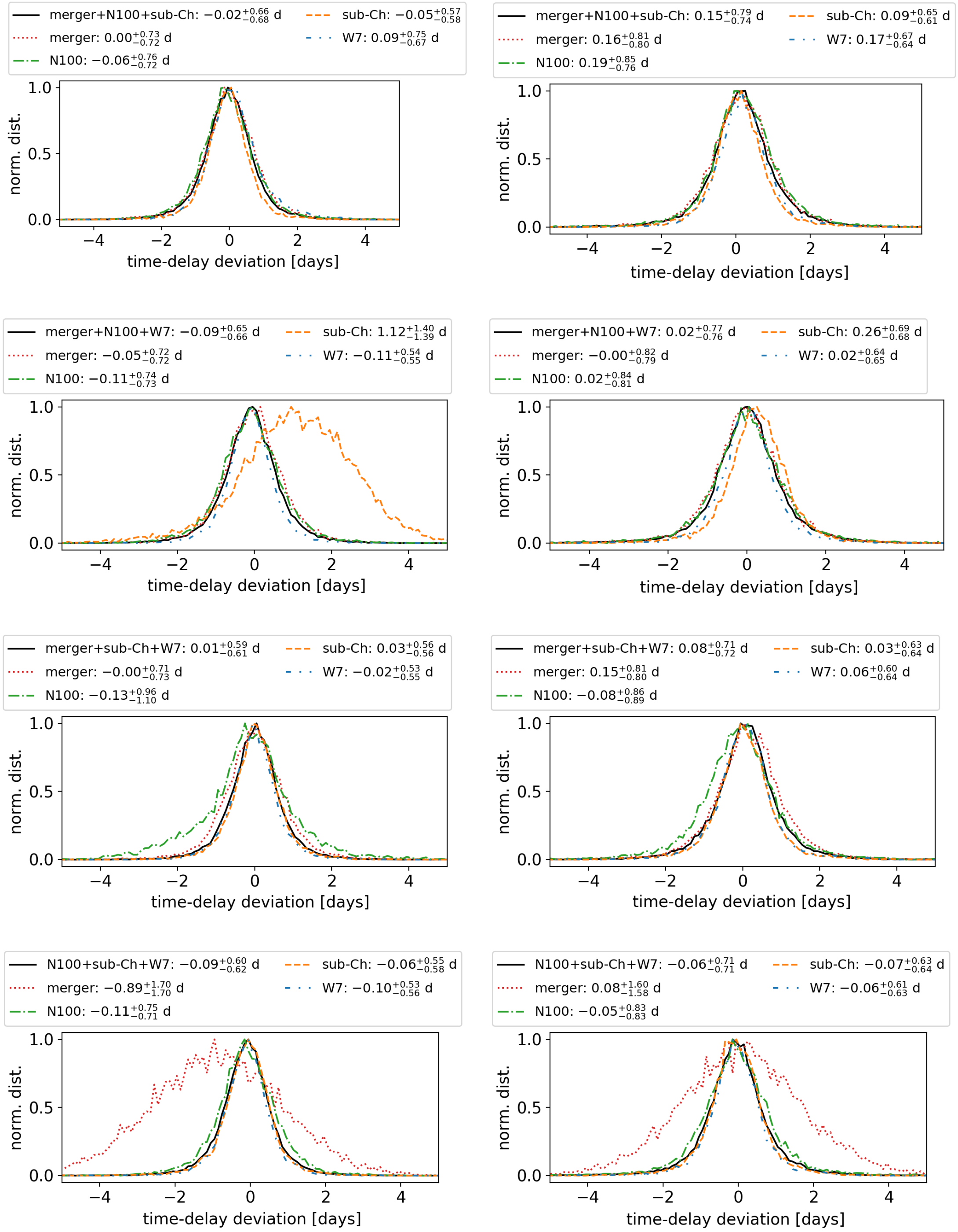


Fig. 4.21. FCNNs trained on three SN Ia models and evaluated on all samples of its corresponding test set, in comparison to the evaluation on all samples from four test sets from the four individual SN Ia models. The left column shows the case where the SN Ia models have been used as shown in Fig. 4.1. The right column contains plots in which the SN Ia models are randomly shifted within ±5d in time and ±0.4 in magnitude. The random shifts make the training set generalize better to test sets that are not represented by the training set.

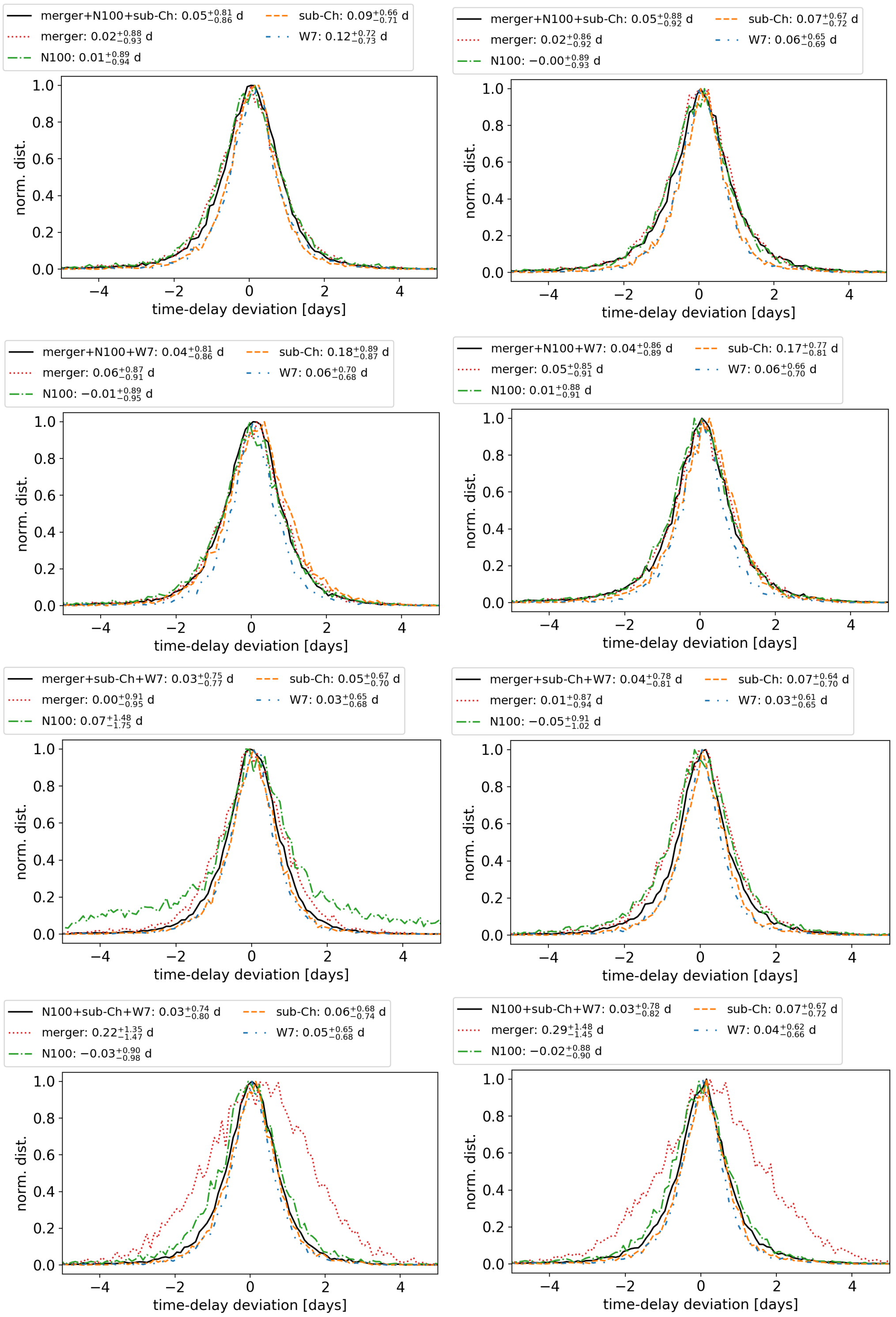


Fig. 4.22. RF networks trained on three SN Ia models and evaluated on all samples of their corresponding test set, in comparison to the evaluation on all samples from four test sets from the four individual SN Ia models. The left column shows the case where the SN Ia models have been used as shown in Fig. 4.1. The right column contains plots where the SN Ia models are randomly shifted within ±5d in time and ±0.4 in magnitude. The random shifts are far less important than for the FCNN approach shown in Fig. 4.21.

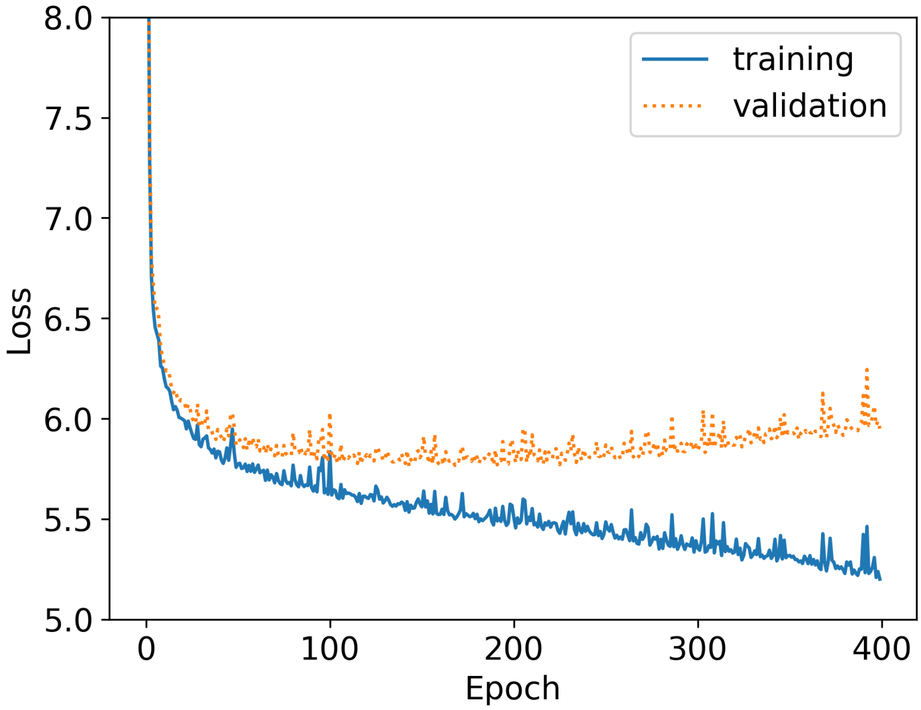


Fig. 4.23. Training and validation loss of the FCNN for 400 training epochs.

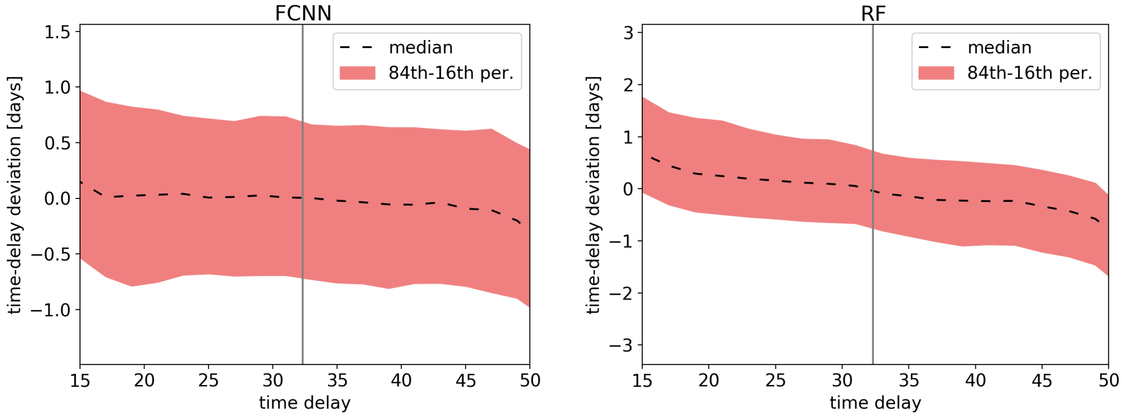


Fig. 4.24. Time-delay deviation, τ_i, from Eq. (4.8) as a function of the time delay, where we have binned up the samples from the corresponding test set with similar time delays (bin size: two days) for the LSN Ia from Fig. 4.4. The vertical gray line marks the true absolute time delay from the mock LSN Ia system. The corresponding test set has LSN Ia systems that span a range of input time-delay values, which include the vertical gray line. Results from the FCNN network are shown in the left-hand panel, and the RF is presented in the right-hand panel.

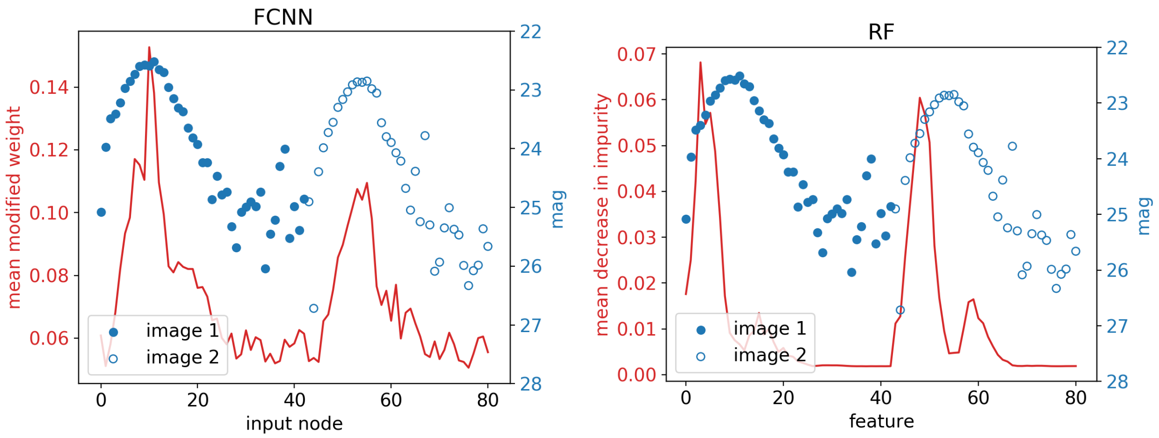


Fig. 4.25. Evaluation of the features (input nodes) for the FCNN (left-hand panel) and the RF (right-hand panel). The features (input nodes) are listed in Eq. (4.4), where feature 1 stands for $m_{i1,1}$ and the last feature (81) is $m_{i2,N_{i2}}$.

further see that the noisy part of the light curves is almost not considered in comparison to the FCNN. Focusing on higher signal-to-noise data points, less on the peak and more on the rise and decline seems to help the RF to perform better on the empirical `SNEMO15` data set in comparison to the FCNN.

4.10.7 Filters for different redshifts

Fig. 4.26 shows the performance of the RF on different filters for $z_s = 0.55$ and $z_s = 0.99$.

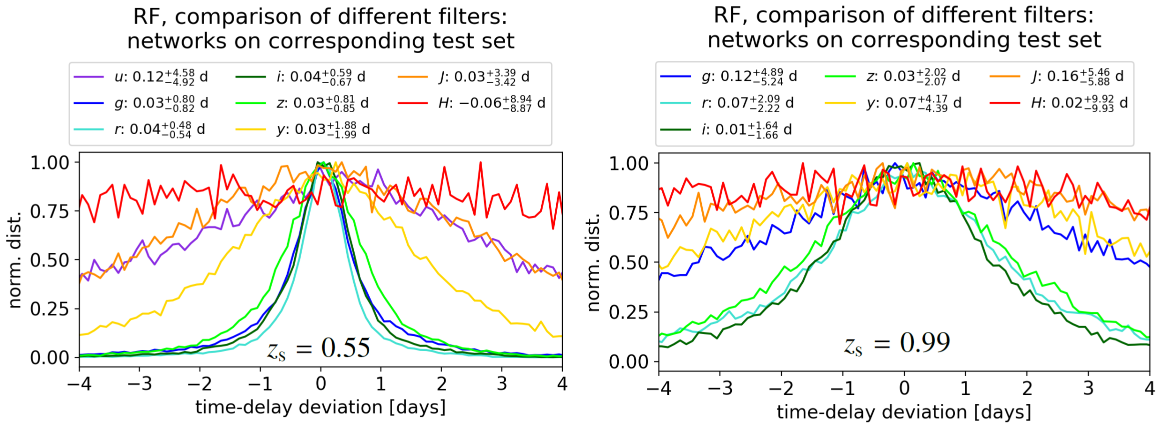


Fig. 4.26. Both panels show different RF models, each trained on a data set from a single band (as indicated in the legend) and evaluated on all samples from the corresponding test set. The left-hand panel shows the case for $z_s = 0.55$, and the right-hand panel shows $z_s = 0.99$, without the u band, which is already too faint for such a source redshift.

4.10.8 Correlation plots

Figures 4.27 and 4.28 show the correlation plots (Foreman-Mackey, 2016) for a quad system of a LSN Ia using a separate RF model per pair of images in comparison to a single RF for all images.

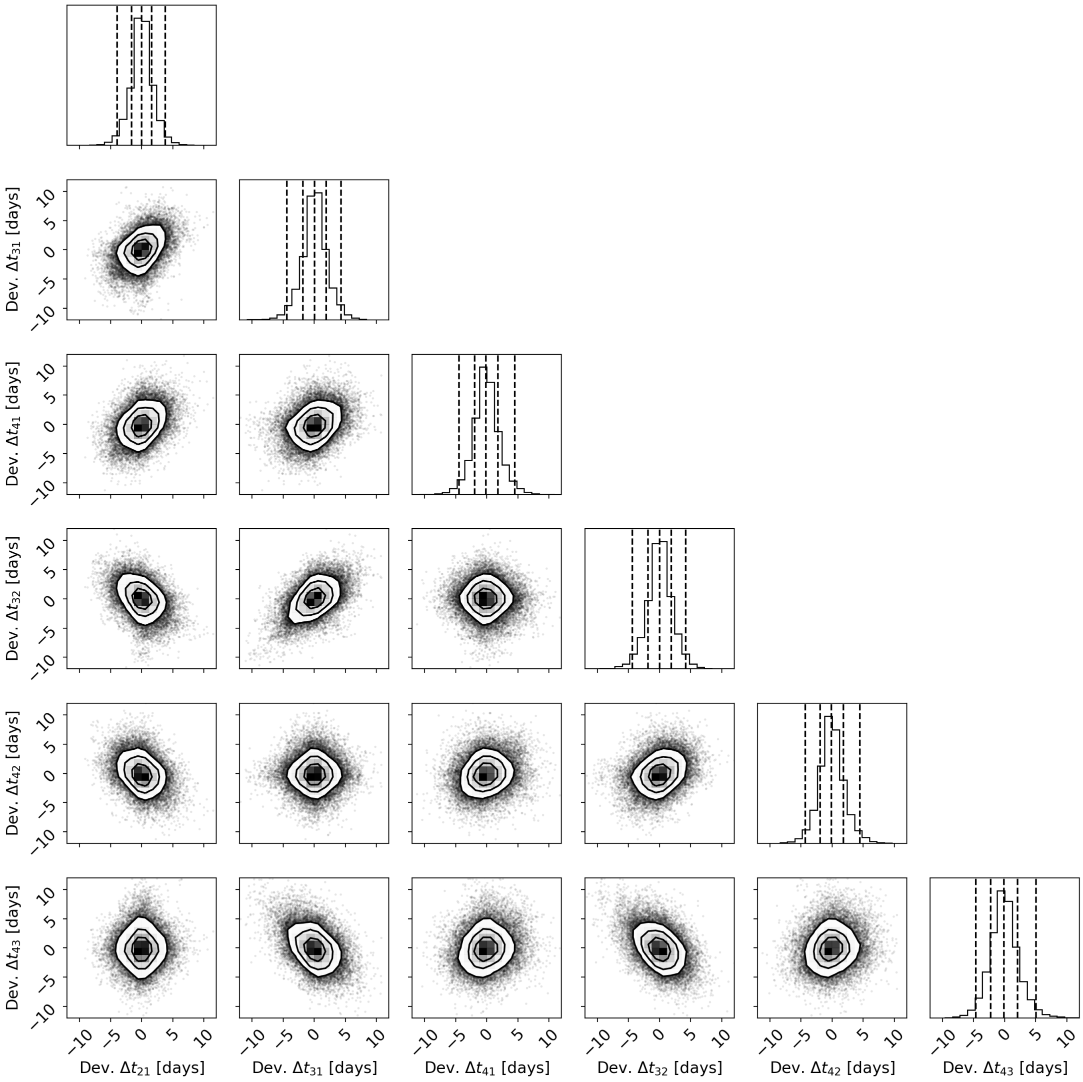


Fig. 4.27. Correlation plots using separate RF models per pair of images for the LSN Ia quad system shown in Fig. 4.18. The contour plot shows the 1σ, 2σ, and 3σ contours. The dashed lines in the histograms correspond to the median and the 1σ and 2σ range.

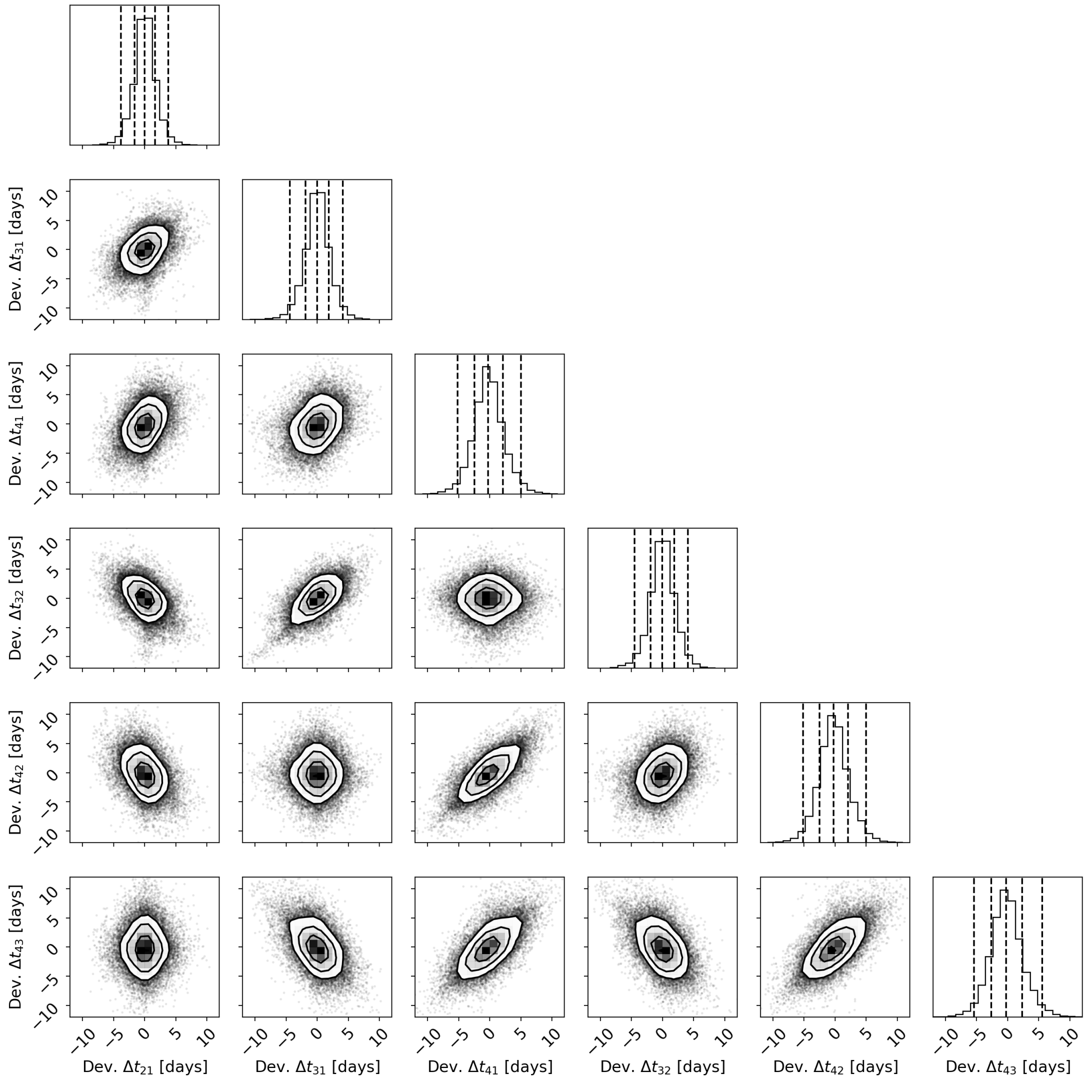


Fig. 4.28. Correlation plots using a single RF model for all images for the LSN Ia quad system shown in Fig. 4.18. The contour plot shows the 1σ, 2σ, and 3σ contours. The dashed lines in the histograms correspond to the median and the 1σ and 2σ range.

5 Time-delay measurement of strongly lensed Type Ia supernovae using long short-term memory network

Note. The content of this chapter will be the basis of a forthcoming paper that will be submitted to an international journal for publication.

From the RF investigated in Chapter 4, we expect slightly more precise time-delay measurements than `PyCS` that we used in Huber et al. (2019). The main disadvantage of the RF is that the same input structure and amount of data points are required, meaning that a model needs to be trained for each observation of a LSNe Ia individually. Therefore, we develop in this chapter a more general method that can handle LSNe Ia of almost any kind and ideally also outperforms the RF. Specifically, we investigate the long short-term memory network (LSTM), which is well suited to handle time-dependent problems (Hochreiter & Schmidhuber, 1997). Here, we investigate only LSNe Ia, although the techniques can also be applied to other types of SNe. Applications to other SN types require, however, mock microlensed light curves of these SNe, which are beyond the scope of this thesis to produce. The number of LSNe II is slightly higher than that of LSNe Ia, making them also promising for time-delay cosmography. A study on LSNe IIP using spectra and color curves for the time-delay measurement was done by Bayer et al. (2021).

5.1 Data set for machine learning

In this section, we explain the production of the data set to train, validate, and test our LSTM model. The calculation of microlensed light curves with observational noise is described in Sect. 2.3. In the following, we explain the production of the data set for the LSTM network, which is more general than the RF from Chapter 4 since LSTM's architecture can handle most LSNe Ia expected from OM10 without requiring specific training for a particular observational pattern. Our training set comprises a sample of 200000 LSNe Ia, and the validation set and test set have a size of 20000. A single sample S contains all observed data points for each of the two images of a LSNe Ia, where a single data point k of image j is represented by

$$S_{jk} := \{t_{jk}, m_{jk}, \sigma_{jk}\}, \tag{5.1}$$

where t_{jk} is the time when the magnitude m_{jk} of image j is observed with the 1σ uncertainty of σ_{jk}. We investigate only the i band, which is, according to Fig. 4.13, the most promising band. Therefore, we drop the filter notation as introduced in Eq. (2.106), and the single sample S, containing the two images (S^*_1 and S^*_2) of a single LSNe Ia, can be written as

$$S := (S^*_1; S^*_2) := (S_{11}, S_{12}, S_{13}, ..., S_{1N_{\mathrm{sl},1}}; S_{21}, S_{22}, ..., S_{2N_{\mathrm{sl},2}}), \tag{5.2}$$

where $N_{\mathrm{sl},1}$ and $N_{\mathrm{sl},2}$ is the number of observed data points as in Eq. (5.1) for images one and two, which we also refer to as sequence length. If we consider multiple samples, we label them by l leading to $S_l = (S^*_{1,l}, S^*_{2,l})$ with $N_{\mathrm{sl},1,l}$ and $N_{\mathrm{sl},2,l}$ for the sequence length of image one and two of the l-th sample.

A random set of LSNe Ia systems used for our training process is shown in Fig. 5.1, where the four left-hand panels show the data without normalization. The right-hand panels show the same four samples with normalization, which will be used in our machine learning network. We use all data points [as in Eq. (5.1)] from the two images as input to predict the time delay Δt between these images. The applied normalization is essential such that the input values for the LSTM network always have the same order of magnitude. Therefore, we normalize the magnitude of both images with respect to the peak of the image under consideration. Given that very noisy data points can be brighter than the peak, we only consider 25 % of the data points with the lowest uncertainty to find the peak (minimum in magnitude). We then subtract the peak magnitude from each data point in the corresponding light curve such that the peak has a magnitude value of zero. Furthermore, we also normalize the time scale t, which in our case is the time after the explosion, by adding a constant offset to t such that the offset $t = 0$ corresponds to the peak determined for the first image. We then divide the time delay Δt and time scale t by 150 days, which is the maximum of the time delays under consideration, such that the output of the LSTM network is restricted to values between 0 and 1.

In the following, we explain how such a random data set is created, where we summarize our assumptions in Table 5.1. To simulate light curves for LSNe Ia, we use the four theoretical models described in Sect. 2.3.1.1 and the empirical `SNEMO15` model (Saunders et al., 2018), which contains 171 SN Ia light curves. To ensure that we have light curves not used in the training process to test the generalizability, we randomly split the whole `SNEMO15` data set into three subsets: (1) `SNEMO15-T` consisting of 87 curves for training, (2) `SNEMO15-V` consisting of 42 curves for validation, and (3) `SNEMO15-E` consisting of 42 curves for testing. Our training, validation, and test set is then composed of 50% light curves from theoretical SN models and 50% of empirical light curves from the corresponding subset of `SNEMO15` light curves. Furthermore, we construct a second test set to investigate the generalizability, namely the `SNEMO15`-only test set, based only on the 42 light curves in `SNEMO15-E`. A summary of all data sets is listed in Table 5.2. Our underlying light curves from theoretical and empirical models cover a range from ~3.4 to 70.0 rest-frame days after explosion.

For each theoretical model, we calculate for a given LSN Ia image, 10000 microlensed light curves from the corresponding microlensing map (set by κ, γ, s, z_s, z_d), as described in Sect. 2.3.2 and Sect. 4.2.1. Since the calculation of microlensed SN Ia light curves is very time-consuming, we only investigate a subset of microlensing magnification maps, while covering a broad range of properties of the LSN Ia images. The left-hand panel of Fig. 5.2 shows the distribution of (κ, γ) values from the OM10 catalog and the six pairs used in this work, which are composed of the median and two systems on the 1σ contours, taken separately for the lensing image types "minimum" and "saddle". Furthermore, we assume three typical s values of 0.4, 0.6, and 0.8 similar to Huber et al. (2019, 2021, 2022) and five different pairs of $(z_\mathrm{s}, z_\mathrm{d})$ as shown in the right-hand panel of Fig. 5.2 also based on the OM10

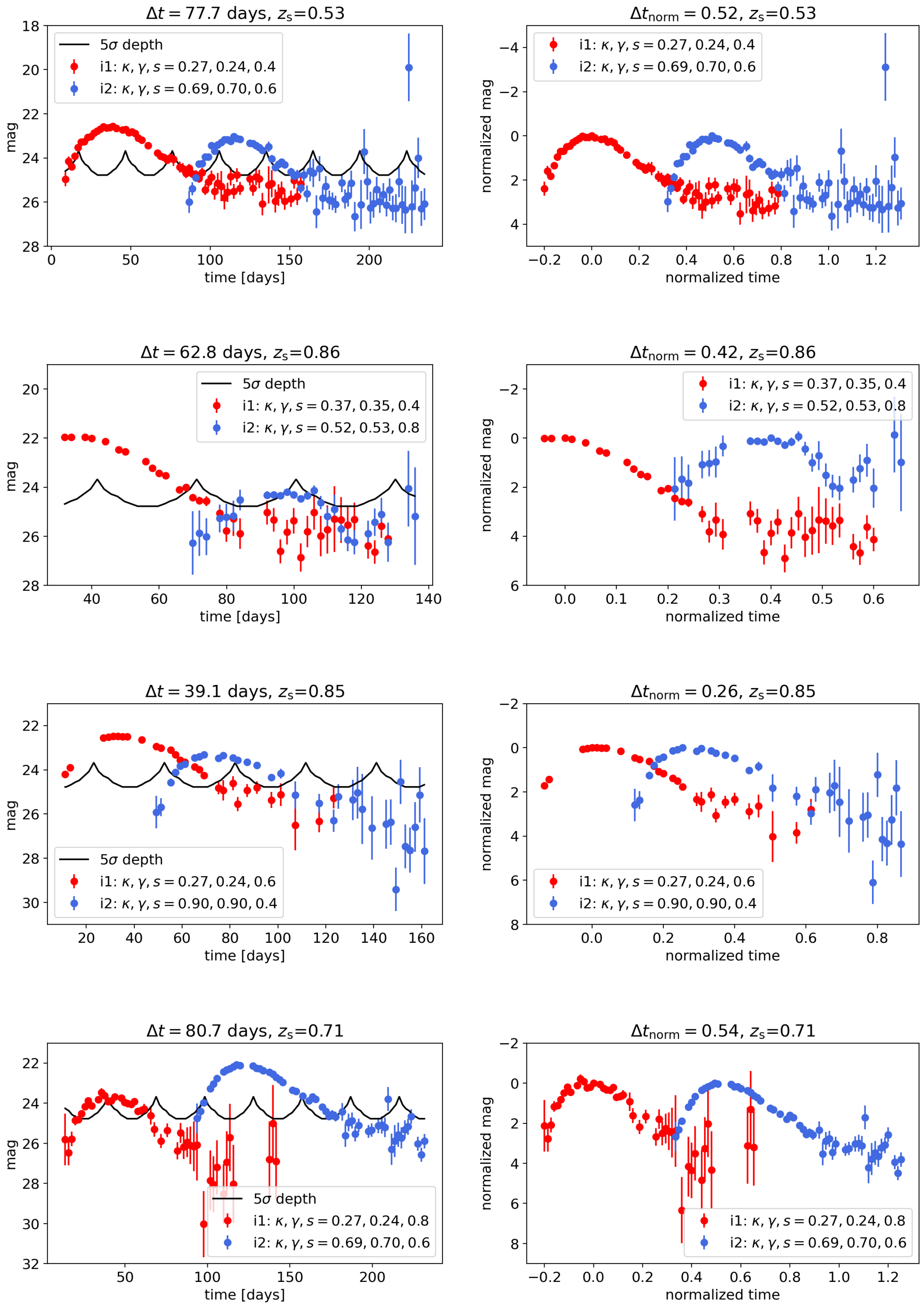


Fig. 5.1. Four random samples of our training set, comparing data without normalization (left-hand panels) and data with normalization (right-hand panels). We will use the normalized data for our machine-learning network. The red and blue data points are the i-band light curves for the first and second SN images. The black curves in the left-hand panels show the i-band 5σ limiting depth, which oscillates with the moon phase. The normalization procedure to obtain the right-hand panels from the left-hand panels is described in Sect. 5.1.

Table 5.1. Summary of our assumptions for the LSTM data set used for training, validation, and testing.

theoretical SN Ia models	W7, N100, Sub-Ch, merger
empirical SN Ia model	`SNEMO15`
light curve coverage	~3.4 to 70.0 rest-frame days after explosion
time delay	0.1 to 150 days
detection of first image	before peak
cadence	2 days
missing data points	0 to 50 %
band / filter	i
mean single-epoch 5σ depth	24.5 mag
max. σ_{jk}	2 mag
moon phase	random
z_{s}	0.17 to 1.34 following OM10
$(z_{\mathrm{s}}, z_{\mathrm{d}})$ just for microlensing	(0.77, 0.30), (0.97, 0.47), (0.53, 0.13), (0.94, 0.20), (0.60, 0.42)
(κ, γ)	(0.37, 0.35), (0.27, 0.24), (0.46, 0.48), (0.69, 0.7), (0.90, 0.90), (0.52, 0.53)
s	0.4, 0.6, 0.8

Table 5.2. Description of the training, validation, test, and `SNEMO15`-only test set used in this chapter. The `SNEMO15` light curves are split randomly into three distinct subsets with no overlap in light curves: `SNEMO15-T` for training, `SNEMO15-V` for validation and `SNEMO15-E` for testing. This ensures that light curves used in one process (e.g., validation) are not used in another process (e.g., training or testing).

Training set	50% light curves from theoretical SN Ia models + 50% light curves from `SNEMO15-T`
Validation set	50% light curves from theoretical SN Ia models + 50% light curves from `SNEMO15-V`
Test set	50% light curves from theoretical SN Ia models + 50% light curves from `SNEMO15-E`
`SNEMO15`-only test set	100% light curves from `SNEMO15-E`

catalog. Therefore we investigate in total $6 \times 3 \times 5 = 90$ different microlensing magnification maps where we calculate for each map 10000 random source positions for all four theoretical models providing $90 \times 10000 \times 4 = 3600000$ microlensed SN Ia light curves. This is more than sufficient given that we found in Fig. 4.12 that microlensing is not the dominant source of uncertainty. For the empirical `SNEMO15` model, no specific intensity profiles are available, and therefore the calculation of microlensed light curves as described in Sect. 2.3.2 is not possible. To nonetheless include microlensing effects, we use the theoretical models where we subtract from a randomly selected microlensed light curve (which include macrolensing plus microlensing) the macrolensed light curve, assuming $\mu_{\rm macro} = 1/((1-\kappa)^2 - \gamma^2)$, to get only the contribution from microlensing which we add to a `SNEMO15` light curve (see Sec. 4.5.2). With the six pairs of κ and γ, we only cover a small part of potential macrolensing factors; however, from Fig. 5.3, we see that the microlensed light curves from these six pairs span a much broader range of potential magnification factors as we would get from macrolensing only, from the whole OM10 sample.

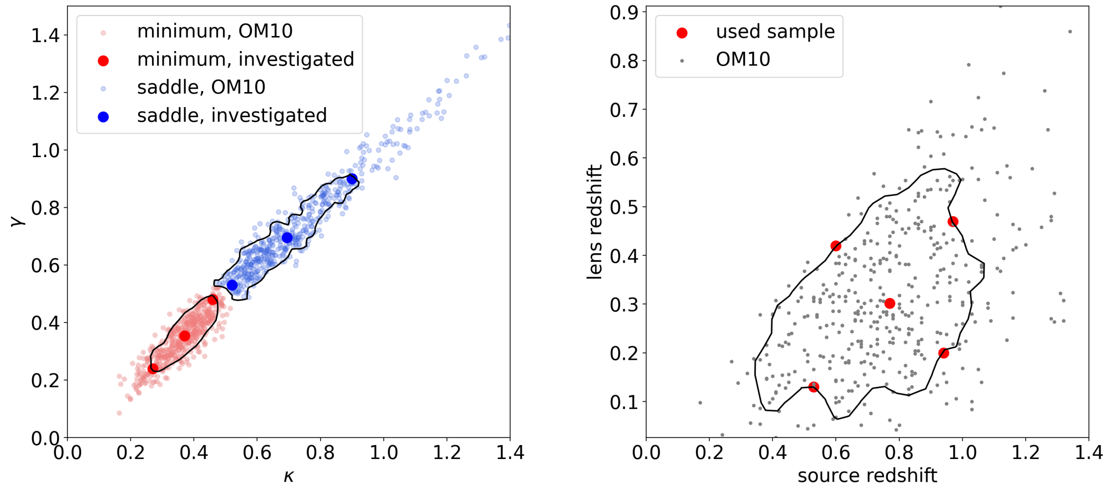


Fig. 5.2. Convergence κ and shear γ (left-hand panel), and source redshift $z_{\rm s}$ and lens redshift $z_{\rm d}$ (right-hand panel) used for the calculation of microlensed light curves drawn from the OM10 catalog. The black lines represent contours that enclose 68% of the LSNe Ia systems in the OM10 catalog.

The source redshift and the lens redshift set the scale of the microlensing map and only weakly affect the microlensing effect on SN Ia [see Fig. 3.8]. However, the source redshift also determines the observed light curve shapes and brightness in a certain band. In terms of noise contributions, we showed in Chapter 4 that observational noise is far more important than microlensing. Therefore, the five source redshifts shown in Fig. 5.2 are sufficient for the microlensing calculation but not for the various brightness and light curve shapes coming from the OM10 sample, which include $z_{\rm s}$ between 0.17 and 1.34. Therefore, we draw randomly pairs of ($z_{\rm s,OM10}$, $z_{\rm d,OM10}$) from the OM10 catalog, and we pick from the right-hand panel of Fig. 5.2 the pair ($z_{\rm s,micro}$, $z_{\rm d,micro}$) from our used sample (in red) which is closest. We then rescale the pre-calculated microlensed flux $F_{\lambda,{\rm o},z_{\rm s,micro}}$ for the source

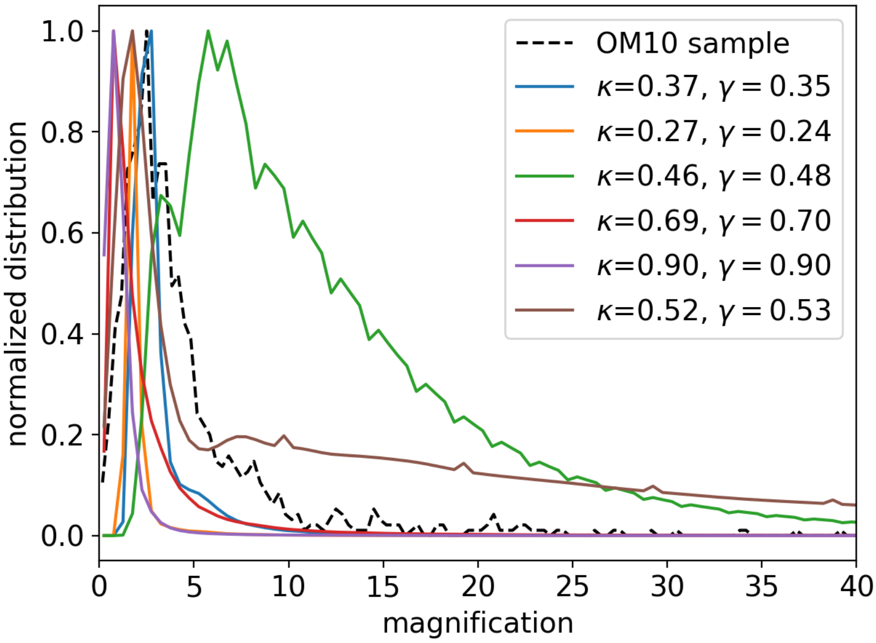


Fig. 5.3. Distribution of magnification factors from the six microlensed magnification maps under investigation in this chapter (solid colored), in comparison to all macro-magnification factors from all LSNe Ia from the OM10 catalog (dashed black). Microlensing magnification maps investigated in this work cover a much broader range of potential magnification factors compared to the macrolensing-only case from the whole OM10 sample.

redshift $z_{\mathrm{s,micro}}$ to $z_{\mathrm{s,OM10}}$, the random source redshift of interest [see in comparison to Eq. (2.103)]:

$$F_{\lambda,\mathrm{o},z_{\mathrm{s,OM10}}} = F_{\lambda,\mathrm{o},z_{\mathrm{s,micro}}} \frac{D^2_{\mathrm{lum},z_{\mathrm{s,micro}}}(1+z_{\mathrm{s,micro}})}{D^2_{\mathrm{lum},z_{\mathrm{s,OM10}}}(1+z_{\mathrm{s,OM10}})}. \tag{5.3}$$

Further, we rescale as well the wavelength [see Eq. (2.28)]

$$\lambda_{z_{\mathrm{s,OM10}}} = \frac{\lambda_{z_{\mathrm{s,micro}}}}{1+z_{\mathrm{s,micro}}}(1+z_{\mathrm{s,OM10}}) \tag{5.4}$$

and time since explosion [see Eq. (2.31)]

$$t_{z_{\mathrm{s,OM10}}} = \frac{t_{z_{\mathrm{s,micro}}}}{1+z_{\mathrm{s,micro}}}(1+z_{\mathrm{s,OM10}}). \tag{5.5}$$

With Eq. (2.101), we can then calculate the microlensed light curves for $z_{\mathrm{s,OM10}}$, providing the right light curve shape and brightness but only approximated microlensing contributions.

For our data set, we further make assumptions on the follow-up strategy, namely a cadence of 2 days to a depth 1 mag deeper than the mean LSST single-epoch 5σ depth. This provides a good compromise between required time on a hypothetical 2m telescope on the one hand and accurate/precise time delays in LSNe Ia systems, on the other hand (Huber et al., 2019). In reality, we will not be able to observe every second day because of, e.g., bad weather. Therefore, we delete randomly between 0 to 50% of the times follow-up observations are assumed. Furthermore, we limit the discoveries of the first image of a LSNe Ia to before or at peak detections, which is the aim of our lens-finding approaches. We train the network in the *i* band, given that this provides the most precise time-delay measurement (see Fig. 4.13). From Eq. (2.106), we see that extremely high uncertainty values are possible, which occur only rarely in practice and can be easily removed from the data. Therefore, we will only take data into account where the standard deviation σ_{ik} of an observed data point m_{ik} is smaller than 2 mag.

The theoretical light curves used in the training process match only partly the empirical `SNEMO15` light curves as can be seen from Fig. 5.4 for typical source redshifts. To increase the variety of the light curve shapes from theoretical models, we multiply a random factor between 0.7 and 1.3 to the time after explosion to stretch and squeeze the light curves and add a random shift in magnitude between −0.4 and 0.4 to vary the brightness. This is motivated by Chapter 4, where we have shown that this helps the machine learning models best to generalize better to SN models not used in the training process.

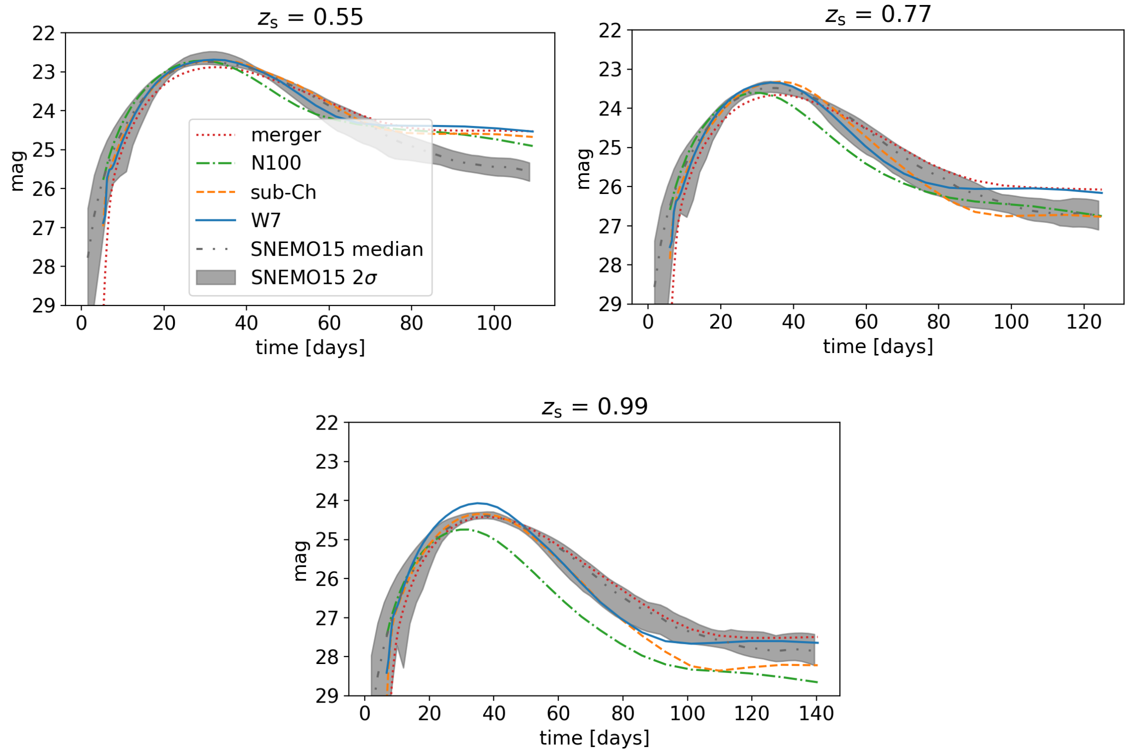


Fig. 5.4. *i* band light curves for typical source redshifts of LSNe Ia. We compare four theoretical models (merger, N100, sub-Ch, and W7) with the empirical `SNEMO15` model.

Finally, to create samples for our training, validation, or test set, we randomly pick a time delay between 0.1 to 150 days and decide if our LSN Ia light curves will be based on a theoretical model or the empirical model (including microlensing from a theoretical model) with equal probability. For the theoretical model, we decide with a 50:50 chance if we use the theoretical models as shown in Fig. 5.4 or if we apply a random stretching factor between 0.7 and 1.3 and a random shift in magnitude between −0.4 and 0.4. Further, we pick randomly for the first image a minimum $[(\kappa,\gamma) = (0.37, 0.35), (0.27, 0.24), \text{or } (0.46, 0.48)]$

and for the second image a saddle [$(\kappa,\gamma) = (0.69, 0.7)$, $(0.9, 0.9)$, or $(0.52, 0.53)$], as well as a source redshift from the OM10 sample, for which we approximate the microlensing contribution following the process described in Eqs. (5.3) to (5.5). Given that the empirical `SNEMO15` model covers just rest-frame wavelengths between 3305 Å to 8586 Å, we cannot calculate i-band light curves for $z_s > 1.0$; therefore, if the empirical model is picked for such a high z_s, we draw the redshift again until we have $z_s \leq 1.0$. Furthermore, we draw a random position in the moon phase (see Fig. 4.2) and assume the detection and follow-up strategy as described previously. For this specific random set-up, we create then four samples as described by Eq. (5.2) for the four theoretical SN Ia models, or if the empirical model is picked, we use, e.g., for the training set, one of the 87 `SNEMO15-T` light curves, where we also create four samples with four different microlensing contributions from the four theoretical models. For this specific set-up, we calculate then the microlensed SN Ia light curves with random noise following Eq. (2.106), leading to samples shown on the left-hand side of Fig. 5.1. We repeat the just described procedure by drawing new time delays until we reach the target size of a certain data set. For the training set, we use 200000 samples, and the validation and test sets contain 20000 samples.

5.2 Machine learning technique

In Sect. 5.2.1 we summarize briefly the basic concept of an LSTM network before we introduce our setup in Sect. 5.2.2.

5.2.1 Basics of long short-term memory network

The LSTM network (Hochreiter & Schmidhuber, 1997; Sak et al., 2014; Sherstinsky, 2020) is motivated by the Recurrent Neural Network (RNN), where both have the purpose of solving a time-dependent problem. For the RNN, the basic setup is illustrated in Fig. 5.5 for the k-th time step. We have an input vector x_k for which we calculate the state A_k by using the information of the previous time step $k-1$ to further calculate the output h_k. In equations, ignoring non-linear activation functions, this can be expressed as

$$A_k = w_A A_{k-1} + w_x x_k \tag{5.6}$$

and

$$h_k = w_h A_k, \tag{5.7}$$

where w_A, w_x, and w_h are the time-step independent weights that are learned in the training process via backpropagation. The main issue with RNNs is that the gradients tend to explode or vanish, which makes it very difficult to learn long-term dependencies (Pascanu et al., 2013; Philipp et al., 2018).

To overcome that issue, LSTM networks have been introduced by Hochreiter & Schmidhuber (1997), where we have, in addition to the hidden state h_k, a cell state c_k, which acts as a memory. Furthermore, the state A_k is replaced by a more complex unit as shown in Fig. 5.6, which contains the following main ingredients with $\odot$ denoting the Hadamard

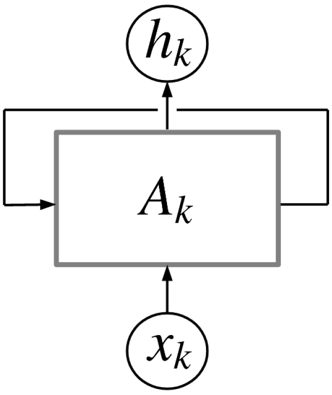


Fig. 5.5. Illustration of an RNN. From the input x_k, one calculates the state A_k of the network at the k-th time step by using the information from the previous $k-1$ time step. The output h_k is then computed from the state A_k. See Eqs. (5.6) and (5.7).

product[36]:

$$h_k = o_k \odot \tanh(c_k), \tag{5.8}$$

$$c_k = f_k \odot c_{k-1} + i_k \odot g_k, \tag{5.9}$$

$$f_k = \text{sigmoid}(w_{xf} x_k + w_{hf} h_{k-1} + b_f), \tag{5.10}$$

$$g_k = \tanh(w_{xg} x_k + w_{hg} h_{k-1} + b_g), \tag{5.11}$$

$$i_k = \text{sigmoid}(w_{xi} x_k + w_{hi} h_{k-1} + b_i), \tag{5.12}$$

$$o_k = \text{sigmoid}(w_{xo} x_k + w_{ho} h_{k-1} + b_o), \tag{5.13}$$

where the sigmoid and tanh activation functions (see Fig 2.13) act on each element of the vector. The input vector at time step k is x_k with size N_{input}, which is, in our case, three, containing the time, magnitude, and uncertainty as listed in Eq. (5.1). The hidden state h_k is a vector with size N_{hidden}. The biases b_f, b_g, b_i, and b_o are vectors with size N_{hidden}, the weights w_{xf}, w_{xg}, w_{xi}, and w_{xo} are $N_{\text{hidden}} \times N_{\text{input}}$ matrices, and the weights w_{hf}, w_{hg}, w_{hi}, and w_{ho} are $N_{\text{hidden}} \times N_{\text{hidden}}$ matrices. The biases and weights are not dependent on the time step and are learned through backpropagation in the training process. The hidden state h_k is the prediction of the LSTM network at time step k and serves as input to calculate the hidden state h_{k+1} corresponding to the time step $k+1$. The cell state c_k acts as a memory and transports information through the unit. The forget gate f_k decides when and which parts to erase from the cell state. The gate g_k removes or adds information to the cell state coming from the input gate i_k. The output gate o_k is responsible for the output / hidden state h_k of

---

[36] The Hadamard product is an element-wise product for two matrices of same dimension.

the LSTM network. Outputs are produced at every time step, but for our case, we will only use the last output at time-step N_{sl}, after the full light curve with the sequence length of N_{sl} was propagated through the network. This corresponds to a many-to-one scenario where we have N_{sl} input data points as in Eq. (5.1) and one output of size N_{hidden}. However, two images need to be taken into account, and we explain our approach to predict one time delay Δt and the uncertainty of the delay $\sigma_{\Delta t}$ in Sect. 5.2.2.

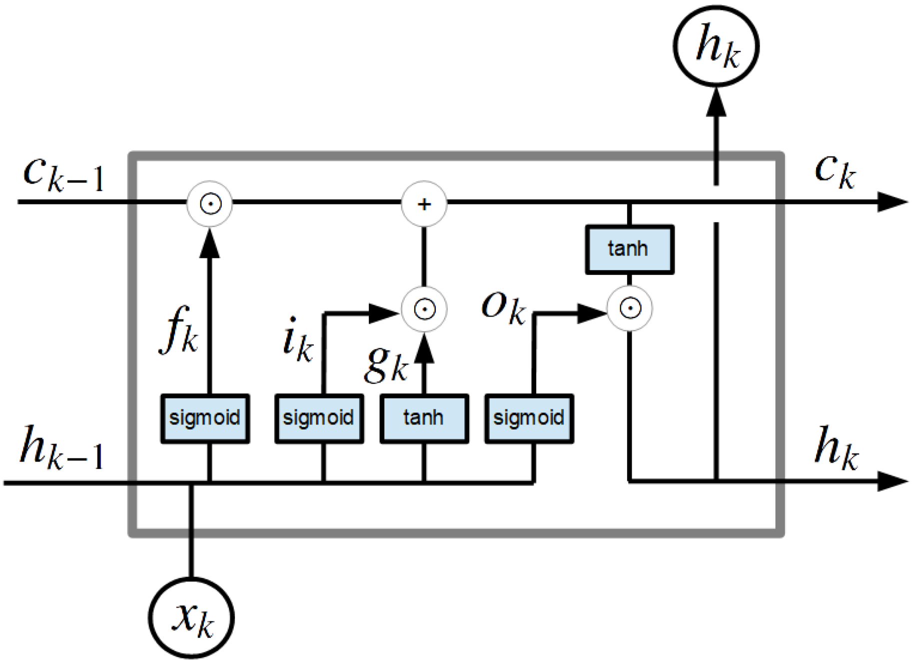


Fig. 5.6. Illustration of an LSTM. The input x_k is used in combinations with the output h_{k-1} and the cell state c_{k-1} from the previous time step $k-1$ to produce the output h_k for the k-th time step. The cell state c_k acts as a memory and transports information through the cell, where the forget gate f_k, input gate i_k, and gate g_k update information in the cell state. The output gate, based on the previous output h_{k-1} and input x_k in combination with information from the cell state, provides the new output h_k. For more information, see Eqs. (5.8) to (5.13).

5.2.2 Modified Siamese long short-term memory network

For our approach, we use a Siamese LSTM network (Bromley et al., 1993). The Siamese neural networks are generally a class of neural network architectures that contain two or more identical sub-networks, which therefore have the same weights and parameters. A Siamese LSTM is also used in the Manhattan LSTM (MaLSTM) network (Mueller & Thyagarajan, 2016) developed to find semantic similarities between sentences. In this network, two sentences with different (or the same) numbers of words are fed through the same LSTM network to predict an output for each sentence. The two outputs are then used to calculate a Manhattan distance which states the similarity between the two input sentences.

Our approach is similar to the MaLSTM, but instead of sentences, we have different images of a LSN Ia, and instead of a single word, we have the time of the observation, the magnitude, and the uncertainty of the measurement, as shown in Eq. (5.1). Our approach produces similar to the MaLSTM two outputs for the two images. Still, instead of a Manhattan distance, we use a FCNN with two hidden layers, given that we want to determine a time delay instead of a similarity score. Our full network is shown in Fig. 5.7, where the calculations of the LSTM network are listed in detail in Eq. (5.8) to (5.13), where we added a second index j to the time step k, leading to h_{jk}, c_{jk}, f_{jk}, g_{jk}, i_{jk}, and o_{jk} in order to distinguish between image one ($j = 1$) and image two ($j = 2$) of our lensed SN system. Our network takes as input the data S^*_{1k} of image one and S^*_{2k} of image two and predicts as an output, the time delay Δt and the uncertainty of the time delay $\sigma_{\Delta t}$. The sub-LSTM networks for both images are identical and composed of two layers LSTM A and LSTM B. The second layer of the LSTM network takes as input the output from the first layer $\tilde{h}_{jk}$, where a dropout rate $p_{\mathrm{dropout}} = 20\%$ is applied, meaning that during the training process, randomly 20% of the input data for the second layer ($\tilde{h}_{jk}$) is set to zero and the other values are rescaled by 1/(1-p_{dropout}). The rescaling is necessary given that the dropout is just applied in the training process but not during the validation and test phase, which would lead to systematically larger outputs for the validation and test set if one would not apply the rescaling in the training process. The dropout helps the network avoid focusing on particular neurons during training to generalize better to data not used in the training process. The last two outputs from the second layer $h_{1N_{\mathrm{sl},1}}$ for image one and $h_{2N_{\mathrm{sl},2}}$ for image two are then concatenated and used as input for the FCNN. For given j and k, the inputs S_{jk} are vectors with size three [see Eq. (5.1)] whereas the outputs from both layers $\tilde{h}_{jk}$ and h_{jk} are vectors with size $N_{\mathrm{hidden}} = 128$. Therefore, the input layer of the FCNN has 256 nodes, which we summarize in a vector d:

$$d := (h_{1N_{\mathrm{sl},1}}, h_{2N_{\mathrm{sl},2}}), \tag{5.14}$$

with dimension 256. The two hidden layers of the FCNN have 128 and 10 nodes, where at each hidden layer, a non-linearity is applied by using a rectified linear units (ReLU) activation function (e.g., Schmidt et al., 1998; Maas et al., 2013), which is the identity function for positive values and zero for all negative values. In the last layer, where final outputs are produced, a sigmoid activation is used, which predicts values between 0 and 1 and, therefore, covers the range of potential normalized time delays. The first hidden layer is represented by the vector η_1 with size 128, and the second hidden layer by the vector η_2 with size 10:

$$\eta_1 = \mathrm{ReLU}(\omega_1 d + \beta_1), \tag{5.15}$$

$$\eta_2 = \mathrm{ReLU}(\omega_2 \eta_1 + \beta_2). \tag{5.16}$$

The outputs of the network are then two scalar values, namely the normalized time delay Δt_{norm} and the normalized variance “Var”:

$$(\Delta t_{\mathrm{norm}}, \mathrm{Var}) = \mathrm{sigmoid}(\omega_3 \eta_2 + \beta_3), \tag{5.17}$$

where, ω_1 (matrix of size 128×256), ω_2 (matrix of size 10×128), and ω_3 (matrix of size 2×10) are the weights, and β_1 (vector of size 128), β_2 (vector of size 10), and β_3 (vector of size 2) are the biases. Weights and biases of the network are learned in the training process. Further, the outputs from Eq. (5.17) need to be translated to a physically meaningful Δt and σ_{Δ_t}. For the time delay, this is straightforward by just removing our normalization factor:

$$\Delta t = \Delta t_{\mathrm{norm}} \cdot 150\,\mathrm{days}, \tag{5.18}$$

but for σ_{Δ_t}, this is not as trivial, and we will explain our approach in Sect. 5.3. In the following, we will refer to our machine learning network composed of a Siamese LSTM network and a FCNN as LSTM-FCNN.

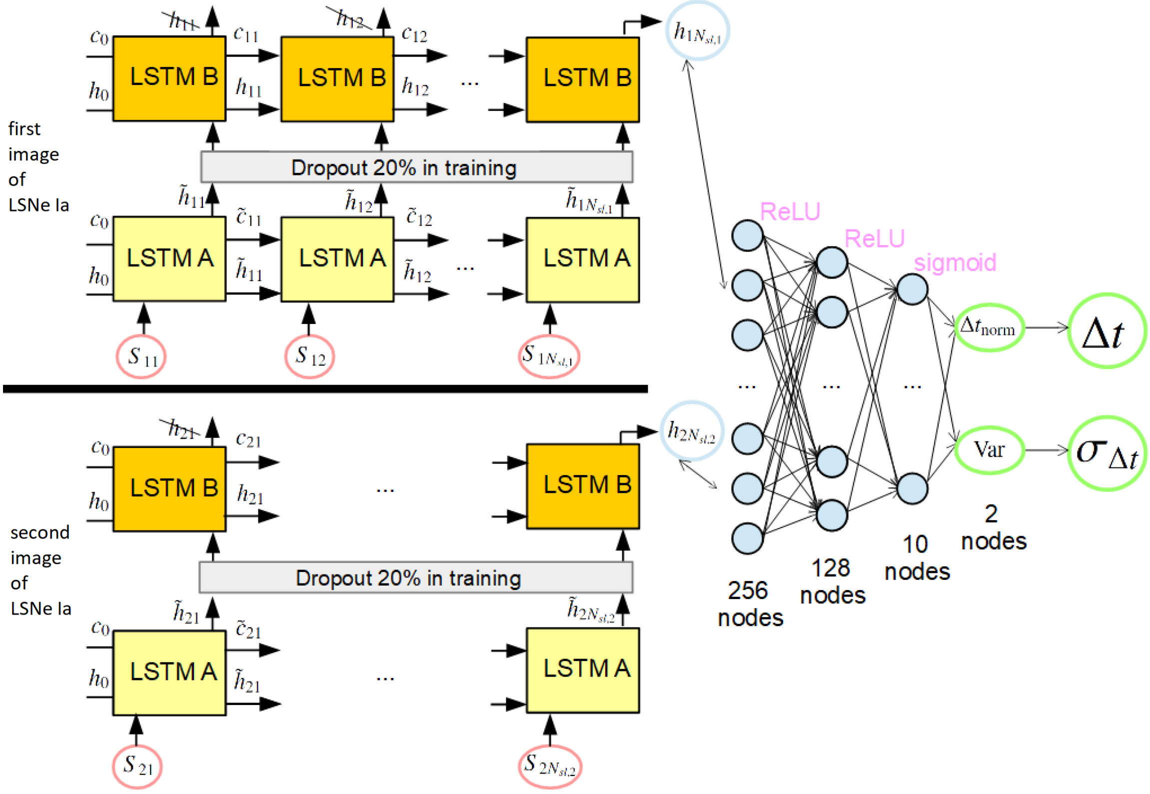


Fig. 5.7. Illustration of the machine learning network used in this chapter, composed of a Siamese LSTM network and a FCNN with two hidden layers. Details about the LSTM cells are shown in Fig. 5.6. Both sub-LSTM networks for image one (top, input denoted by S_{1k}) and image two (bottom, input denoted by S_{2k}) are the same LSTM network (i.e., with the same weights in the neural network) composed of 2 layers (LSTM A and LSTM B). During the training process, we use a dropout rate of 20%. From the second layer, only the last outputs $h_{1N_{\mathrm{sl},1}}$ and $h_{2N_{\mathrm{sl},2}}$ are further used, which are concatenated to be the input for the FCNN. The FCNN has two hidden layers using ReLU activation functions and a sigmoid activation for the final output. Given that the data is normalized, the output of the network t_{norm} and Var needs to be converted to get the time delay Δt [see Eq. (5.18)] and the uncertainty of the time delay $\sigma_{\Delta t}$ (see Sect. 5.3). The initial cell state c_0 and hidden state h_0 have the size N_{hidden} with all elements being equal to zero.

We use the machine learning library `PyTorch` (Paszke et al., 2019) to implement the LSTM-FCNN, where we train for a certain number of epochs until a generalization gap becomes visible (more details in Sect. 5.3). In each epoch, we subdivide our training set randomly into mini-batches with size $N_{\mathrm{batch}} = 128$. Given that image one and image two of the 128 samples (labeled as l) of a mini-batch have a different sequence length $N_{\mathrm{sl},1,l}$ and $N_{\mathrm{sl},2,l}$, we pad them with zeros at the end until all mini-batch samples of image one have a shape of $[\max(N_{\mathrm{sl},1,l}), 3]$ and all samples of image two have a shape of $[\max(N_{\mathrm{sl},2,l}), 3]$, where the 3 is coming from the three inputs from Eq. (5.1). The padding is required such that the `PyTorch` class `torch.nn.LSTM` can handle variable sequence lengths within a mini batch[37]. As input for the FCNN, we use for each sample l the last *unpadded* output at time step $N_{\mathrm{sl},1,l}$ for image one and $N_{\mathrm{sl},2,l}$ for image two, which are then combined to predict Δt and $\sigma_{\Delta t}$. For each sample of the mini-batch, we can then compute the performance of the LSTM-FCNN by using the Gaussian negative log-likelihood (Nix & Weigend, 1994) as the loss function:

$$\mathrm{loss} = 0.5\left(\log(\max(\mathrm{Var}, \epsilon)) + \frac{(\Delta t_{\mathrm{norm}} - \Delta t_{\mathrm{norm,true}})^2}{\max(\mathrm{Var}, \epsilon)}\right), \tag{5.19}$$

where $\epsilon = 10^{-6}$, is used for stability. The true normalized time delay of the training sample is $\Delta t_{\mathrm{norm,true}}$, and Δt_{norm} is the time delay predicted from the network. Furthermore, we use the second "raw" output of the LSTM-FCNN as the variance "Var" as in Equation (5.17). The advantage of the Gaussian negative log-likelihood loss compared to the mean squared error loss [see Eq. (2.109)] used in Chapter 4 is that we get an approximation of the uncertainty for each sample while the predicted time delay is optimized. The loss is then used for the backpropagation to update the weights and biases of the LSTM-FCNN. For this we use the Adaptive Moment Estimation (Adam) optimizer (Kingma & Ba, 2014) with a learning rate $\alpha = 0.00001$.

The network with our specific choices/values of N_{hidden}, N_{batch}, α, number of LSTM layers, dropout rate, FCNN with two hidden layers, and so on, is somewhat arbitrary. However, the setup of the FCNN was motivated by our investigations in Chapter 4. Additionally, we investigated for a smaller training sample other setups, e.g., $N_{\mathrm{hidden}} = 64, 256, 512$, $N_{\mathrm{batch}} = 64, 256, 512$, just one LSTM layer without dropout and so on. We found that the performance of these networks was typically very similar or ~10 percent worse than our chosen setup. Therefore, given that the training of a single setup of such an LSTM network takes several weeks on a GPU node and that we found only minor performance differences for a few tested configurations of the LSTM-FCNN, we skipped a huge grid search for all potential hyperparameters, which is also justified by the performance of our LSTM-FCNN already achieving our target precision and accuracy.

5.3 Results

To evaluate the performance of our LSTM-FCNN on a given data set (e.g., training set, validation set, test set), we calculate for each sample S_l as in Eq. (5.2), labeled with l, the time delay deviation via,

---

[37] To implement this, the `PyTorch` functions `torch.nn.utils.rnn.pad_sequence`, `torch.nn.utils.rnn.pack_padded_sequence`, and `torch.nn.utils.rnn.pad_packed_sequence` are used.

$$\tau_l = \Delta t_{\mathrm{true},l} - \Delta t_{\mathrm{pred},l}, \tag{5.20}$$

comparing the true time delay $\Delta t_{\mathrm{true},l}$ to the predicted one from the network $\Delta t_{\mathrm{pred},l}$. Given that our data is normalized in time for the input/output of our LSTM-FCNN, we need to undo the normalization as in Eq. (5.18) to get the time delays in days. We can then estimate the bias (also referred to as accuracy) by using the median $\tilde{\tau}$ and the precision by using the 84th percentile $\tau_{p,84}$ and the 16th percentile $\tau_{p,16}$ from the whole sample (200000 for the training data and 20000 for the other data sets). The results are then summarized in the following form:

$$\tilde{\tau}^{\tau_{p,84}-\tilde{\tau}}_{\tau_{p,16}-\tilde{\tau}}. \tag{5.21}$$

We train our LSTM-FCNN as described in Sect. 5.2.2 with the data set as described in Sect. 5.1 using light curves from theoretical and the empirical `SNEMO15` models as listed in Table 5.2. An approach where we use only theoretical SN Ia models in the training process is shown in Appendix 7.1.

Our training process lasts for 10400 epochs, and we use the LSTM-FCNN at the epoch where the validation set has the minimum loss value calculated from Eq. (5.19). In addition, we require that the bias as defined in Eq. (5.21) is lower than 0.05-day to ensure that at least for the validation set, a time delay longer than 5 days would be enough to achieve a bias lower than 1%. The loss curve for the training and validation set as a function of the training epoch is shown in Fig. 5.8, where we see the generalization gap beyond epoch ~5000, and we reach the lowest validation loss with low bias at epoch 7313, which is the status of the LSTM-FCNN as used further in this work.

For our LSTM-FCNN we calculate then the time-delay deviation for the training set, validation set, test set, and `SNEMO15`-only test set, which is shown in Fig. 5.9. Our network performs very well with almost zero bias and precision on the order of 0.7-day, which can also be achieved on light curve shapes never used in the training process, as represented by the `SNEMO15`-only test set. In comparison to the results from Appendix 7.1, where we only train on theoretical SN Ia models, we see that using at least parts of the `SNEMO15` light curves in the training process helps to reduce the bias on unseen light curves in the training process, although biases for a network just trained on theoretical models are still on a negligible level if one restricts its application to LSN Ia systems with time delays longer than 20 days.

In order to relate the raw Var output from our LSTM-FCNN also listed in Eq. (5.19) to a meaningful $\sigma_{\Delta t}$, we use the validation set as shown in Fig. 5.9. To convert the Var output, we first calculate the rescaled uncertainty prediction from the network

$$\sigma_{\mathrm{res}} = \sqrt{\mathrm{Var}} \cdot 150\,\mathrm{day}, \tag{5.22}$$

where the factor 150 comes from the normalization of our time scale as described in Sect. 5.1. We then compare this to the 1σ width of the time-delay deviation for the validation data set, which we approximate to be Gaussian by assuming as standard deviation $(\tau_{p,84} - \tau_{p,16})/2$. Then, to find a relation, we fit a second-order polynomial as shown in Fig. 5.10,

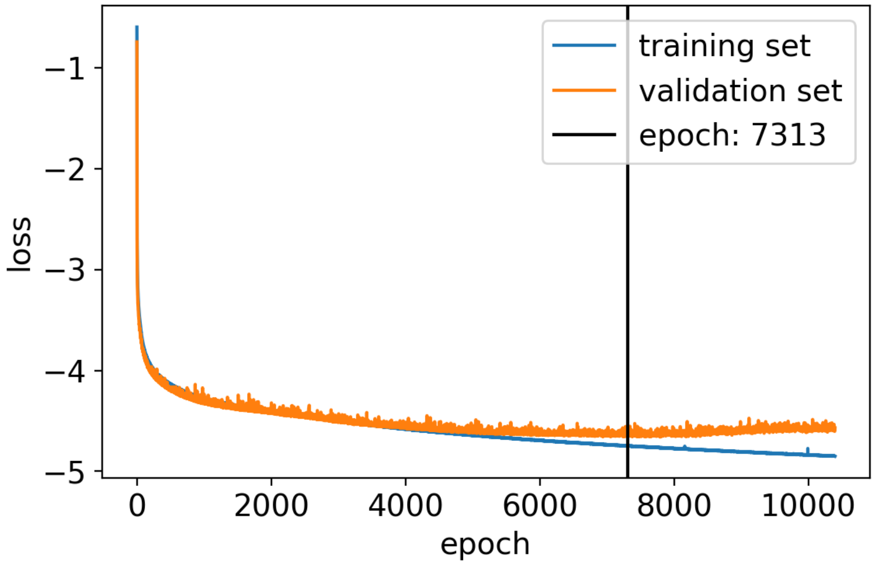


Fig. 5.8. Training and validation loss of the LSTM-FCNN as a function of the training epoch. The vertical line marks the state at which we use the LSTM-FCNN where the loss is minimal with the additional requirement that the bias is lower than 0.05-day. The plot shows improving performance with more training epochs. Around epoch 7300, the validation loss increases again in contrast to the training loss because the LSTM-FCNN starts to overfit, meaning that, e.g., it fits specific noise patterns just present in the training set.

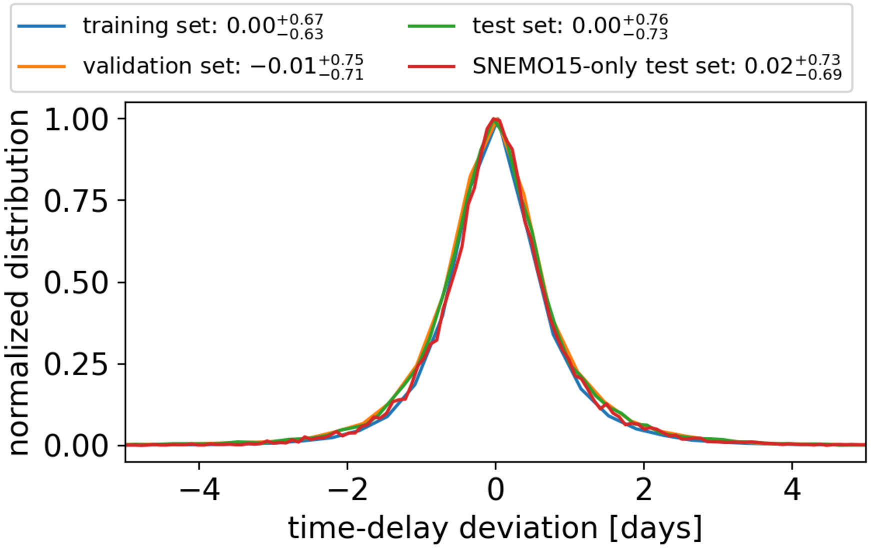


Fig. 5.9. Evaluation of the LSTM-FCNN on the training set, validation set, test set, and `SNEMO15`-only test set. With a 1σ spread of ~0.7 and almost no bias even on the two test sets, the LSTM-FCNN seems to be very promising for future measurements of time delays from real LSNe Ia.

which yields

$$\sigma_{\Delta t} = 0.14\sigma_{\mathrm{res}}^2 + 0.75\sigma_{\mathrm{res}} + 0.03. \tag{5.23}$$

In Fig. 5.11, we show the predicted uncertainties of the LSTM-FCNN following Eq. (5.23) for the test set and `SNEMO15`-only test set. We find that the median values are within $\lesssim$0.05 day of the corresponding precision from Fig. 5.9; therefore, our uncertainty output gives a reasonable estimate of the actual underlying uncertainty.

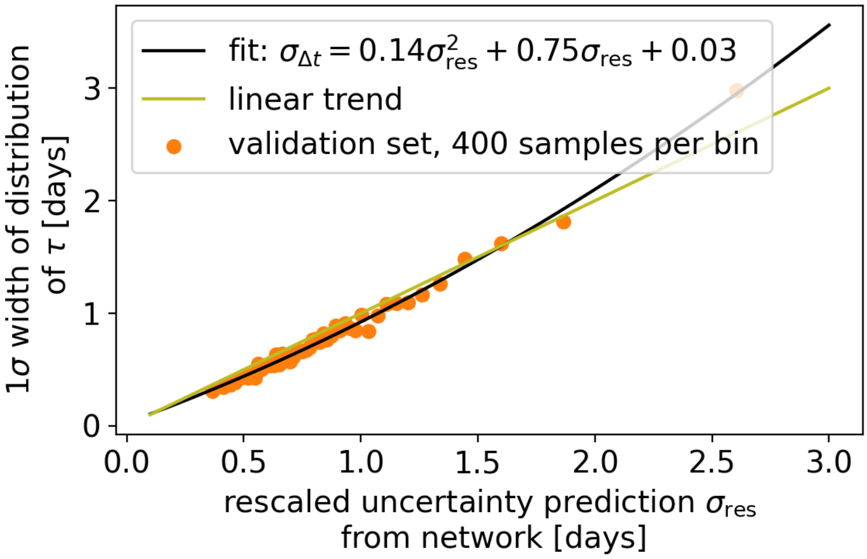


Fig. 5.10. 1σ width of the validation set, approximated as $(\tau_{p,84} - \tau_{p,16})/2$, in comparison to the rescaled uncertainty prediction from the network σ_{res} for the same data set. The fit is used to translate, via Eq. (5.22), the "Var" output of the LSTM-FCNN to a meaningful 1σ uncertainty $\sigma_{\Delta t}$ of the time-delay measurement Δt.

5.4 Specific applications of the LSTM-FCNN and comparison to other methods

In this Section, we apply the LSTM-FCNN to more specific examples to compare the performance to the results from Chapter 4. In particular, we create a data set as described in Sect. 5.1, but we fix $z_{\mathrm{s}}, z_{\mathrm{d}}, \kappa, \gamma$, and s to very specific values as used in Chapter 4, where a system was picked from the OM10 catalog with a source redshift $z_{\mathrm{s}} = 0.76$ that is closest to the median source redshift $z_{\mathrm{s}} = 0.77$ of the OM10 catalog. The specific system is a double LSNe Ia with $(\kappa_1, \gamma_1, s_1) = (0.25, 0.27, 0.6)$ for the first SN image and $(\kappa_2, \gamma_2, s_2) = (0.83, 0.81, 0.6)$ for the second SN image. The lens is located at $z_{\mathrm{d}} = 0.252$. For this system, we investigated in Chapter 4 different scenarios at peak and before peak detection and also varied the source redshifts.

Fig. 5.12 shows our LSTM-FCNN on a `SNEMO15`-only test set, based on $(\kappa_1, \gamma_1, s_1) = (0.25, 0.27, 0.6)$ and $(\kappa_2, \gamma_2, s_2) = (0.83, 0.81, 0.6)$ for three different z_{s} (~median and 16th and 84th percentile of OM10). We first see that even though the specific κ and γ values are

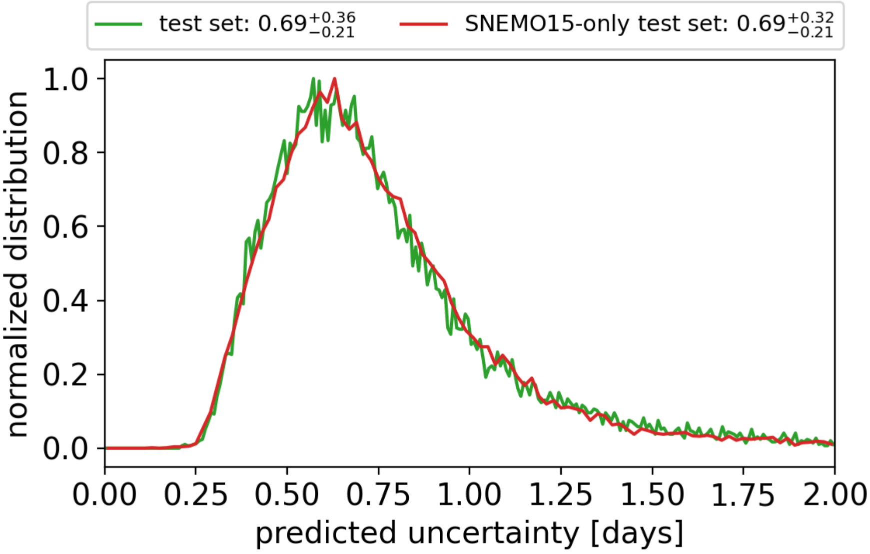


Fig. 5.11. Predicted uncertainty of the LSTM-FCNN on the test set and `SNEMO15`-only test set. The median values of both test sets match well, within $\lesssim$0.05 day, the spread shown in Fig. 5.9.

not part of our training data (Table 5.1), the network predicts very accurate and precise time delays. Further, for source redshifts lower than the median of the OM10 catalog ($z_s = 0.77$), we find that the performance of our LSTM-FCNN is similarly leading to very precise time-delay measurements of ~0.55 day. If we go to redshifts of ~1.0, the precision drops to ~0.75 day.

In Table 5.3, we compare the results from the LSTM-FCNN to the results from the RF taken from Table 4.5, where we see that the predictions from the LSTM-FCNN are far more precise (by more than a factor of 2) than the ones from the RF. One reason for the worse performance of the RF is that the RF was only trained on theoretical models, whereas the evaluation was done on the empirical `SNEMO15` model. Although the RF did generalize well to the light curve shapes not used in the training process in terms of bias, the precision dropped by ~0.5 days in comparison to a test set based on theoretical light curves as used in the training process (see Fig. 4.8). However, even if we would improve the precision by ~0.5 days by including `SNEMO15` light curves in the training process of the RF, the LSTM-FCNN would still outperform the RF by almost a factor of two in terms of precision. The second reason for the much better performance seems to be that the LSTM structure is better suited to the applied problem, given that it was built to solve time-dependent problems.

In a second specific example, we assume different times of detection in addition to the aforementioned $(\kappa_1, \gamma_1, s_1, \kappa_2, \gamma_2, s_2, z_d)$ values for $z_s = 0.76$. Therefore, we fix in Table 5.1 in addition to $(\kappa_1, \gamma_1, s_1, \kappa_2, \gamma_2, s_2, z_s, z_d)$, the detection of the first image to very specific values. From Fig. 5.13, we see that the precision drops with fewer data points before the peak, although at peak detection still yields an uncertainty of ~1.0 day, which is still a precise time-delay measurement. Furthermore, we find that the accuracy deviates from zero, showing that the spread detected in Fig. 5.9 not only comes from random noise and microlensing positions

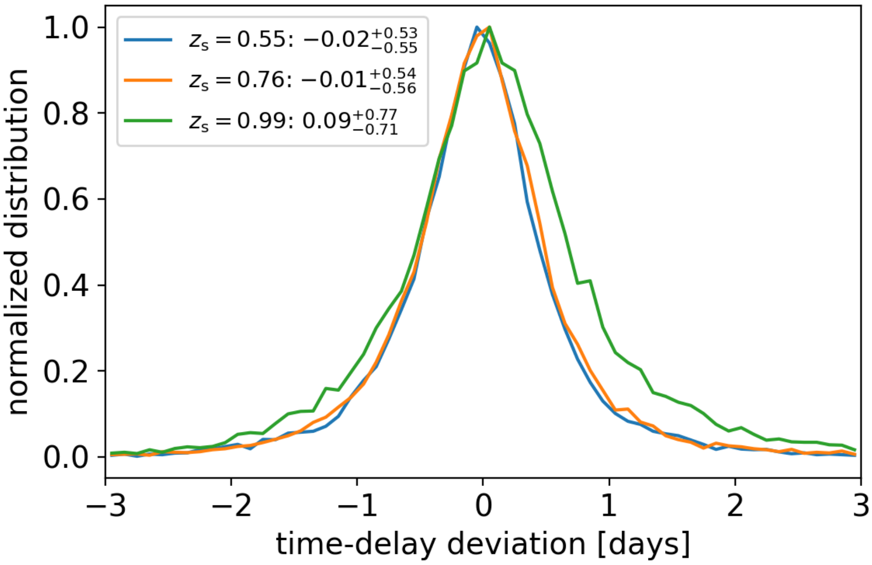


Fig. 5.12. Evaluation of the LSTM-FCNN on `SNEMO15`-only test sets assuming a specific LSN Ia system with $(\kappa_1, \gamma_1, s_1) = (0.25, 0.27, 0.6)$ for image one and $(\kappa_2, \gamma_2, s_2) = (0.83, 0.81, 0.6)$ for image two as in Chapter 4 and three different z_s. The precision for a large LSNe Ia sample for source redshifts lower than the median source redshift of OM10 ($z_s = 0.77$) is around ~0.55 day and drops to ~0.75 day, which is still very precise, for redshifts of one.

Table 5.3. Time-delay deviations of the LSTM-FCNN on `SNEMO15`-only test sets in comparison to the RF (results are taken from Table 4.5 for `SNEMO15` data sets) for a LSN Ia with $(\kappa_1, \gamma_1, s_1) = (0.25, 0.27, 0.6)$ for image one and $(\kappa_2, \gamma_2, s_2) = (0.83, 0.81, 0.6)$ for different assumptions on the source redshift z_s.

	LSTM-FCNN	RF
$z_s = 0.55$	$-0.02^{+0.53}_{-0.55}$	$0.01^{+1.29}_{-1.25}$
$z_s = 0.76$	$-0.01^{+0.54}_{-0.56}$	$0.02^{+1.38}_{-1.42}$
$z_s = 0.99$	$0.09^{+0.77}_{-0.71}$	$0.02^{+2.1}_{-2.16}$

– for very specific systems, the LSTM-FCNN slightly over or underestimates the predicted time delay, which averages out if large samples of LSNe Ia are investigated.

In Table 5.4, we compare the results from Fig. 5.13 to similar runs for the RF and PyCS (see Figs. 4.17 and 4.19) and we find that the accuracy of the LSTM-FCNN is comparable to PyCS and overall better than the RF, especially if fewer data points are available before peak. However, if we look at the precision, we see that the LSTM-FCNN outperforms the RF and PyCS by far. PyCS provides by far the worst precision, which is not surprising, given that splines are used, and therefore any knowledge about SN Ia light curves is completely ignored. The performance differences between the LSTM-FCNN and the RF are comparable to the ones presented and discussed in Table 5.3. Because of the much better performance of the LSTM-FCNN in comparison to PyCS, we can expect more LSNe Ia with well-measured time delay (accuracy better than 1% and precision better than 5%) than that predicted by Huber et al. (2019).

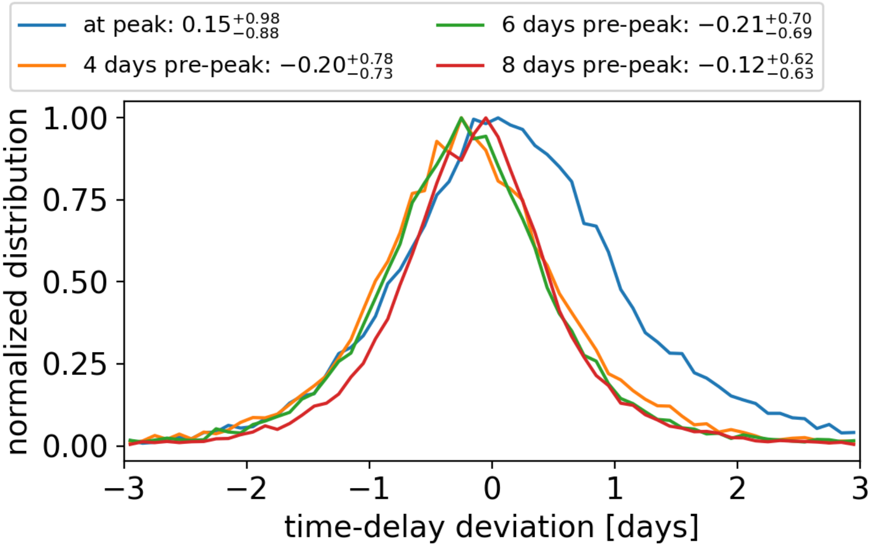


Fig. 5.13. Evaluation of the LSTM-FCNN on `SNEMO15`-only test sets assuming a specific LSN Ia system at $z_s = 0.76$ with $(\kappa_1, \gamma_1, s_1) = (0.25, 0.27, 0.6)$ for image one and $(\kappa_2, \gamma_2, s_2) = (0.83, 0.81, 0.6)$ for image two as in Chapter 4 for at peak detection and three cases where the first SN image is detected before peak (in observer-frame days). As expected, the precision drops with fewer available data points pre-peak but is still very good with ~1.0 day for an at-peak detection. The accuracy is worse than in Fig. 5.9, which shows that the spread presented there is not only due to various noise and microlensing configurations but also comes from different specific LSNe Ia, although the bias cancels out if a large variety of systems is investigated.

Table 5.4. Time-delay deviations of the LSTM-FCNN on `SNEMO15`-only test sets in comparison to the RF and `PyCS` (results for `SNEMO15` data sets are taken from Figs. 4.17 and 4.19) for a LSN Ia with $(\kappa_1, \gamma_1, s_1) = (0.25, 0.27, 0.6)$ for image one and $(\kappa_2, \gamma_2, s_2) = (0.83, 0.81, 0.6)$ for image two for $z_s = 0.76$ and various detection times (in observer days) before the first SN image peaks in brightness.

	LSTM-FCNN	RF	`PyCS`
at peak detection	$0.15^{+0.98}_{-0.88}$	$0.83^{+3.33}_{-2.05}$	$-0.16^{+3.99}_{-3.52}$
4 days pre-peak	$-0.20^{+0.78}_{-0.73}$	$0.22^{+2.07}_{-1.62}$	$0.12^{+2.89}_{-2.22}$
6 days pre-peak	$-0.21^{+0.70}_{-0.69}$	$0.17^{+1.61}_{-1.48}$	$0.15^{+2.25}_{-1.57}$
8 days pre-peak	$-0.12^{+0.62}_{-0.63}$	$0.09^{+1.46}_{-1.42}$	$0.09^{+1.87}_{-1.39}$

6 Summary and outlook

Gravitational lensing arises as a natural consequence of spacetime curvature from general relativity. This theory of gravity has its origin in inconsistencies between Newtonian mechanics and the unknown frame of reference for the laws of electrodynamics. This changed with Einstein's groundbreaking publication "Zur Elektrodynamik bewegter Körper" (Einstein, 1905). Maybe with the current tension in H_0 we are standing again right in front of groundbreaking new physics.

However, the tension first needs to be confirmed by independent measurements, and time-delay cosmography with strongly lensed SNe Ia can be one of them. This method is based on three pillars. A time-delay measurement in combination with a lens mass model and a reconstruction of mass perturbations along the line of sight gives a direct measurement of H_0. Time-delay cosmography has been applied already successfully to strongly lensed quasars (e.g., Wong et al., 2020; Birrer et al., 2020) by the TDCOSMO organization (Millon et al., 2020), consisting of H0LiCOW (Suyu et al., 2017), COSMOGRAIL (Courbin et al., 2018), SHARP (Chen et al., 2019), and STRIDES (Shajib et al., 2020). However, LSNe have several advantages over lensed quasars. The characteristic light curve shapes typically allow time-delay measurements on shorter time scales. Further, SNe fade with time, enabling follow-up observations of the lensing galaxy with multiple images of the SN host galaxy. This can be used to derive a more precise lens model, free of contamination from bright point sources (Ding et al., 2021). Further, the follow-up observations allow measuring stellar kinematics (Barnabè et al., 2011a; Yıldırım et al., 2017; Shajib et al., 2018; Yıldırım et al., 2020) for breaking model degeneracies, such as the mass-sheet transformation. LSNe Ia are even more promising than other lensed SNe, given that SNe Ia are standardizable candles and, therefore, provided that microlensing effects can be mitigated (Oguri & Kawano, 2003; Yahalomi et al., 2017; Foxley-Marrable et al., 2018; Huber et al., 2019; Pierel & Rodney, 2019; Huber et al., 2021; Weisenbach et al., 2021), allow the determination of the lensing magnifications and therefore yield an additional way to break degeneracies in the lens mass model.

The main scope of this thesis was to develop tools to measure time delays in LSNe Ia and investigate main uncertainties. First, we investigated the color curves of LSNe Ia qualitatively to mitigate the uncertainties of microlensing. Goldstein et al. (2018) showed that for the W7 model (Nomoto et al., 1984), where synthetic observables have been calculated with the radiative transfer code `SEDONA` (Kasen et al., 2006), the effect of microlensing in the color curves up to three rest-frame weeks after the explosion is almost negligible. This is referred to as the achromatic phase, where specific intensity profiles for different bands are very similar, and therefore, the microlensing cancels out in the color curves. To investigate this further we used four theoretical models, namely, W7, N100 (Seitenzahl et al., 2013), a sub-Ch (Sim et al., 2010) model, and a merger model (Pakmor et al., 2012). To calculate synthetic observables for these models, we used the radiative transfer code `ARTIS` (Kromer & Sim, 2009). From our investigation, we found as well an achromatic phase of around three rest-frame weeks after the explosion in good agreement with Goldstein et al. (2018).

However, our findings show that for typical source redshifts, color curves with robust features for delay measurement, require broadband follow-up in r, i, z, y, J, and H, which will be challenging in practice. Therefore we developed our methods for measuring time delays based on light curves, although microlensing uncertainties will be higher.

In the first approach, we used two simple machine learning techniques, a FCNN and RF, to measure time delays in LSNe Ia. These machine learning models always require the same amount of input data points, and therefore, a model must be trained for each observed LSNe Ia separately. However, the advantage is that we train our systems very specifically for a certain observation, and thus, accurate and precise results are expected. To train the FCNN and RF for a mock LSNe Ia, we simulated realistic light curves with observational noise based on the four theoretical models to include the effect of microlensing. Although both networks performed well on a test set based on four theoretical models, only the RF provided bias-free results on a data set based on empirical `SNEMO15` (Saunders et al., 2018) light curves. Therefore the RF should be used in practice. For our mock LSNe Ia at a redshift of 0.76, with a follow-up cadence of two days to a 5σ point-source depth of 25.7, 25.3, 24.7, and 23.8 for g, r, i, and z, we achieved an ~1.5 day uncertainty in the i band. In order to combine multiple filters, we find that it is best to train a RF for each filter separately. From three filters we expect a ~1.0 day uncertainty, where gri bands are preferred for $z_{\mathrm{s}} \lesssim 0.6$ and riz bands for higher z_{s}. Further, our results show that only a detection before the peak assures low biases. In terms of uncertainties, microlensing is not negligible (factor of ~2 compared to no microlensing), but the dominant source of uncertainty is the observational noise, which makes the follow-up depth a crucial parameter.

Our second approach was to develop a more general method to measure time delays from any LSNe Ia detected by LSST, for which we plan to trigger follow-up observations with a 2-day cadence, one magnitude deeper than the LSST-like 5σ depth. The main goal is to get a flexible machine-learning model working with a variable number of input data points. Therefore, we use a Siamese long short-term memory network combined with a FCNN (LSTM-FCNN). For this model, we simulated a data set of 200000 pairs of LSNe Ia images, where we use only i band as the most promising band and further assume a before peak detection. Despite that, we cover a broad range of potential systems covering source redshifts from 0.17 to 1.34 and time delays from 0.1 to 150 days (Oguri & Marshall, 2010). For the network training, we use light curves from four theoretical models and the empirical `SNEMO15` model, although one-half of the empirical light curves are used only in the validation and testing process. To include microlensing for the empirical SN Ia light curves, we add the microlensing-only contribution from the theoretical models. After successful training of the LSTM-FCNN, we apply the network to a test set of 20000 LSNe Ia composed of empirical `SNEMO15` light curves not used in the training process. In this test of generalizability, we find excellent results with an uncertainty of ~0.7 days and almost zero bias; therefore, the LSTM-FCNN outperforms the RF. Another advantage is that the LSTM-FCNN outputs a time delay and an estimate of the uncertainty. However, for an observed LSNe Ia it might still be beneficial to train a RF in addition to the LSTM-FCNN because the RF is trained explicitly for a specific observation pattern, microlensing, and noise configuration, which will help to identify potential biases.

Although the LSTM-FCNN has a far broader application than the RF, the LSTM-FCNN cannot be blindly applied to any LSNe Ia. If follow-up strategies deviate significantly from our assumptions (e.g., detection, filter, cadence), a new LSTM-FCNN needs to be trained. Furthermore, we assumed only positive time delays, meaning that image one should be correctly identified as the first appearing image and image two as the second. This is not always possible for very short time delays of a couple of days, and therefore, our LSTM-FCNN cannot be applied to these systems. In practice, this is not a problem as typically only the image pairs of a LSNe Ia with longer time delays are of interest for cosmography. To avoid biases Huber et al. (2019) recommends focusing on LSNe Ia with time delays longer than 20 days, which agrees well with our results from the RF or LSTM-FCNN.

Although the precision in the i band is already quite good, additional bands, as we saw for the RF should further improve the precision. Here we saw that especially the r band is very promising. In the future, it might also be interesting to train a LSTM-FCNN for additional bands, especially r. However, using at least some of the light curves from empirical models in the training process of the LSTM-FCNN is helpful to reduce the overall bias on a very large sample of LSNe Ia from 0.2-day to 0.02-day. The `SNEMO15` model covers only a wavelength range from 3305 Å to 8586 Å and therefore r band light curves are only possible up to a redshift of 0.62. The empirical `SALT2` model (Guy et al., 2007) enables with a range of 2500 Å to 8000 Å, r band light curves up to a redshift of 1.15 and is a potential source for a future LSTM-FCNN on r band.

However, the LSTM-FCNN on a single band outperformed the RF and `PyCS`, as used in Huber et al. (2019), by far. For LSNe Ia at $z_{\rm s} = 0.76$, detected four to eight days before peak, `PyCS` provides a time-delay precision of $\sim$2.0 days, the RF of $\sim$1.6 days and the LSTM-FCNN of $\sim$0.7 days. Therefore, we can expect more LSNe Ia with accurate and precise time-delay measurements compared to our predictions in Huber et al. (2019). Based on results from Huber et al. (2019), Suyu et al. (2020) forecasted cosmological constraints from strongly lensed SNe. For the baseline observing strategy with follow-up observations, one magnitude deeper than LSST, one expects 30 LSNe Ia with time delays longer than 20 days (criterion to avoid biases). From these 30 LSNe Ia, roughly 20 will have a time-delay precision better than five percent based on results from `PyCS` in Huber et al. (2019). In a conservative estimate, Suyu et al. (2020) assumed that all 20 LSNe Ia have a time-delay uncertainty of 5%, a lens mass model uncertainty of 3%, and a lens environment uncertainty of 3%. Adding these values in quadrature leads to a 6.56% uncertainty on $D_{\Delta t}$ for each LSNe Ia. From this Suyu et al. (2020) predicted for the 20 LSNe Ia in a flat ΛCDM a 1.52 % uncertainty on H_0 with only weak constraints on $\Omega_{\rm m}$ ($\sim$50% uncertainty). Given the weak constraints on $\Omega_{\rm m}$ the uncertainty of $D_{\Delta t}$ is almost exclusively represented in H_0: $\delta H_0/H_0 \approx 6.56\%/\sqrt{20} = 1.47\%$. Now with an improved time-delay precision by a factor of two, which is achievable with the LSTM-FCNN, we get 4.92% uncertainty on $D_{\Delta t}$ for each LSNe Ia and therefore $\delta H_0/H_0 \approx 4.92\%/\sqrt{20} = 1.10\%$ in flat ΛCDM. Even though we assumed that $D_{\Delta t}$ constrains only H_0, this is still a rather conservative estimate, given that all 30 LSNe Ia (although with worse precision) will contribute to the H_0 measurement and that the improvement of the LSTM-FCNN in comparison to `PyCS` is larger than two. Therefore, it seems possible to push the uncertainty on H_0 below 1%, making time-delay cosmography with LSNe Ia a highly competitive H_0 probe. Given the promising results presented in this thesis, it will be interesting in the future to train a LSTM-FCNN to measure time delays in lensed core-collapse SN.

Personal note. If independent measurements confirm the H_0 tension, theoretical work will be demanded to modify our cosmological models or even our fundamental laws of gravity. This might arise in new exciting physics like the theories of relativity in the early 20th century, which always fascinated me and were the reason why I started to study physics. However, such a development could take a very long time and might by far exceed my lifetime. Therefore I would be delighted to take a spaceship close to the end of my life, speed it up to almost the speed of light, and come back a few moments later when thousands of years have passed on earth. Hopefully, I will find a better world than we live in today and one of my great × 100-grandchildren, who will explain to me all the exciting physical developments I have missed.

Acknowledgments

First and foremost, many thanks go to my supervisor Sherry Suyu for the opportunity to work on such an exciting research project. Sherry always provided great support while creating an environment where I could pursue my ideas. I also thank her for our regular meetings, where we discussed my project, such that I was able to learn a lot from her. I am exceedingly grateful for the last five years, including my master's studies! Further, I would like to thank all our current and former group members for a great time at the institute and beyond. I especially enjoyed our group activities like laser tag, exit rooms, kayaking, and the archery event. Many thanks go to Stefan Schuldt for always giving good advice, which helped me a lot to dig through all the bureaucratic work related to a PhD. I also thank him for all our pleasant conversations while enjoying free pizza and chocolate after the lunch talks at ESO. Further, I am grateful to Stefan Taubenberger for always providing help and sharing his knowledge about supernovae. I also thank him for trying to organize several observing trips. Unfortunately, the coronavirus had other plans. The one time we were finally able to go observing, we went to Asiago, but unfortunately, we had to leave after one night, again, because of the virus. Still, I am thankful we could at least observe for one night. Further, I am grateful to Debarghya Ghoshdastidar, Tim Meinhardt, and Laura Leal-Taixé for providing help with machine-learning-related problems. I thank James Chan for teaching me how to calculate microlensing-magnification maps and Markus Kromer and Stuart Sim for providing ARTIS simulations. I also like to thank Ulrich Nöbauer for his help during my master's studies, from which I still profit today. Furthermore, I thank Andreas Weiss, Andreas Breitfeld, and Goran Toth for their computational support at MPA. For proofreading this manuscript and providing valuable comments, I like to thank Sherry and Stefan T. again and also Raoul Cañameras. Many thanks go as well to the Technische Universität München and the Max-Planck-Institut für Astrophysik for enabling this PhD project! I also like to thank Stephanie Rausch and my parents, Ingrid and Franz Huber, for always supporting me!

References

Abbott, B. P. et al. 2017, Nature, 551, 85, [arXiv:1710.05835]

Allen, S. W., Schmidt, R. W., & Fabian, A. C. 2002, MNRAS, 334, L11, [arXiv:astro-ph/0205007]

Alzubaidi, L., Zhang, J., Humaidi, A. J., et al. 2021, J. Big Data, 8, 53

Anand, G. S., Tully, R. B., Rizzi, L., Riess, A. G., & Yuan, W. 2022, ApJ, 932, 15, [arXiv:2108.00007]

Ashman, K. M. 1992, PASP, 104, 1109

Aylor, K., Joy, M., Knox, L., et al. 2019, ApJ, 874, 4, [arXiv:1811.00537]

Bacon, D. J., Refregier, A. R., & Ellis, R. S. 2000, MNRAS, 318, 625, [arXiv:astro-ph/0003008]

Barnabè, M., Czoske, O., Koopmans, L. V. E., Treu, T., & Bolton, A. S. 2011a, MNRAS, 415, 2215, [arXiv:1102.2261]

Barnabè, M., Czoske, O., Koopmans, L. V. E., Treu, T., & Bolton, A. S. 2011b, MNRAS, 415, 2215, [arXiv:1102.2261]

Baron, E., Bongard, S., Branch, D., & Hauschildt, P. H. 2006, ApJ, 645, 480, [arXiv:astro-ph/0603101]

Bartelmann, M. & Schneider, P. 2001, Phys. Rept., 340, 291, [arXiv:astro-ph/9912508]

Baxter, E. J. & Sherwin, B. D. 2021, MNRAS, 501, 1823, [arXiv:2007.04007]

Bayer, J., Huber, S., Vogl, C., et al. 2021, A&A, 653, A29, [arXiv:2101.06229]

Benhammou, Y., Achchab, B., Herrera, F., & Tabik, S. 2020, Neurocomput., 375, 9–24

Bernal, J. L., Verde, L., & Riess, A. G. 2016, JCAP, 10, 019, [arXiv:1607.05617]

Bessell, M. & Murphy, S. 2012, PASP, 124, 140, [arXiv:1112.2698]

Birrer, S., Welschen, C., Amara, A., & Refregier, A. 2017, JCAP, 04, 049, [arXiv:1610.01599]

Birrer, S. et al. 2019, MNRAS, 484, 4726, [arXiv:1809.01274]

Birrer, S. et al. 2020, A&A, 643, A165, [arXiv:2007.02941]

Blakeslee, J. P., Jensen, J. B., Ma, C.-P., Milne, P. A., & Greene, J. E. 2021, ApJ, 911, 65, [arXiv:2101.02221]

Bolton, A. S., Burles, S., Schlegel, D. J., Eisenstein, D. J., & Brinkmann, J. 2004, Astron. J., 127, 1860, [arXiv:astro-ph/0311055]

Bond, I. A., Udalski, A., Jaroszyński, M., et al. 2004, ApJL, 606, L155, [arXiv:astro-ph/0404309]

Bonvin, V. & Millon, M. 2020, H0LiCOW H0 tension plotting notebook

Bonvin, V., Tewes, M., Courbin, F., et al. 2016, A&A, 585, A88, [arXiv:1506.07524]

Bonvin, V. et al. 2018, A&A, 616, A183, [arXiv:1804.09183]

Bonvin, V. et al. 2019, A&A., 629, A97, [arXiv:1905.08260]

Breiman, L. 2001, Mach. Learn., 45, 5

Bromley, J., Bentz, J. W., Bottou, L., et al. 1993, Int. J. Pattern Recognit. Artif. Intell., 07, 669, [https://doi.org/10.1142/S0218001493000339]

Brout, D. et al. 2021, [arXiv:2112.03864]

Browne, I. W. A., Wilkinson, P. N., Jackson, N. J. F., et al. 2003, MNRAS, 341, 13, [arXiv:astro-ph/0211069]

Buitinck, L., Louppe, G., Blondel, M., et al. 2013, in ECML PKDD Workshop: Languages for Data Mining and Machine Learning, 108–122

Burke, W. L. 1981, ApJL, 244, L1

Cañameras, R., Schuldt, S., Suyu, S. H., et al. 2020, A&A, 644, A163, [arXiv:2004.13048]

Canameras, R. et al. 2021, A&A, 653, L6, [arXiv:2107.07829]

Carlesi, E., Hoffman, Y., & Libeskind, N. I. 2022, MNRAS, 513, 2385, [arXiv:2204.03334]

Carroll, S. 2013, Spacetime and Geometry: An Introduction to General Relativity, Always learning (Pearson Education, Limited)

Chan, J. H. H., Rojas, K., Millon, M., et al. 2021, A&A, 647, A115, [arXiv:2007.14416]

Chang, K. & Refsdal, S. 1979, Nature, 282, 561

Chapelle, O., Schölkopf, B., & Zien, A., eds. 2006, Semi-Supervised Learning (The MIT Press)

Chen, G. C.-F., Chan, J. H. H., Bonvin, V., et al. 2018, MNRAS, 481, 1115, [arXiv:1804.09390]

Chen, G. C.-F. et al. 2019, MNRAS, 490, 1743, [arXiv:1907.02533]

Chornock, R. et al. 2013, ApJ, 767, 162, [arXiv:1302.0009]

Clowe, D., Bradac, M., Gonzalez, A. H., et al. 2006, ApJL, 648, L109, [arXiv:astro-ph/0608407]

Colgate, S. A. & McKee, C. 1969a, ApJ, 157, 623

Colgate, S. A. & McKee, C. 1969b, ApJ, 157, 623

Cooke, J. H. & Kantowski, R. 1975, ApJL, 195, L11

Courbin, F., Bonvin, V., Buckley-Geer, E., et al. 2018, A&A, 609, A71

Courbin, F., Saha, P., & Schechter, P. L. 2002, Lect. Notes Phys., 608, 1, [arXiv:astro-ph/0208043]

Crawford, M., Khoshgoftaar, T. M., Prusa, J. D., Richter, A. N., & Al Najada, H. 2015, J. Big Data, 2, 23

Cybenko, G. 1989, Mathematics of Control, Signals and Systems, 2, 303

Cypriano, E. S., Laerte Sodré, J., Kneib, J.-P., & Campusano, L. E. 2004, ApJ, 613, 95

Davis, T. M. & Scrimgeour, M. I. 2014, MNRAS, 442, 1117, [arXiv:1405.0105]

Deldjoo, Y., Elahi, M., Cremonesi, P., et al. 2016, JoDS, 5, 99

Denzel, P., Coles, J. P., Saha, P., & Williams, L. L. R. 2021, MNRAS, 501, 784, [arXiv:2007.14398]

Dessart, L. & Hillier, D. J. 2006, A&A, 447, 691, [arXiv:astro-ph/0510526]

Dessart, L., Hillier, D. J., Blondin, S., & Khokhlov, A. 2014, MNRAS, 441, 3249, [arXiv:1308.6352]

Dessart, L. et al. 2008, ApJ, 675, 644, [arXiv:0711.1815]

Dhawan, S. et al. 2020, MNRAS, 491, 2639, [arXiv:1907.06756]

Ding, X., Liao, K., Birrer, S., et al. 2021, MNRAS, 504, 5621, [arXiv:2103.08609]

Eddington, A. S. 1919, The Observatory, 42, 119

Einstein, A. 1905, Annalen der physik, 4

Einstein, A. 1915, Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften (Berlin), 844

Einstein, A. 1916, Annalen der Physik, 354, 769, [https://onlinelibrary.wiley.com/doi/pdf/10.1002/andp.19163540702]

Einstein, A. 1917, Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften (Berlin), 142

Einstein, A. 1936, Science, 84, 506

Eliasdottir, A., Hjorth, J., Toft, S., Burud, I., & Paraficz, D. 2006, ApJS, 166, 443, [arXiv:astro-ph/0606649]

Faber, S. M. & Gallagher, J. S. 1979, Annu. Rev. Astron. Astrophys., 17, 135

Falco, E. E., Gorenstein, M. V., & Shapiro, I. I. 1985, ApJL, 289, L1

Filippenko, A. V. 1997, Annu. Rev. Astron. Astrophys., 35, 309

Fink, M., Hillebrandt, W., & Roepke, F. K. 2007, A&A, 476, 1133, [arXiv:0710.5486]

Fink, M., Röpke, F. K., Hillebrandt, W., et al. 2010, A&A, 514, A53, [arXiv:1002.2173]

Foreman-Mackey, D. 2016, JOSS, 1, 24

Foxley-Marrable, M., Collett, T. E., Vernardos, G., Goldstein, D. A., & Bacon, D. 2018, MNRAS, 478, 5081, [arXiv:1802.07738]

Freedman, W. L. 2021, ApJ, 919, 16, [arXiv:2106.15656]

Freedman, W. L., Madore, B. F., Hoyt, T., et al. 2020, ApJ, 891, 57, [arXiv:2002.01550]

Freedman, W. L. et al. 2019, arXiv e-prints (arXiv:1907.05922), [arXiv:1907.05922]

Friedmann, A. 1922, Zeitschrift fur Physik, 10, 377

Friedmann, A. 1924, Zeitschrift fur Physik, 21, 326

Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2021, A&A, 649, A1, [arXiv:2012.01533]

Gall, E. E. E., Taubenberger, S., Kromer, M., et al. 2012, MNRAS, 427, 994, [arXiv:1208.5949]

Gayathri, V., Healy, J., Lange, J., et al. 2021, ApJL, 908, L34, [arXiv:2009.14247]

Geiger, B. & Schneider, P. 1998, MNRAS, 295, 497, [arXiv:astro-ph/9707044]

Glorot, X., Bordes, A., & Bengio, Y. 2010, JMLR, 15

Goldstein, D. A. & Nugent, P. E. 2017, ApJL, 834, L5, [arXiv:1611.09459]

Goldstein, D. A., Nugent, P. E., Kasen, D. N., & Collett, T. E. 2018, ApJ, 855, 22, [arXiv:1708.00003]

Goobar, A. et al. 2017, Science, 356, 291, [arXiv:1611.00014]

Goobar, A. A., Johansson, J., Dhawan, S., et al. 2022, Transient Name Server AstroNote, 180, 1

Greene, Z. S. et al. 2013, ApJ, 768, 39, [arXiv:1303.3588]

Grillo, C., Christensen, L., Gallazzi, A., & Rasmussen, J. 2013, MNRAS, 433, 2604, [arXiv:1305.5844]

Grillo, C., Rosati, P., Suyu, S. H., et al. 2020, ApJ, 898, 87, [arXiv:2001.02232]

Grillo, C. et al. 2016, ApJ, 822, 78, [arXiv:1511.04093]

Grillo, C. et al. 2018, ApJ, 860, 94, [arXiv:1802.01584]

Guy, J. et al. 2007, A&A, 466, 11, [arXiv:astro-ph/0701828]

Hamana, T. et al. 2003, ApJ, 597, 98, [arXiv:astro-ph/0210450]

Harvey, G. M. 1979, The Observatory, 99, 195

Heaviside, O. 1922, Electromagnetic theory, Vol. 3 (Benn brothers)

Hillebrandt, W., Kromer, M., Röpke, F. K., & Ruiter, A. J. 2013, Front. Phys.(Beijing), 8, 116, [arXiv:1302.6420]

Hillebrandt, W. & Niemeyer, J. C. 2000, Annu. Rev. Astron. Astrophys., 38, 191, [astro-ph/0006305]

Hobson, M. P., Efstathiou, G. P., & Lasenby, A. N. 2006, General Relativity: An Introduction for Physicists (Cambridge University Press)

Hochreiter, S. & Schmidhuber, J. 1997, Neural computation, 9, 1735

Hornik, K., Stinchcombe, M., & White, H. 1989, Neural Networks, 2, 359

Hoyle, F. & Fowler, W. A. 1960, ApJ, 132, 565

Hubble, E. 1929, Proceedings of the National Academy of Sciences, 15, 168, [https://www.pnas.org/doi/pdf/10.1073/pnas.15.3.168]

Hubble, E. P. 1925, The Observatory, 48, 139

Hubble, E. P. 1926, ApJ, 64, 321

Huber, S., Suyu, S. H., Ghoshdastidar, D., et al. 2022, A&A, 658, A157, [arXiv:2108.02789]

Huber, S., Suyu, S. H., Noebauer, U. M., et al. 2021, A&A, 646, A110, [arXiv:2008.10393]

Huber, S. et al. 2019, A&A, 631, A161, [arXiv:1903.00510]

Huchra, J., Gorenstein, M., Kent, S., et al. 1985, AJ, 90, 691

Irwin, M. J., Webster, R. L., Hewett, P. C., Corrigan, R. T., & Jedrzejewski, R. I. 1989, AJ, 98, 1989

Ivezić, v. et al. 2019, ApJ, 873, 111, [arXiv:0805.2366]

Jacobs, C., Collett, T., Glazebrook, K., et al. 2019, ApJS, 243, 17, [arXiv:1905.10522]

Jacobs, C., Collett, T., Glazebrook, K., et al. 2019, MNRAS, 484, 5330, [arXiv:1811.03786]

Jee, M., White, R. L., Benitez, N., et al. 2004, ApJ, 618, 46, [arXiv:astro-ph/0409304]

Jeffery, D. J., Leibundgut, B., Kirshner, R. P., et al. 1992, ApJ, 397, 304

Jiménez-Vicente, J., Mediavilla, E., Kochanek, C. S., & Muñoz, J. A. 2015, ApJ, 799, 149, [arXiv:1407.3955]

Kasen, D., Thomas, R. C., & Nugent, P. 2006, ApJ, 651, 366, [arXiv:astro-ph/0606111]

Kawamata, R., Oguri, M., Ishigaki, M., Shimasaku, K., & Ouchi, M. 2016, ApJ, 819, 114, [arXiv:1510.06400]

Kayser, R. & Refsdal, S. 1983, A&A, 128, 156

Kayser, R., Refsdal, S., & Stabell, R. 1986, A&A, 166, 36

Kelly, P., Zitrin, A., Oguri, M., et al. 2022, Transient Name Server AstroNote, 169, 1

Kelly, P. L., Fox, O. D., Filippenko, A. V., et al. 2014, ApJ, 790, 3, [arXiv:1403.4250]

Kelly, P. L. et al. 2015, Science, 347, 1123, [arXiv:1411.6009]

Kelly, P. L. et al. 2016a, ApJ, 819, L8, [arXiv:1512.04654]

Kelly, P. L. et al. 2016b, ApJ, 831, 205, [arXiv:1512.09093]

Kerzendorf, W. E. & Sim, S. A. 2014, MNRAS, 440, 387, [arXiv:1401.5469]

Khetan, N. et al. 2021, A&A, 647, A72, [arXiv:2008.07754]

King, L. & Schneider, P. 2001, A&A, 369, 1, [arXiv:astro-ph/0012202]

Kingma, D. P. & Ba, J. 2014, arXiv e-prints (arXiv:1412.6980), [arXiv:1412.6980]

Kochanek, C. S. 2020, MNRAS, 493, 1725, [arXiv:1911.05083]

Kochanek, C. S., White, M., Huchra, J., et al. 2003, ApJ, 585, 161, [arXiv:astro-ph/0208168]

Kormann, R., Schneider, P., & Bartelmann, M. 1994, A&A, 284, 285

Krishnan, C., Mohayaee, R., Colgáin, E. Ó., Sheikh-Jabbari, M. M., & Yin, L. 2021, CQG, 38, 184001, [arXiv:2105.09790]

Krishnan, C., Mohayaee, R., Colgáin, E. O., Sheikh-Jabbari, M. M., & Yin, L. 2022, Phys. Rev. D, 105, 063514, [arXiv:2106.02532]

Kromer, M., Fremling, C., Pakmor, R., et al. 2016, MNRAS, 459, 4428, [arXiv:1604.05730]

Kromer, M. & Sim. 2009, MNRAS, 398, 1809, [arXiv:0906.3152]

Kuchner, M. J., Kirshner, R. P., Pinto, P. A., & Leibundgut, B. 1994, ApJL, 426, 89

Kunugise, T. & Iwamoto, K. 2007, Publications of the Astronomical Society of Japan, 59, L57, [https://academic.oup.com/pasj/article-pdf/59/6/L57/18531195/pasj59-0L57.pdf]

Leavitt, H. S. 1908, Annals of Harvard College Observatory, 60, 87

Leavitt, H. S. & Pickering, E. C. 1912, Harvard College Observatory Circular, 173, 1

Lee, M. G., Freedman, W. L., & Madore, B. F. 1993, ApJ, 417, 553

Lemaître, G. 1927, Annales de la Société Scientifique de Bruxelles, 47, 49

Lentz, E. J., Baron, E., Branch, D., Hauschildt, P. H., & Nugent, P. E. 2000, ApJ, 530, 966, [arXiv:astro-ph/9906016]

Li, W., Bloom, J. S., Podsiadlowski, P., et al. 2011, Nature, 480, 348, [arXiv:1109.1593]

Liao, K., Biesiada, M., & Zhu, Z.-H. 2022, [arXiv:2207.13489]

Livio, M. & Mazzali, P. 2018, Phys. Rept., 736, 1, [arXiv:1802.03125]

Lochner, M., Scolnic, D. M., Awan, H., et al. 2018, arXiv e-prints, [arXiv:1812.00515]

Lochner, M. et al. 2022, ApJS, 259, 58, [arXiv:2104.05676]

LSST Science Collaboration. 2009, arXiv e-prints (arXiv:0912.0201), [arXiv:0912.0201]

Lucy, L. B. 1999, A&A, 344, 282

Lucy, L. B. 2002, A&A, 384, 725, [arXiv:astro-ph/0107377]

Lucy, L. B. 2003, A&A, 403, 261, [arXiv:astro-ph/0303202]

Lucy, L. B. 2005, A&A, 429, 19, [arXiv:astro-ph/0409249]

Maas, A. L., Hannun, A. Y., & Ng, A. Y. 2013, in ICML Workshop on Deep Learning for Audio, Speech and Language Processing

Mattingly, D. 2005, Living Rev. Rel., 8, 5, [arXiv:gr-qc/0502097]

Maxwell, J. C. 1873, A treatise on electricity and magnetism, Vol. 1 (Clarendon press)

McCully, C., Keeton, C. R., Wong, K. C., & Zabludoff, A. I. 2017, ApJ, 836, 141, [arXiv:1601.05417]

McKenzie, R. H. 1985, J. Math. Phys. (N.Y.); (United States), 26:7

Mediavilla, E., Muñoz, J. A., Garzón, F., & Mahoney, T. J. 2016, Astrophysical Applications of Gravitational Lensing

Mellier, Y. 1999, Ann. Rev. Astron. Astrophys., 37, 127, [arXiv:astro-ph/9812172]

Mihalas, D. & Mihalas, B. W. 1984, Foundations of radiation hydrodynamics

Millon, M. et al. 2020, A&A, 639, A101, [arXiv:1912.08027]

Minkowski, R. 1941, PASP, 53, 224

Moll, R. & Woosley, S. E. 2013, ApJ, 774, 137, [arXiv:1303.0324]

More, A., Suyu, S. H., Oguri, M., More, S., & Lee, C.-H. 2017, ApJ, 835, L25, [arXiv:1611.04866]

Mörtsell, E., Johansson, J., Dhawan, S., et al. 2020, MNRAS, 496, 3270, [arXiv:1907.06609]

Moustakas, L., Anguita, T., Chartas, G., et al. 2019, BAAS, 51, 487

Mueller, J. & Thyagarajan, A. 2016, Proceedings of the AAAI Conference on Artificial Intelligence, 30

Mukherjee, S., Wandelt, B. D., Nissanke, S. M., & Silvestri, A. 2021, Phys. Rev. D, 103, 043520, [arXiv:2007.02943]

Nielsen, M. T. B., Gilfanov, M., Bogdán, A., Woods, T. E., & Nelemans, G. 2014, MNRAS, 442, 3400, [arXiv:1402.2896]

Niikura, H., Takada, M., Yasuda, N., et al. 2017, [arXiv:1701.02151]

Nikolaus, K. S. & Hundertmark, M. 2018, arXiv e-prints, arXiv:1804.10136, [arXiv:1804.10136]

Nix, D. & Weigend, A. 1994, in Proceedings of 1994 IEEE International Conference on Neural Networks (ICNN'94), Vol. 1, 55–60 vol.1

Noebauer, U. M., Kromer, M., Taubenberger, S., et al. 2017, MNRAS, 472, 2787, [arXiv:1706.03613]

Nomoto, K. 1982, ApJ, 253, 798

Nomoto, K., Thielemann, F.-K., & Yokoi, K. 1984, ApJ, 286, 644

Nugent, P., Baron, E., Branch, D., Fisher, A., & Hauschildt, P. H. 1997, ApJ, 485, 812

Nwankpa, C., Ijomah, W., Gachagan, A., & Marshall, S. 2018, arXiv e-prints, arXiv:1811.03378, [arXiv:1811.03378]

Oguri, M. 2010, PASJ, 62, 1017, [arXiv:1005.3103]

Oguri, M. 2019, Rept. Prog. Phys., 82, 126901, [arXiv:1907.06830]

Oguri, M. 2021, PASP, 133, 074504, [arXiv:2106.11464]

Oguri, M. & Kawano, Y. 2003, MNRAS, 338, L25, [arXiv:astro-ph/0211499]

Oguri, M. & Marshall, P. J. 2010, MNRAS, 405, 2579, [arXiv:1001.2037]

Oguri, M., Rusu, C. E., & Falco, E. E. 2014, MNRAS, 439, 2494, [arXiv:1309.5408]

Paczynski, B. 1986, ApJ, 304, 1

Pakmor, R., Hachinger, S., Röpke, F. K., & Hillebrandt, W. 2011, A&A, 528, A117, [arXiv:1102.1354]

Pakmor, R., Kromer, M., Taubenberger, S., et al. 2012, ApJL, 747, L10, [arXiv:1201.5123]

Pakmor, R., Kromer, M., Taubenberger, S., & Springel, V. 2013, ApJL, 770, L8, [arXiv:1302.2913]

Pakmor, R., Röpke, F., Hillebrandt, W., et al. 2010, in Progenitors and Environments of Stellar Explosions, 62

Pankey, Jr., T. 1962, PhD thesis, HOWARD UNIVERSITY.

Pascanu, R., Mikolov, T., & Bengio, Y. 2013, in ICML

Paszke, A., Gross, S., Massa, F., et al. 2019, in Advances in Neural Information Processing Systems 32, ed. H. Wallach, H. Larochelle, A. Beygelzimer, F. Alche-Buc, E. Fox, & R. Garnett (Curran Associates, Inc.), 8024–8035

Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, JMLR, 12, 2825

Penzias, A. A. & Wilson, R. W. 1965, ApJ, 142, 419

Perlmutter, S., Aldering, G., Goldhaber, G., et al. 1999, ApJ, 517, 565, [arXiv:astro-ph/9812133]

Persic, M., Salucci, P., & Stel, F. 1996, MNRAS, 281, 27, [arXiv:astro-ph/9506004]

Pesce, D. W. et al. 2020, ApJL, 891, L1, [arXiv:2001.09213]

Petrillo, C. E., Tortora, C., Chatterjee, S., et al. 2019, MNRAS, 482, 807, [arXiv:1807.04764]

Philipp, G., Song, D. X., & Carbonell, J. G. 2018, ArXiv, abs/1712.05577

Phillips, M. M. 1993, ApJL, 413, L105

Pierel, J. D. R. & Rodney, S. 2019, ApJ, 876, 107, [arXiv:1902.01260]

Pietrzyński, G., Graczyk, D., Gallenne, A., et al. 2019, Nature, 567, 200, [arXiv:1903.08096]

Planck Collaboration. 2014, A&A., 571, A16, [arXiv:1303.5076]

Planck Collaboration. 2016, A&A., 594, A13, [arXiv:1502.01589]

Planck Collaboration. 2020, A&A., 641, A1, [arXiv:1807.06205]

Quimby, R. M., Oguri, M., More, A., et al. 2014, Science, 344, 396, [arXiv:1404.6014]

Quimby, R. M. et al. 2013, ApJL, 768, L20, [arXiv:1302.2785]

Refsdal, S. 1964, MNRAS, 128, 307

Reid, M. J., Pesce, D. W., & Riess, A. G. 2019, ApJL, 886, L27, [arXiv:1908.05625]

Riess, A. G., Casertano, S., Yuan, W., et al. 2021, ApJL, 908, L6, [arXiv:2012.08534]

Riess, A. G., Casertano, S., Yuan, W., Macri, L. M., & Scolnic, D. 2019, ApJ, 876, 85, [arXiv:1903.07603]

Riess, A. G. et al. 1998, Astron. J., 116, 1009, [arXiv:astro-ph/9805201]

Riess, A. G. et al. 2016, ApJ, 826, 56, [arXiv:1604.01424]

Riess, A. G. et al. 2018, ApJ, 861, 126, [arXiv:1804.10655]

Riess, A. G. et al. 2022, ApJL, 934, L7, [arXiv:2112.04510]

Rizzi, L., Tully, R. B., Makarov, D., et al. 2007, ApJ, 661, 815, [arXiv:astro-ph/0701518]

Rodney, S. A., Brammer, G. B., Pierel, J. D. R., et al. 2021, arXiv e-prints (arXiv:2106.08935), [arXiv:2106.08935]

Rubin, V. C., Ford, W. K., J., & Thonnard, N. 1980, ApJ, 238, 471

Rusu, C. E., Fassnacht, C. D., Sluse, D., et al. 2017, MNRAS, 467, 4220, [arXiv:1607.01047]

Rusu, C. E. et al. 2019, arXiv e-prints (arXiv:1907.05922), [arXiv:1905.09338]

Saeed, M. M., Al Aghbari, Z., & Alsharidah, M. 2020, PeerJ Comput. Sci., 6, e321

Sak, H., Senior, A., & Beaufays, F. 2014, arXiv preprint arXiv:1402.1128

Sandage, A. 1958, ApJ, 127, 513

Saunders, C. et al. 2018, ApJ, 869, 167, [arXiv:1810.09476]

Schechter, P. L., Pooley, D., Blackburne, J. A., & Wambsganss, J. 2014, ApJ, 793, 96, [arXiv:1405.0038]

Schmidt, B. P. et al. 1998, ApJ, 507, 46, [arXiv:astro-ph/9805200]

Schmidt, R. W. & Wambsganss, J. 2010, Gen. Relativ. Gravit., 42, 2127

Schneider, P. 1987, ApJ, 319, 9

Schneider, P., Meylan, G., Kochanek, C., et al. 2006, Gravitational Lensing: Strong, Weak and Micro: Saas-Fee Advanced Course 33, Saas-Fee Advanced Course (Springer Berlin Heidelberg)

Schneider, P. & Sluse, D. 2013, A&A, 559, A37, [arXiv:1306.0901]

Schneider, P. & Sluse, D. 2014, A&A, 564, A103, [arXiv:1306.4675]

Schuldt, S., Cañameras, R., Shu, Y., et al. 2022a, [arXiv:2206.11279]

Schuldt, S., Chirivì, G., Suyu, S. H., et al. 2019, A&A, 631, A40, [arXiv:1901.02896]

Schuldt, S., Suyu, S. H., Canameras, R., et al. 2022b, [arXiv:2207.10124]

Scolnic, D. et al. 2021, [arXiv:2112.03863]

Seitenzahl, I. R., Ciaraldi-Schoolmann, F., Röpke, F. K., et al. 2013, MNRAS, 429, 1156, [arXiv:1211.3015]

Shajib, A. et al. 2020, MNRAS, 494, 6072, [arXiv:1910.06306]

Shajib, A. J., Treu, T., & Agnello, A. 2018, MNRAS, 473, 210, [arXiv:1709.01517]

Shapiro, I. I. 1964, PRL, 13, 789

Sherstinsky, A. 2020, Physica D: Nonlinear Phenomena, 404, 132306

Shu, Y., Bolton, A. S., Mao, S., et al. 2018, ApJ, 864, 91, [arXiv:1803.07569]

Shu, Y., Cañameras, R., Schuldt, S., et al. 2022, arXiv preprint arXiv:2201.11135

Sim, S. A., Röpke, F. K., Hillebrandt, W., et al. 2010, ApJL, 714, L52, [arXiv:1003.2917]

Sim, S. A., Seitenzahl, I. R., Kromer, M., et al. 2013, MNRAS, 436, 333, [arXiv:1308.4833]

Slipher, V. M. 1917, Proceedings of the American Philosophical Society, 56, 403

Sluse, D., Rusu, C. E., Fassnacht, C. D., et al. 2019, MNRAS, 490, 613, [arXiv:1905.08800]

Soltis, J., Casertano, S., & Riess, A. G. 2021, ApJL, 908, L5, [arXiv:2012.09196]

Sonnenfeld, A., Chan, J. H. H., Shu, Y., et al. 2018, PASJ, 70, S29, [arXiv:1704.01585]

Sonnenfeld, A., Verma, A., More, A., et al. 2020, A&A, 642, A148, [arXiv:2004.00634]

Sun, S., Cao, Z., Zhu, H., & Zhao, J. 2019, arXiv e-prints, arXiv:1906.06821, [arXiv:1906.06821]

Suyu, S. H., Auger, M. W., Hilbert, S., et al. 2013, ApJ, 766, 70, [arXiv:1208.6010]

Suyu, S. H. & Halkola, A. 2010, A&A, 524, A94, [arXiv:1007.4815]

Suyu, S. H., Hensel, S. W., McKean, J. P., et al. 2012, ApJ, 750, 10, [arXiv:1110.2536]

Suyu, S. H., Marshall, P. J., Auger, M. W., et al. 2010, ApJ, 711, 201, [arXiv:0910.2773]

Suyu, S. H., Treu, T., Hilbert, S., et al. 2014, ApJL, 788, L35, [arXiv:1306.4732]

Suyu, S. H. et al. 2017, MNRAS, 468, 2590, [arXiv:1607.00017]

Suyu, S. H. et al. 2020, A&A, 644, A162, [arXiv:2002.08378]

Tewes, M., Courbin, F., & Meylan, G. 2013, A&A, 553, A120, [arXiv:1208.5598]

Tihhonova, O. et al. 2018, MNRAS, 477, 5657, [arXiv:1711.08804]

Tisserand, P. et al. 2007, A&A, 469, 387, [arXiv:astro-ph/0607207]

Truran, J. W., Arnett, W. D., & Cameron, A. G. W. 1967, Canadian Journal of Physics, 45, 2315

Tucker, W., Blanco, P., Rappoport, S., et al. 1998, ApJL, 496, L5, [arXiv:astro-ph/9801120]

Turatto, M. 2003, Lect. Notes Phys., 598, 21, [arXiv:astro-ph/0301107]

Turner, M. S. 1999, ASP Conf. Ser., 165, 431, [arXiv:astro-ph/9811454]

Valentino, E. D., Mena, O., Pan, S., et al. 2021, CQG, 38, 153001

Verde, L., Treu, T., & Riess, A. G. 2019, arXiv e-prints (arXiv:1907.10625), [arXiv:1907.10625]

Vernardos, G. & Fluke, C. J. 2013, MNRAS, 434, 832, [arXiv:1306.3722]

Vernardos, G. & Fluke, C. J. 2014, Astron. Comput., 6, 1, [arXiv:1406.0559]

Vernardos, G., Fluke, C. J., Bate, N. F., & Croton, D. 2014, ApJS, 211, 16, [arXiv:1401.7711]

Vernardos, G., Fluke, C. J., Bate, N. F., Croton, D., & Vohl, D. 2015, ApJS, 217, 23, [arXiv:1503.00770]

Vikhlinin, A., Kravtsov, A., Forman, W., et al. 2006, ApJ, 640, 691, [arXiv:astro-ph/0507092]

Vogl, C. 2020, PhD thesis, Munich University of Technology, Germany

Vogl, C., Kerzendorf, W. E., Sim, S. A., et al. 2020, A&A, 633, A88, [arXiv:1911.04444]

Vogl, C., Sim, S. A., Noebauer, U. M., Kerzendorf, W. E., & Hillebrandt, W. 2019, A&A, 621, A29, [arXiv:1811.02543]

Walker, E. S., Hachinger, S., Mazzali, P. A., et al. 2012, MNRAS, 427, 103, [arXiv:1208.4130]

Walsh, D., Carswell, R. F., & Weymann, R. J. 1979, Nature, 279, 381

Wambsganss, J. 2006, in 33rd Advanced Saas Fee Course on Gravitational Lensing: Strong, Weak, and Micro, 453–540, [arXiv:astro-ph/0604278]

Wambsganss, J., Witt, H. J., & Schneider, P. 1992, A&A, 258, 591

Weisenbach, L., Schechter, P. L., & Pontula, S. 2021, ApJ, 922, 70, [arXiv:2105.08690]

Whelan, J. & Iben, Jr., I. 1973, ApJ, 186, 1007

Wilson, A. G. & Izmailov, P. 2020, arXiv e-prints, arXiv:2002.08791, [arXiv:2002.08791]

Wittman, D. M., Tyson, J. A., Kirkman, D., Dell'Antonio, I., & Bernstein, G. 2000, Nature, 405, 143, [arXiv:astro-ph/0003014]

Wojtak, R., Hjorth, J., & Gall, C. 2019, MNRAS, 487, 3342, [arXiv:1903.07687]

Wong, K. C. et al. 2020, MNRAS, 498, 1420, [arXiv:1907.04869]

Woosley, S. E. & Weaver, T. A. 1986, Annu. Rev. Astron. Astrophys., 24, 205

Wulczyn, E., Steiner, D. F., Xu, Z., et al. 2020, PLoS One, 15, e0233678

Yahalomi, D. A., Schechter, P. L., & Wambsganss, J. 2017, arXiv e-prints (arXiv:1711.07919), [arXiv:1711.07919]

Yıldırım, A., Suyu, S. H., & Halkola, A. 2020, MNRAS, 493, 4783, [arXiv:1904.07237]

Yıldırım, A., van den Bosch, R. C. E., van de Ven, G., et al. 2017, MNRAS, 468, 4216, [arXiv:1701.05898]

Zwicky, F. 1937, Phys. Rev., 51, 290

7 Appendix

7.1 Results for a LSTM-FCNN trained only on theoretical SN Ia models

In this Appendix, we show a LSTM-FCNN trained on data as in Chapter 5, with the exception that we use only light curves from theoretical models. We summarize our assumptions on the data sets in Tab. 7.1 in contrast to Tab. 5.2. Further, we use all empirical light curves from the `SNEMO15` model to test the generalizability. The results are presented in Fig. 7.1, where we see that the bias increases due to the not fully ideal generalization of the theoretical SN Ia models to the empirical `SNEMO15` model. However, the bias is still on a controllable level if only LSNe Ia with time delays longer than 20 days are considered. Still, the comparison to Fig. 5.9 shows that the performance of the LSTM-FCNN improves if 50% of the `SNEMO15` light curves are used in the training process.

Table 7.1. Description of the training set, validation set, test set, and `SNEMO15` data set used in this appendix in contrast to Table 5.2 and the rest of this work.

Training set	Light curves based on theoretical SN Ia models
Validation set	Light curves based on theoretical SN Ia models
Test set	Light curves based on theoretical SN Ia models
`SNEMO15` data set	Light curves based on the 171 `SNEMO15` light curves

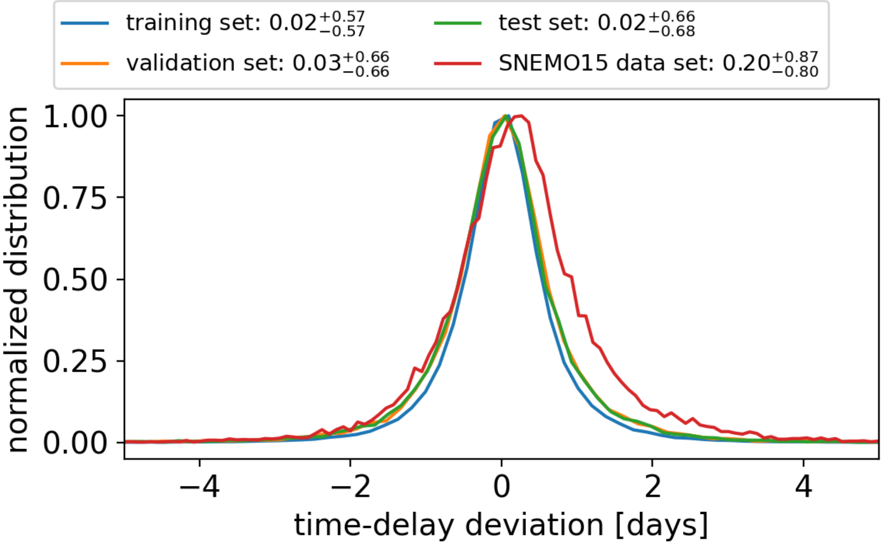


Fig. 7.1. A LSTM-FCNN as in Fig. 5.9, but this time only trained on theoretical models. The composition of all listed data sets is explained in Table 7.1.