

Dynamics Discovery: a Next Generation Dynamical Astronomical Event Model and Classifier for Next Generation Surveys

JARED HAND¹

¹ *Department of Physics and Astronomy,
University of Pittsburgh,*

1. INTRODUCTION, PAST PROJECTS, AND CURRENT RESEARCH

Next generation surveys will observe millions of dynamic astronomical events over the coming decades and transform time domain (TD) astronomy. This transformation will occur only after we accurately and efficiently model, and then classify, these newly observed TD events. Such a demand is placing strain on existing tools, so achieving any scientific leap requires an innovative advance in modeling. It is this type of innovation I demonstrate in my current research and that I propose for this 2023 NASA Hubble Fellowship application: a novel TD event dynamical discovery model to directly obtain these evolving systems' underlying dynamics.

1.1. *Past Research: Understanding Supernova Cosmology Systematics*

SN Ia are standardizable candles: correlations between observed properties and peak brightness are used to reduce peak brightness dispersion via standardization. Exploiting these empirical relations led directly to the discovery of dark energy (Riess et al. 1998; Perlmutter et al. 1999). Ideally, correlations between SN Ia and intrinsic or extrinsic modes of spectral variation are accounted for during this standardization procedure, but established empirical relationships between host galaxy properties such as stellar mass, and peak SN Ia brightness appear after standardization in a phenomenon coined the mass step (Sullivan et al. 2010; Rigault et al. 2020). Specifically, the difference between current best-fit cosmological model's predicted peak brightness and standardized peak brightness co-varies with host mass in a step-like fashion.

My first project quantified the influence observation method (broad-band photometry or spectroscopy) or host galaxy property fitting technique had on the host bias measurement (Hand et al. 2022). Given our ignorance of the host bias and its source, considering how we observe SN Ia was an important place to start our search. With the PISCO SN Ia host observation subsample (Galbany et al. 2018) that had corresponding SDSS and GALEX photometric coverage (Alam et al. 2015; Morrissey et al. 2007), we derived independent stellar mass and star formation rate estimates to compare to our standardized SN Ia properties. We found no evidence that neither observation technique or fitting method biases or produces the SN Ia host bias and subsequent mass step.

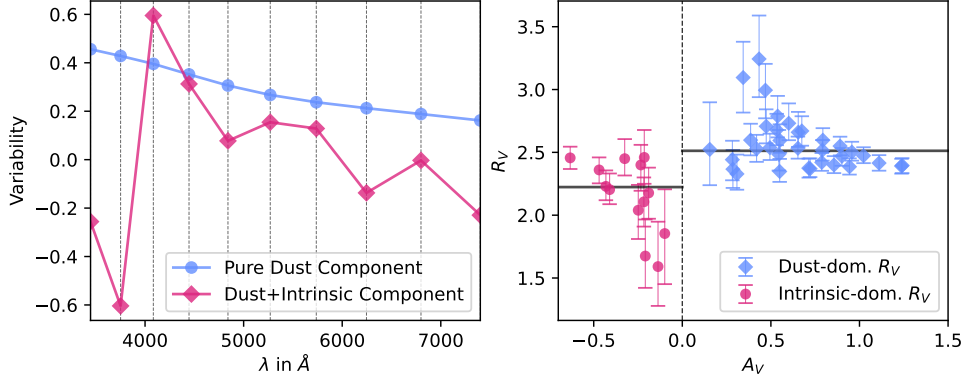
1.2. *Current Research: a New Type Ia Supernova Model*

As a recipient of a 2019 Department of Energy Office of Science SCGSR Award, I relocated to the Lawrence Berkeley National Laboratory in August of that year to work under Dr. Alex Kim developing a new empirical SN Ia model. SN Ia theory is insufficiently developed to provide parametric models that account for observed SN Ia spectral variation, so the SN Ia researchers develop increasingly sophisticated empirical models trained on sets of observation time series.

Such empirical models must account for intrinsic and extrinsic modes of SN Ia population variability. Unlike intrinsic variation, extrinsic variation is variation that is independent of the progenitor or explosion physics, largely arising from differing dust attenuation. Extrinsic variability is phase-independent, so most SN Ia empirical models assume a single-component effective attenuation curve (Jha et al. 2007; Burns et al. 2011; Mandel et al. 2022), or group all phase-independent variation (intrinsic or extrinsic) into a single template (Guy et al. 2007; Kenworthy et al. 2021). Neither approach is satisfactory — physical consideration alone requires at least two dust attenuation curves: one quantifying optical depth and the other extracting wavelength dependence (or curve steepness) (Weingartner & Draine 2001). Effective dust attenuation models or agnostic color laws ignore per-SN variability in curve slope, a constraint that may contribute to the mass step (Brout et al. 2019; Popovic et al. 2021). Cosmological parameter uncertainty derived from SN Ia distance ladders is now dominated by SN Ia model systematics (Scolnic et al. 2018), so improvements in SN Ia cosmology will follow superior SN Ia empirical modeling.

Dr. Kim and I developed a SN Ia model that simultaneously fits two physics-agnostic, phase-independent color templates and a single phase-dependent variation template using Stan (Carpenter et al. 2017). This model is the first SN Ia model architecture to successfully fit two physics-agnostic, phase-independent spectral variation templates, better constraining SN Ia dust attenuation than previous efforts. We make innovative use of bivectors to represent the model’s two-dimensional wavelength subspace. We further exploit our model’s geometry to decompose the best-fit bivector into physically interpretable components with an application of geometric algebra. This is the first-ever application of its kind and allows us to intersect our two-dimensional wavelength subspace with other two-dimensional dust attenuation parameterizations, naturally extracting the intrinsic and extrinsic components of our wavelength subspace. I developed and implemented the bivector representation and geometric algebra techniques, both of which were fundamental to finally separating intrinsic and extrinsic modes of phase-independent spectral variation.

Instead of extracting two dust-sourced color templates, we recover one dust-like component and another that is a nontrivial mixture of extrinsic and intrinsic features (left-hand plot Figure 1). The phase-independent intrinsic features (notably, Ca H&K) are relatively constant during the first ten weeks of a SN Ia’s evolution (Branch et al. 1993). In recovering effectively both intrinsic and extrinsic phase-independent variation, our results bring into question the narrative that unmodeled

**Figure 1.**

SN Ia systematics results from incomplete extrinsic variation modeling (Brout et al. 2019).

These two extracted color templates allow us to explore dust variation between SNe Ia via another hierarchical Bayesian dust model. We find that SN Ia dominated by intrinsic phase-independent variation (SNe Ia with little to no dust attenuation) systematically fit for steeper attenuation curves than their dust-dominated counterparts. Nominally quantified with the total-to-selective ratio R_V , past estimates of SN Ia effective attenuation curves have consistently been steeper (or have a lower R_V) than the Milky Way sight line average of $R_V = 3.1$. Our results find having intrinsic-dominated SN Ia in one's sample biases your effective R_V estimate lower than it otherwise should be, although this does not entirely account for suspiciously low SN Ia R_V values. Figure 1, right-hand plot compares recovered absolute attenuation (A_V) to R_V , while the black solid lines in correspond to these respective subsample's discrepant median R_V values.

A paper summarizing our new model is undergoing internal review within the Nearby Supernova Factory group. Meanwhile, Dr. Kim and I are increasing this model's wavelength resolution and integrating a two-component dust attenuation recipe directly into its architecture. This higher resolution will enable quantification of spectral feature properties and a built-in attenuation model will better separate intrinsic phase-independent variation from its extrinsic counterpart. This newer model, by directly extracting per-SN R_V estimates, will allow for a first-ever comparison of SN-derived R_V values to local environment R_V estimates. Such a comparison will be a valuable asset for answering outstanding SN Ia dust attenuation questions.

2. PROPOSAL: DISCOVERING TIME DOMAIN ASTRONOMY'S UNDERLYING DYNAMICS

SN Ia science is one part of the much larger field of TD astronomy, a field to be revolutionized by NASA's Roman Space Telescope, set to launch in 2026 (Akeson et al. 2019), and the Rubin Observatory's Legacy Survey of Space and Time (LSST), set for first light expected in 2023. LSST alone is expected to produce ≈ 20 TB/night

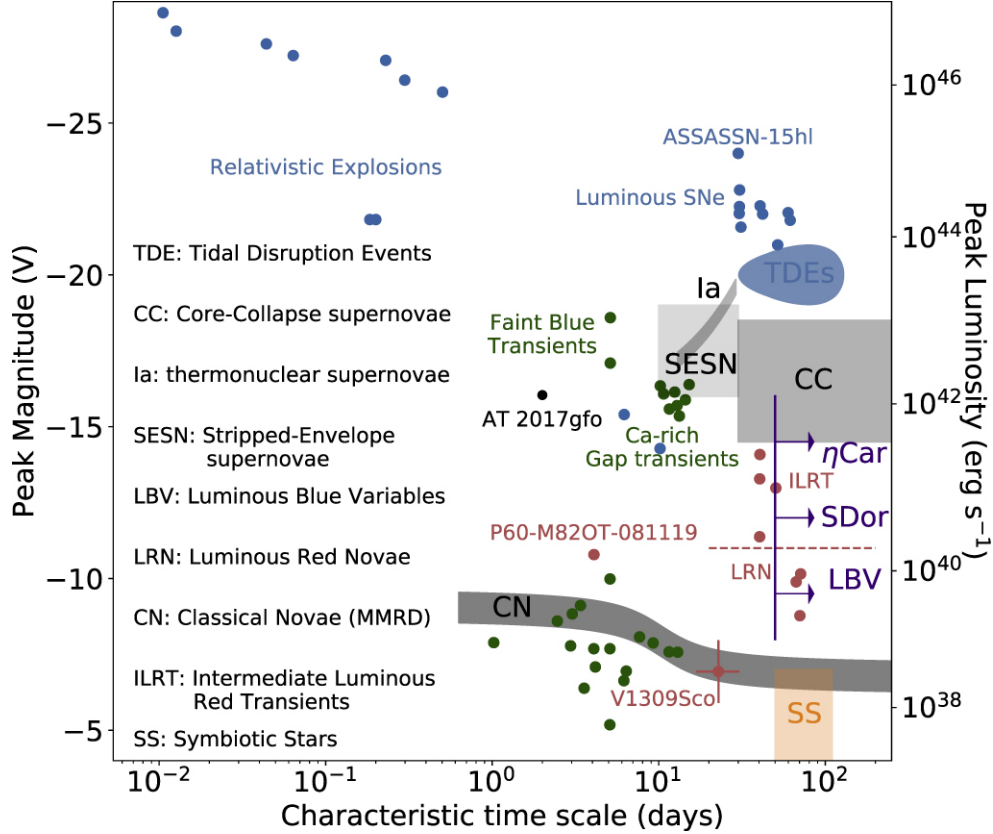


Figure 2.

of raw data (Ivezić et al. 2019). Figure 2 is a phase space diagram from Ivezić et al. (2019) comparing timescales and peak brightness variation between different transient classes to be observed by LSST and Roman. Not included in said figure are equally interesting periodic events such as variable stars (such as Cepheids or RR Lyrae stars), transiting exoplanets, or microlensing events. All TD classes will see orders of magnitude increases in discovery and observation rates, and accurately distinguishing between observations will require a TD model that directly captures these events' underlying dynamical properties. Many TD event types are not well understood because we have yet to fully observed their evolution, or because we have observed too few examples, and the statistics provided by Roman and LSST will rapidly fill these gaps in our empirical knowledge. These surveys will also allow us to discover phenomena in currently inaccessible regions of TD phase space. This will only happen after we accurately *and* efficiently characterize and classify the millions of expected per-day TD observations (Narayan et al. 2018).

Research groups are developing alert-broker pipelines to flag TD events in Roman's and LSST's data streams and then distribute classified alerts to interested science teams (Narayan et al. 2018; Raen 2022). Roman and LSST will observe too many TD events per night to classify via spectroscopic followup or human investigation, instead relying on classifier algorithms. These algorithms require TD observation time series

be represented by some model (such as a template or wavelet decomposition model), with representation components then serving as the classifier’s feature inputs (Malz et al. 2019). There is ample room for classification algorithm improvement. For example, stripped envelope core-collapse SNe classification false negative rates sit at $\approx 10\%$ with current LSST algorithms (Malz et al. 2019). Improving this number is an important precursor to LSST cosmology data product efficiency.

Current representation models ignore something very fundamental: TD events are dynamical systems. Figure 2 makes clear that dynamical properties, such as amplitude and timescale, are defining features that differentiate types of TD events. Therefore, I propose discovering each TD event’s underlying dynamics from observations. Referred to here as Dynamics Discovery, this approach is a rapidly maturing field of study (Mai et al. 2016; Raissi & Karniadakis 2018; Goyal & Benner 2022). Not only would this technique provide a physically-motivated representation for classification, but this representation would be a powerful analysis tool for probing TD event’s underlying physical processes.

2.1. *Connecting with NASA Astrophysics Research*

A Dynamics Discovery model and subsequent classifier will be directly applicable to Roman’s TD program (specifically, Roman’s High Latitude Time Domain Survey), improving its SN Ia and exoplanet observation utility to better reach NASA ‘Physics of the Cosmos’ and ‘Exoplanet Exploration’ research goals. Through LSST, this proposal will provide methods and software directly utilizable by Roman, and it is valid to read ‘Roman’ in place of ‘LSST’ throughout. Both LSST and Roman will transform TD astronomy, but LSST commissioning two years before Roman motivates my initial focus on the LSST over Roman.

2.2. *A Dynamics Discovery Approach to Time Domain Modeling and Classification*

Visualized with Figure 3, Dynamics Discovery works backwards from a traditional dynamical systems model workflow that finds the best-fit parameters from an assumed dynamical model. Dynamics Discovery instead obtains the governing dynamics via a denoised implicit representation of the measurements. Although we lack physical or empirical dynamical systems models for many TD events, but will have plenty of measurements to discover the underlying dynamics.

LSST will observe the night sky with six photometric filters *ugrizy*, each centered at different effective wavelengths. Consider a six-component vector \mathbf{F}_I , each component being a TD event’s true (or implicit) flux for one of the mentioned filters. Formally, Dynamics Discovery promotes a TD event’s flux derivative to a differential equation $\dot{\mathbf{F}}_I = g(\mathbf{F}_I, t|\boldsymbol{\Theta})$ whose functional form $g(\mathbf{F}_I, t|\boldsymbol{\Theta})$ is obtained from a set of observations $\{\mathbf{F}_O\}$ via the simultaneously obtained implicit flux evolution $\mathbf{F}_I(t)$. Here, $\boldsymbol{\Theta}$ is a parameter vector describing the discovered dynamics’ functional form.

Dynamics Discovery models the implicit representation $\mathcal{F}(t) \approx \mathbf{F}_I(t)$, where for an event’s i th observation $\mathbf{F}_I(t_i) = \mathbf{F}_{O,i} + \text{error}_i$. The dynamics is simultaneously

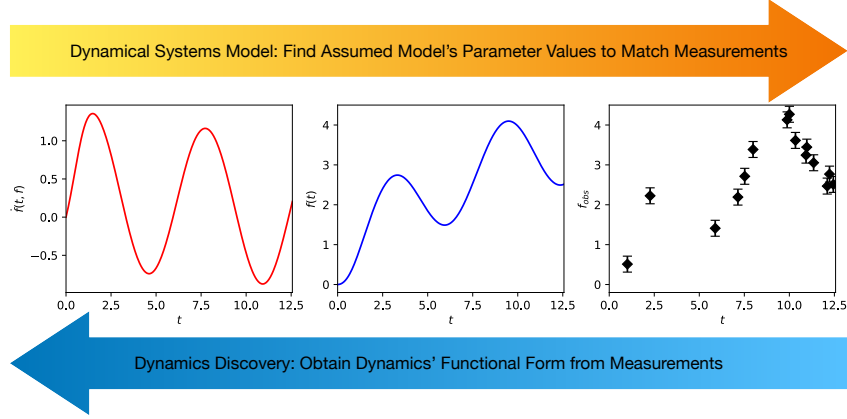


Figure 3.

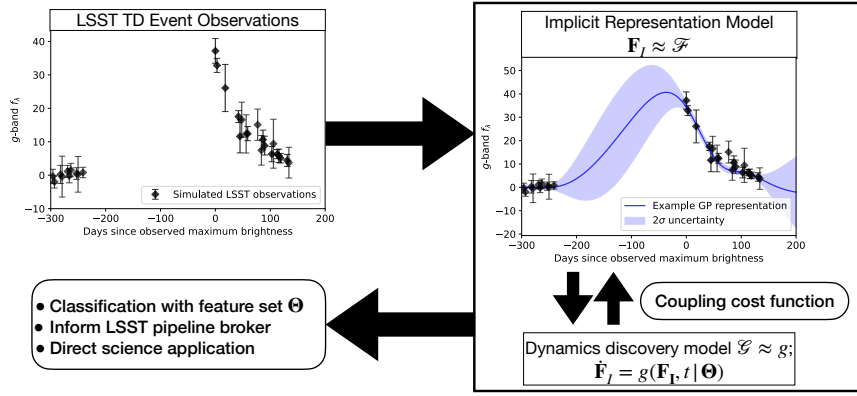


Figure 4.

modeled with a representation $\mathcal{G} \approx g(\mathbf{F}_I, t | \Theta)$. \mathcal{F} is obtained by minimizing the loss between observations and their implicit representation, while \mathcal{G} is obtained by minimizing loss between \mathcal{F} derivatives and the modeled dynamics. \mathcal{F} and \mathcal{G} are further coupled by minimizing loss between predicted time-integrated evolution to the implicit representation expected value. Figure 4 visually summarizes this workflow and Section 3 of Goyal & Benner (2022) further details a purely neural network (NN) implementation (see Raissi & Karniadakis (2018) for a Gaussian process (GP) alternative). Figure 4 visualizes the proposed model architecture with an example GP implicit representation for a simulated LSST transient’s g -band photometric time series.

Fundamental TD event properties, like time scales and amplitudes, are naturally captured by a Dynamics Discovery model. Many periodic stars exhibit transient flare-ups, while the transition from $^{56}\text{Ni} \rightarrow ^{56}\text{Co}$ to $^{56}\text{Co} \rightarrow ^{56}\text{Fe}$ decay partway through SN evolution changes the physical decay timescale of the system. Unlike the current TD models, discovered dynamics impartially accounts for multiple timescales or timescale evolution, especially if higher-order derivatives are considered. Discovered dynamics can simultaneously account for periodic, quasiperiodic, and exponential time scales,

or, unlike wavelength decomposition, nonlinear combinations thereof. Stochastic dynamics is also discoverable, which will help better us discriminate irregular TD events from noisy observations.

2.3. Year One Goals

There are two LSST-specific software deliverables expected by this project: a Dynamics Discovery model to represent TD events, and a classifier using the representation’s features as inputs. TD classification is a prerequisite to any science output: we need to know what we are looking at before we can study it. The proposed model will be developed using simulated LSST data and Zwicky Transient Facility TD observations, per common LSST practice (Malz et al. 2019; Dekany et al. 2020).

Year one will initially focus on design choice and implementation for both the \mathcal{F} and \mathcal{G} model components. Implementing \mathcal{F} with GPs has the particular advantage in simplifying error propagation and measurement derivative estimation, while a NN-based \mathcal{F} architecture may be more computationally efficient, especially for periodic objects with many observations. Selecting \mathcal{F} ’s architecture requires simultaneous development of the dynamics model \mathcal{G} . One can obtain $g(\mathbf{F}, t|\Theta)$ using GP or NN architectures (Raissi & Karniadakis 2018; Goyal & Benner 2022). Sparse basis learning from sparse dynamics function libraries is another technique that obtains the dynamics functional form from a sparse library of basis functions (Mai et al. 2016). This approach works because differential equations are almost always sparse compared to an infinite-dimensional function space basis. All of these methods are well-developed, which will considerably reduce \mathcal{G} ’s development time.

Using a Dynamics Discovery model representation requires a \mathcal{G} architecture that allows for a consistent comparison between TD events since the parameters Θ (or derived quantifies derived) will serve as a classifier’s features input. This requirement is satisfied by a sparse basis learning approach, but not for a general GP or NN implementation. Spectral submanifold (SSM) learning provides a potential \mathcal{G} NN or GP framework satisfying this consistency requirement by instead discovering the dominant phase space submanifold embedding (Cenedese et al. 2022). Derived submanifold properties, such as dimensionality or curvature, would then serve as feature inputs. Periodic events currently can be represented with existing SSM models, although some development is required for transient events (Cenedese, private communications).

2.3.1. Further Goals

Numerous classifier algorithms have been tested using template or wavelet decomposition models, with boosted decision trees currently performing best for SN classification (Malz et al. 2019). More novel classifiers, such as Dirichlet processes (Professor Chad Schafer and Dr. Kim, private communications) may be explored if time permits, but priority will be developing the Dynamics Discovery model. Integrating the developed Dynamics Discovery model into LSST’s alert-broker pipeline and the

LINCC Frameworks will be year two’s first objective; it will hopefully be started earlier. LINCC Frameworks is an initiative developing the cloud-based software and database infrastructure needed to address the unprecedented scale of LSST’s data products. A Dynamics Discovery end-user implementation is such a product LINCC Frameworks will deliver. I also will integrate the proposed model into Roman’s upcoming pipelines before its launch.

I intend to develop a generative Dynamics Discovery mixture model by exploiting the tremendous volume of raw LSST data to simultaneously obtain a mixture of generative TD class models for improved classification. Such a technique would combine Dynamics Discovery’s physics-agnosticism with informative generative models obtained from LSST’s vast pool of observations to create a particularly powerful discriminator between pre-characterized TD events. It would also allow TD subtype classification within LSST’s alert-broker, enabling rapid identification of SN Ia subtypes to decrease cosmology analysis time and largely removing time-consuming manual identification of nuanced periodic star observations as two notable LSST improvements. This tool would also be particularly enabling for Roman TD science output as well.

2.4. Carnegie Mellon University and Leadership Initiatives

Carnegie Mellon University (CMU) is home to the McWilliams Center for Cosmology with professors holding leadership positions in LSST’s Dark Energy Science Collaboration (LSST-DESC). Furthermore, CMU is a hub and founding member of the LINCC Frameworks and my presence working with Professor Chad Schafer would facilitate the proposed methodology’s integration into LSST’s software stack. Roman Telescope science is also represented at CMU. University of Pittsburgh faculty also hold LSST-DESC leadership positions and are closely connected with CMU’s LINCC Frameworks initiative, particularly LSST’s TD analysis software development. Professor Wood-Vasey’s group at the University of Pittsburgh leads development of a candidate LSST alert-broker (Raen 2022), and my being at CMU provides the proximity to more effectively integrate a Dynamics Discovery classifier into their pipeline.

My time in graduate school has been spent engaging with and leading incoming students while also advocating for historically marginalized groups and for those living with mental illness. As a person with obsessive compulsive disorder, I recognize the importance of peer support in creating a healthy environment for learning and discovery. Additionally, I have been active in the Pittsburgh area, tutoring high school students while also mentoring University of Pittsburgh physics graduate students. I started the Pitt Stan workshop to introduce students and faculty to the Stan programming language (Carpenter et al. 2017), and have been an active committee member for the University’s APS Bridge program. I look forward to continuing my work in mental health advocacy and educational leadership with the support of NASA.

REFERENCES

- Akeson, R., Armus, L., Bachelet, E., et al. 2019, arXiv e-prints, arXiv:1902.05569. <https://arxiv.org/abs/1902.05569>
- Alam, S., Albareti, F. D., Allende Prieto, C., et al. 2015, *ApJS*, 219, 12, doi: [10.1088/0067-0049/219/1/12](https://doi.org/10.1088/0067-0049/219/1/12)
- Branch, D., Fisher, A., & Nugent, P. 1993, *AJ*, 106, 2383, doi: [10.1086/116810](https://doi.org/10.1086/116810)
- Brout, D., Scolnic, D., Kessler, R., et al. 2019, *ApJ*, 874, 150, doi: [10.3847/1538-4357/ab08a0](https://doi.org/10.3847/1538-4357/ab08a0)
- Burns, C. R., Stritzinger, M., Phillips, M. M., et al. 2011, *AJ*, 141, 19, doi: [10.1088/0004-6256/141/1/19](https://doi.org/10.1088/0004-6256/141/1/19)
- Carpenter, B., Gelman, A., Hoffman, M. D., et al. 2017, *Journal of statistical software*, 76
- Cenedese, M., Ax  s, J., Yang, H., Eriten, M., & Haller, G. 2022, *Philosophical Transactions of the Royal Society of London Series A*, 380, 20210194, doi: [10.1098/rsta.2021.0194](https://doi.org/10.1098/rsta.2021.0194)
- Dekany, R., Smith, R. M., Riddle, R., et al. 2020, *PASP*, 132, 038001, doi: [10.1088/1538-3873/ab4ca2](https://doi.org/10.1088/1538-3873/ab4ca2)
- Galbany, L., Anderson, J. P., S  nchez, S. F., et al. 2018, *ApJ*, 855, 107, doi: [10.3847/1538-4357/aaaf20](https://doi.org/10.3847/1538-4357/aaaf20)
- Goyal, P., & Benner, P. 2022, arXiv e-prints, arXiv:2205.09479. <https://arxiv.org/abs/2205.09479>
- Guy, J., Astier, P., Baumont, S., et al. 2007, *A&A*, 466, 11, doi: [10.1051/0004-6361:20066930](https://doi.org/10.1051/0004-6361:20066930)
- Hand, J., Liu, S., Galbany, L., et al. 2022, *ApJ*, 925, 115, doi: [10.3847/1538-4357/ac389f](https://doi.org/10.3847/1538-4357/ac389f)
- Ivezi  ,   ., Kahn, S. M., Tyson, J. A., et al. 2019, *ApJ*, 873, 111, doi: [10.3847/1538-4357/ab042c](https://doi.org/10.3847/1538-4357/ab042c)
- Jha, S., Riess, A. G., & Kirshner, R. P. 2007, *ApJ*, 659, 122, doi: [10.1086/512054](https://doi.org/10.1086/512054)
- Kenworthy, W. D., Jones, D. O., Dai, M., et al. 2021, *ApJ*, 923, 265, doi: [10.3847/1538-4357/ac30d8](https://doi.org/10.3847/1538-4357/ac30d8)
- Mai, M., Shattuck, M. D., & O’Hern, C. S. 2016, arXiv e-prints, arXiv:1605.05420. <https://arxiv.org/abs/1605.05420>
- Malz, A. I., Hlo  ek, R., Allam, T., J., et al. 2019, *AJ*, 158, 171, doi: [10.3847/1538-3881/ab3a2f](https://doi.org/10.3847/1538-3881/ab3a2f)
- Mandel, K. S., Thorp, S., Narayan, G., Friedman, A. S., & Avelino, A. 2022, *MNRAS*, 510, 3939, doi: [10.1093/mnras/stab3496](https://doi.org/10.1093/mnras/stab3496)
- Morrissey, P., Conrow, T., Barlow, T. A., et al. 2007, *ApJS*, 173, 682, doi: [10.1086/520512](https://doi.org/10.1086/520512)
- Narayan, G., Zaidi, T., Soraisam, M. D., et al. 2018, *ApJS*, 236, 9, doi: [10.3847/1538-4365/aab781](https://doi.org/10.3847/1538-4365/aab781)
- Perlmutter, S., Aldering, G., Goldhaber, G., et al. 1999, *ApJ*, 517, 565, doi: [10.1086/307221](https://doi.org/10.1086/307221)
- Popovic, B., Brout, D., Kessler, R., Scolnic, D., & Lu, L. 2021, arXiv e-prints, arXiv:2102.01776. <https://arxiv.org/abs/2102.01776>
- Raen, T. 2022. <http://d-scholarship.pitt.edu/43428/>
- Raissi, M., & Karniadakis, G. E. 2018, *Journal of Computational Physics*, 357, 125, doi: [10.1016/j.jcp.2017.11.039](https://doi.org/10.1016/j.jcp.2017.11.039)
- Riess, A. G., Nugent, P., Filippenko, A. V., Kirshner, R. P., & Perlmutter, S. 1998, *ApJ*, 504, 935, doi: [10.1086/306106](https://doi.org/10.1086/306106)
- Rigault, M., Brinnel, V., Aldering, G., et al. 2020, *A&A*, 644, A176, doi: [10.1051/0004-6361/201730404](https://doi.org/10.1051/0004-6361/201730404)
- Scolnic, D. M., Jones, D. O., Rest, A., et al. 2018, *ApJ*, 859, 101, doi: [10.3847/1538-4357/aab9bb](https://doi.org/10.3847/1538-4357/aab9bb)
- Sullivan, M., Conley, A., Howell, D. A., et al. 2010, *Monthly Notices of the Royal Astronomical Society*, 406, 782, doi: [10.1111/j.1365-2966.2010.16731.x](https://doi.org/10.1111/j.1365-2966.2010.16731.x)
- Weingartner, J. C., & Draine, B. T. 2001, *ApJ*, 548, 296, doi: [10.1086/318651](https://doi.org/10.1086/318651)