Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **549,** 1–13 (2026)
Advance Access publication 2026 June 5

https://doi.org/10.1093/mnras/stag1066

SNID–SAGE: a modern framework for interactive supernova classification and spectral analysis

Fiorenzo Stoppa ★ and Stephen J. Smartt

Astrophysics sub-Department, Department of Physics, University of Oxford, Denys Wilkinson Building, Keble Road, Oxford OX1 3RH, UK

Accepted 2026 June 2. Received 2026 May 17; in original form 2026 March 30

ABSTRACT

We present SNID–SAGE (SuperNova IDentification–Spectral Analysis and Guided Exploration), a framework for supernova spectral classification with both a fully interactive graphical interface and a scriptable command-line pipeline for large-scale processing. The pipeline combines deterministic spectral pre-processing, FFT-based cross-correlation against a curated template library, ranking of candidate matches using a composite quality metric, and consolidation of redshift and classification solutions into a single result with associated quality and confidence estimates. SNID–SAGE includes an upgradeable template library (about 6000 spectra), interactive line identification with velocity measurements, and optional natural-language summaries of classification results. We evaluate SNID–SAGE using two complementary tests: (i) Leave-one-out cross-validation, in which each template spectrum is matched against the remainder of the library; and (ii) large-scale application to WISeREP spectra with valid coverage across the 4000–7000 Å interval, irrespective of spectral type, comprising approximately 46 000 spectra, with redshift validation against known host-galaxy measurements where available. The full validation results and the SNID–SAGE framework are publicly available, supporting integration into spectroscopic survey workflows.

Key words: methods: data analysis – methods: statistical – techniques: spectroscopic – surveys – software: development – (stars:) supernovae: general.

1 INTRODUCTION

The number and diversity of astronomical transients discovered each year are growing rapidly, driven by wide-field time-domain surveys such as Pan-STARRS (H. A. Flewelling et al. 2020), ATLAS (J. L. Tonry et al. 2018), BlackGEM (P. J. Groot et al. 2024), ZTF (E. C. Bellm et al. 2019), and GOTO (D. Steeghs et al. 2022), with further increases expected from the Rubin Observatory's Legacy Survey of Space and Time (Ž. Ivezić et al. 2019).

As discovery rates increase, the demand for spectroscopic follow-up to enable transient classification and redshift estimation grows accordingly. Classification is currently supported by a wide range of spectroscopic facilities spanning apertures from roughly 1.5–10 m, with corresponding variation in spectral resolution, sensitivity, and available follow-up time. Major contributors include Keck/LRIS (J. B. Oke et al. 1995), VLT/X-shooter (J. Vernet et al. 2011), VLT/FORS (I. Appenzeller et al. 1998), Gemini/GMOS (I. M. Hook et al. 2004), GTC/OSIRIS (J. Cepa 1998), NTT/EFOSC2 through PESSTO (S. J. Smartt et al. 2015), UH88/SNIFS (M. A. Tucker et al. 2022), Palomar/SEDM (N. Blagorodnova et al. 2018), LT/SPRAT (A. S. Piascik et al. 2014), NOT/ALFOSC (A. A. Djupvik & J. Andersen 2010), and LCO/FLOYDS (T. M. Brown et al. 2013). New multiplexed facilities such as WEAVE (S. Jin et al. 2024), 4MOST/TiDES (R. S. Jong et al. 2019; C. Frohmaier et al. 2025), and Subaru/PFS (N. Tamura et al. 2016) will further increase the volume of available spectra, shifting the bottleneck from discovery towards scalable and reproducible classification.

Over the past two decades, a small number of community-standard tools have underpinned supernova spectral classification and redshift inference. Widely used approaches include template-based matching methods, such as those implemented in SuperNova IDentification (SNID; S. Blondin & J. L. Tonry 2007) and in χ^2 fitting frameworks with host-galaxy mixing (e.g. SUPERFIT and its Python re-implementation NGSF; D. A. Howell et al. 2005; O. Yaron & A. Gal-Yam 2012), as well as web-based spectral matching services like GELATO (A. H. Harutyunyan et al. 2008) and machine-learning classifiers such as DASH (D. Muthukrishna, D. Parkinson & B. E. Tucker 2019). These tools have proven highly effective in traditional follow-up settings; however, the scale, cadence, and heterogeneity of current and forthcoming spectroscopic programmes place new demands on classification pipelines, requiring methods that remain robust across diverse transient classes and readily extensible to incorporate new or rare types without retraining or reconfiguration.

We introduce SNID–SAGE (Spectral Analysis and Guided Exploration), a spectroscopic classification framework that retains the template-based cross-correlation pioneered by SNID (J.

★ E-mail: fiorenzo.stoppa@physics.ox.ac.uk

Tonry & M. Davis 1979; S. Blondin & J. L. Tonry 2007) and adds a new core metric together with a redshift-space Gaussian mixture model (GMM) clustering to consolidate information from multiple template matches in a structured way. SNID–SAGE is provided as both a fully interactive graphical interface and a scriptable command-line pipeline. In addition to classification, it integrates tools for template curation, emission- and absorption-line analysis, and natural-language summaries, the latter remaining strictly descriptive and independent of the numerical inference.

This paper is organized as follows: Section 2 describes the SNID–SAGE classification pipeline, including pre-processing, cross-correlation, similarity metrics, and clustering-based inference. Section 3 outlines the construction and composition of the template library. Section 4 presents validation and performance analyses based on leave-one-out testing and large-scale application to WISeREP spectra. Finally, Section 5 summarizes the main results and discusses limitations and future directions.

2 METHODS

SNID–SAGE implements a deterministic spectroscopic classification pipeline with two interfaces, an interactive graphical client and a batch-oriented command-line tool. At a high level, the pipeline ingests a wavelength-calibrated one-dimensional spectrum, applies a fixed pre-processing sequence, compares the conditioned spectrum against a library of templates, and consolidates the resulting ensemble of matches into a type, phase, and redshift inference. Alongside the primary classification, the pipeline produces diagnostic outputs and alternative solutions to support interpretation and quality control.

In this section, we describe the pre-processing, cross-correlation, similarity metrics, and clustering-based consolidation steps used to derive type, redshift, and phase (days relative to maximum light) estimates from an observed spectrum.

2.1 Step 1: pre-processing

Pre-processing follows the principles of the classical SNID approach (S. Blondin & J. L. Tonry 2007), transforming an observed spectrum into a pre-processed representation suitable for Fast Fourier Transform (FFT)-based cross-correlation.

The first stage addresses data quality. Obvious artefacts such as cosmic-ray spikes and isolated outliers are identified automatically and masked prior to further processing. The user may additionally define wavelength regions to exclude, such as telluric absorption bands (e.g. the A-band at 7550–7700 Å) or known instrumental artefacts. Optional Savitzky–Golay smoothing (A. Savitzky & M. J. E. Golay 1964) can be applied to reduce high-frequency noise while preserving spectral features. In the second stage, the observed spectrum is rebinned on to the default logarithmic wavelength grid used by the template library. The templates are pre-defined on a common $\ln\lambda$ grid spanning 2500–10000 Å with 1024 uniformly spaced bins, and the input spectrum is interpolated on to this same grid prior to cross-correlation. Working in $\ln\lambda$ ensures that a cosmological redshift z corresponds to an additive shift $\Delta\ln\lambda = \ln(1+z)$; on a grid with spacing $\Delta\ln\lambda$, this is a shift of $\Delta k = \ln(1+z)/\Delta\ln\lambda$ bins in correlation space, enabling efficient FFT-based matching (J. Tonry & M. Davis 1979). The third stage removes the broad-band continuum. A low-order spline model is fitted to the log-rebinned spectrum and divided out, isolating the line structure that drives the correlation and reducing sensitivity to broad-band differences between the target and templates, consistent with the established SNID continuum treatment (S. Blondin & J. L. Tonry 2007). After continuum division, unity is subtracted to centre the normalized continuum at zero, and the spectrum is then tapered at the edges with a cosine window to suppress discontinuities that would otherwise introduce spurious high-frequency power in the Fourier domain (F. J. Harris 1978). The result is a zero-mean, flattened spectrum on a regular $\ln\lambda$ grid, ready for the cross-correlation stage. Fig. 1 illustrates the input spectrum and its final pre-processed form. Although the pre-processing is generally well behaved, continuum fitting can, in principle, modify very broad spectral features; however, the use of a low-order spline and its consistent application to both target and template spectra helps ensure that residual distortions are treated consistently between target and template spectra, reducing their impact on the cross-correlation.

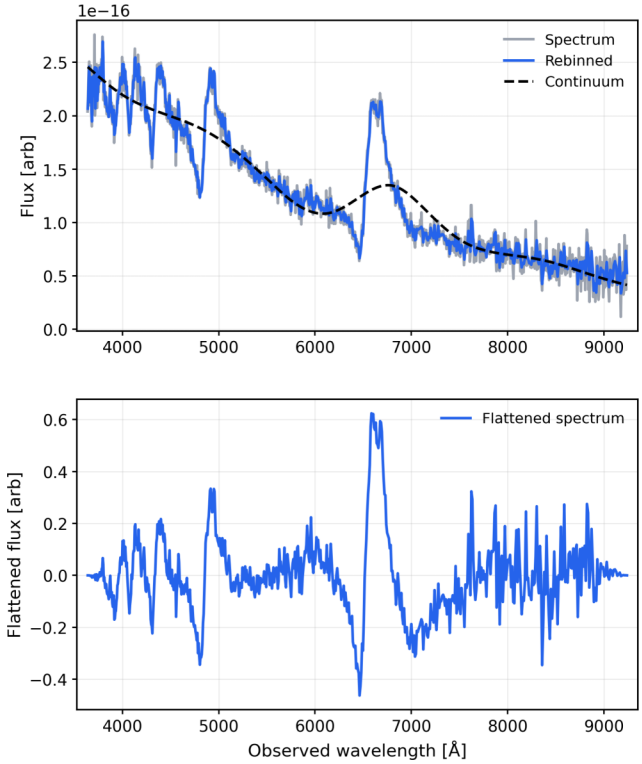

Figure 1. Pre-processing pipeline illustrated for SN 2018bif. *Top panel:* raw input spectrum (grey) and spectrum after logarithmic-wavelength rebinning (blue), with the fitted broad-band continuum overlaid (dashed black). *Bottom panel:* continuum-divided flattened spectrum ready for the cross-correlation stage.

2.2 Step 2: cross-correlation and similarity metrics

After pre-processing, the flattened target spectrum is compared with each rest-frame template using the Fourier cross-correlation technique introduced by J. Tonry & M. Davis (1979) and adopted by SNID (S. Blondin & J. L. Tonry 2007). Working on the common $\ln\lambda$ grid described in the previous section, redshift determination reduces to identifying the shift that maximizes the alignment of spectral features between the target and each template.

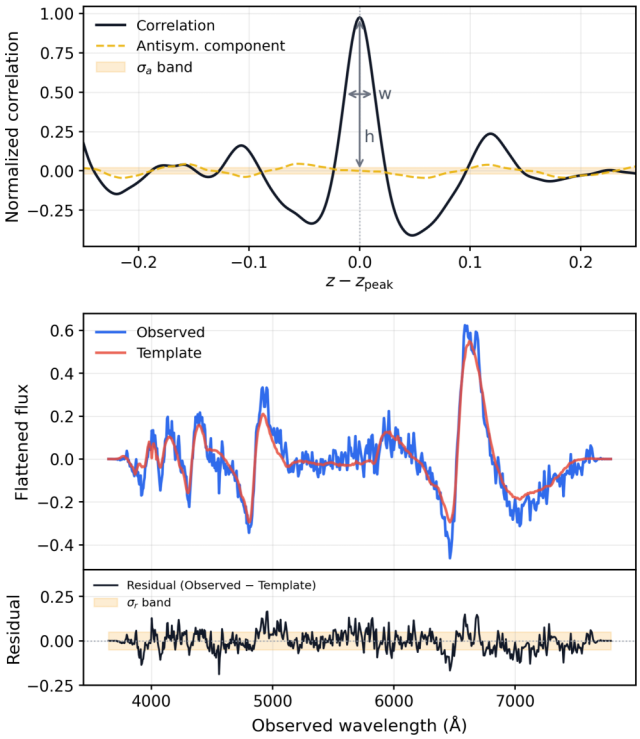

Figure 2. Illustration of cross-correlation components and match-quality diagnostics for SN 2018bif. *Top panel:* normalised cross-correlation function between the observed spectrum and the best-matching template, shown as a function of redshift offset relative to the peak ($z - z_{peak}$). The solid black curve shows the filtered correlation, while the dashed orange curve shows the antisymmetric component used in the classical Tonry–Davis R statistic. The shaded band indicates the rms of the antisymmetric component. The peak height, h and width, w characterize the strength of the alignment and the sharpness of the redshift constraint. *Middle panel:* Continuum-flattened, apodized observed spectrum (blue) and the best-matching template (red) at the peak redshift, illustrating the degree of spectral agreement that drives the correlation. *Bottom panel:* residual spectrum (observed minus template) at the peak redshift, with the shaded band indicating the standard deviation of the residuals. In SNID–SAGE, this residual scatter provides an explicit estimate of mismatch, complementing the peak width in the redshift uncertainty estimate and replacing reliance on the antisymmetric correlation component alone.

To suppress spurious correlations driven by residual continuum structure or high-frequency noise, the cross-correlation is computed with a smooth raised-cosine band-pass filter in Fourier space, following standard practice (F. J. Harris 1978; S. Blondin & J. L. Tonry 2007). The resulting cross-correlation function typically exhibits one or more peaks, each corresponding to a candidate redshift (Fig. 2, top panel). For a given template, we only take the highest peak as the preferred redshift solution, and its centre position is further refined via local quadratic interpolation to achieve sub-grid accuracy.

In the original SNID formulation, the significance of a cross-correlation peak is quantified using the Tonry–Davis R statistic, defined as the peak height divided by the rms of the antisymmetric component of the correlation function ($R = h\,/(\sqrt{2}\sigma_a)$, components shown in Fig. 2). This is combined with the fractional wavelength overlap between the target and the redshifted template to form the standard quality metric RLAP $= R \times$ lap (S. Blondin & J. L. Tonry 2007). As illustrated in the top panel of Fig. 2, the antisymmetric component is intended to provide a noise estimate against which the peak height is normalized. However, the antisymmetric component of the cross-correlation is not, in general, a reliable estimator of the effective noise in the match. As noted in several studies, it can be influenced by spectral structure, windowing effects, and filtering choices rather than noise alone (A. F. Heavens 1993; F. Bouchy, F. Pepe & D. Queloz 2001; S. Zucker 2003; S. Zamora & A. I. Díaz 2023). As a result, RLAP can overestimate the quality of matches in cases where the correlation peak is narrow but driven by a limited subset of features. Furthermore, cross-correlation is intrinsically insensitive to overall amplitude and offset mismatches, as it quantifies relative alignment but not agreement in flux scale.

In SNID–SAGE, we introduce two modifications intended to mitigate these limitations. First, rather than using the antisymmetric rms as a proxy for noise, we characterize peak sharpness and spectral mismatch explicitly. For each template, we define a local scale parameter

$$\sigma = w\,\sigma_{\rm r},$$

where w is the full width at half-maximum (FWHM) of the cross-correlation peak in redshift space and $\sigma_{\rm r}$ is the standard deviation of the residual spectrum (observed minus template) evaluated at the peak redshift (Fig. 2, bottom panel). This scale increases for broad or poorly matched solutions and moderates the ranking accordingly. Secondly, to account for amplitude and offset mismatches that cross-correlation alone cannot capture, we incorporate the Lin concordance correlation coefficient (L. I.-K. Lin 1989), CCC, which measures agreement in both correlation and flux scaling between the target and the redshifted template.

Template matches in SNID–SAGE are therefore ranked using a composite statistic

$$\mathrm{H}\sigma\mathrm{LAP\text{-}CCC} = \frac{h\ \mathrm{LAP\ CCC}}{\sqrt{\sigma}},$$

where h is the cross-correlation peak height, LAP the fractional wavelength overlap, CCC the concordance correlation coefficient, and σ the local sharpness mismatch scale defined above. The square root dependence on σ introduces a moderate penalty for broad or spectrally inconsistent solutions while preserving the primary influence of the alignment quality terms in the numerator. As with RLAP, this statistic is heuristic and intended for stable ranking rather than direct probabilistic interpretation.

For each template, the procedure therefore yields a best-fitting redshift, the corresponding match score HσLAP-CCC, and associated metadata (type, subtype, and phase). Templates with HσLAP-CCC ≤ 1.5 are discarded at this stage, so that only meaningful matches are propagated to the consolidation step.

2.3 Step 3: clustering and inference

In the original SNID framework, classification is based on the ranking of individual template matches, typically using the RLAP statistic, and the final type and redshift are inferred from the highest-scoring templates or from simple aggregation over the top-ranked subset (S. Blondin & J. L. Tonry 2007). While effective in many cases, this approach treats template matches

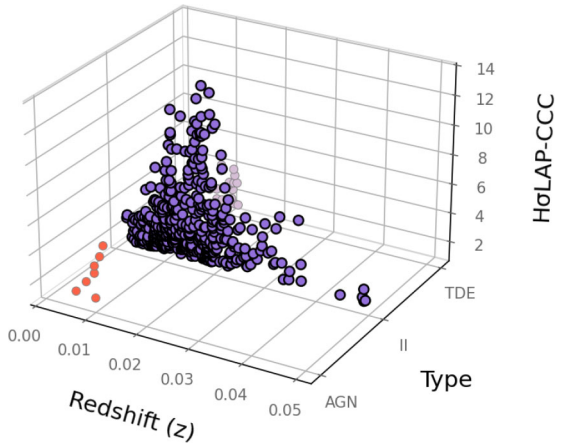

Figure 3. Example of redshift-space clustering and cluster selection in SNID–SAGE. Visualization of template matches for 2018bif shown as a function of redshift, template type, and $H\sigma$ LAP-CCC. Each point corresponds to a single template match; clusters form naturally as concentrations of matches at similar redshift within a given type. The highlighted cluster corresponds to the highest-ranked solution according to the cluster quality score Q.

largely independently and does not explicitly account for the presence of multiple, potentially distinct solutions in redshift space. SNID–SAGE instead combines information from multiple template matches in a structured way to derive the final classification.

For each template type (currently 15 categories: AGN, CV, GAP, Galaxy, II, Ia, Ib, Ibn, Ic, Icn, KN, LFBOT, SLSN, Star, and TDE), we collect the redshift estimates of the retained template matches and cluster them using a one-dimensional GMM (G. J. McLachlan, S. X. Lee, & S. I. Rathnayake 2019). This procedure yields distinct redshift clusters that represent alternative candidate solutions within a given type. The number of Gaussian components is selected using the Bayesian Information Criterion (G. Schwarz 1978), supplemented by an elbow criterion to avoid overfragmentation in sparsely populated regions of redshift space. To ensure physically meaningful solutions, clusters are required to correspond to contiguous structures in redshift space: Widely separated groups of matches are split into distinct clusters rather than merged into a single solution. An example of this clustering and subsequent cluster selection is shown in Fig. 3.

For each cluster, we define a quality score Q based on its highest-ranked member templates. Specifically, we compute the mean of the top five $H\sigma$ LAP-CCC scores within the cluster (or all members if fewer than five are present) and apply a linear penalty when fewer than five templates support the solution

$$Q = \left(\frac{n}{5}\right) \bar{q}_{\rm top}, \qquad n = \min(5, N_{\rm cl}),$$

where $\bar{q}_{\rm top}$ is the mean of the top n match $H\sigma$ LAP-CCC scores and $N_{\rm cl}$ is the cluster size. This definition favours solutions that are both strong and supported by multiple independent templates.

Across all clusters and types, the cluster with the highest Q is adopted as the primary solution, while lower-ranked clusters are retained as explicit alternatives. For reporting, Q is mapped to qualitative classes using fixed thresholds: *Very Low* for $Q < 2.5$, *Low* for $2.5 \le Q < 5$, *Medium* for $5 \le Q < 8$, and *High* for $Q \ge 8$.

We additionally compute a relative type confidence based on the fractional improvement in Q over the next-best competing type,

$$C_{\rm type} = \frac{Q_1 - Q_2}{Q_2} \times 100 \text{ per cent},$$

where Q_1 is the winning type's best-cluster score and Q_2 is the best-cluster score of the runner-up type. This is mapped to discrete confidence classes using fixed thresholds: *High* for $C_{\rm type} \ge 75$ per cent, *Medium* for $C_{\rm type} \ge 25$ per cent, *Low* for $C_{\rm type} \ge 5$ per cent, and *Very Low* otherwise. When a runner-up type is absent after the match-retention and clustering steps, no finite value of $C_{\rm type}$ is assigned. These cases are flagged separately as *No Comp*, indicating that no competing alternative survived the quality criteria.

Each redshift cluster typically contains template matches belonging to multiple subtypes of the same type (see Table 1 for the lists of subtypes). Within the highest-ranked cluster, subtype inference is performed by grouping member templates by subtype and applying the same top-n scoring procedure to each subgroup. The subtype with the highest resulting Q is selected as the preferred subtype, and a relative subtype confidence is computed from the fractional improvement over the next-best subtype within the same cluster.

With the preferred subtype established, we estimate its redshift and phase by combining information from all templates of that subtype within the winning cluster. For each contributing template i, we retain the redshift z_i, phase t_i, and match score q_i, where q_i is the template's $H\sigma$ LAP-CCC value (Section 2.2). Templates are weighted according to their relative match quality using exponential weights,

$$w_i = \exp\left(\frac{q_i}{\tau}\right),$$

which favours higher-quality matches while retaining contributions from the full subtype sample. In practice, we adopt $\tau = 1$, which provides stable behaviour across the template library. The parameter τ controls the concentration of the weights: When one match is clearly superior, it dominates the estimate, whereas comparable matches contribute collectively when scores are similar. This weighting is heuristic and designed for robust aggregation rather than formal probabilistic inference.

We then estimate redshift and phase as weighted means,

$$\hat{z} = \frac{\sum_i w_i z_i}{\sum_i w_i}, \qquad \hat{t} = \frac{\sum_i w_i t_i}{\sum_i w_i},$$

and report internal-scatter measures defined as the unbiased weighted dispersion,

$$\sigma_{\hat{x}} = \left[\frac{\sum_i w_i (x_i - \hat{x})^2}{\sum_i w_i} \times \frac{N_{\rm eff}}{N_{\rm eff} - 1}\right]^{1/2},$$

where $x = z$ (with $x_i = z_i$) gives $\sigma_{\hat{z}}$ and $x = t$ (with $x_i = t_i$) gives $\sigma_{\hat{t}}$, and where

$$N_{\rm eff} = \frac{\left(\sum_i w_i\right)^2}{\sum_i w_i^2}.$$

This correction intentionally penalizes solutions that are dominated by only one or a few templates: When a single match carries most of the weight, $N_{\rm eff} \to 1$ and the factor $N_{\rm eff}/(N_{\rm eff} - 1)$ increases the reported scatter, reflecting the fact that a visually strong best match is less reliable when it is not supported by a broader set of comparably good templates. In Section 4.1, we validate these redshift and phase estimators using leave-one-out

Table 1. Composition of the 698-template default optical library, grouped by transient type. Subtype counts are shown in parentheses. The total number of individual spectra (epochs) per type is given in the final column.

Type	Subtypes (*N*)	Templates	Spectra
Ia	Ia-norm (160), Ia-91T (36), Ia-91bg (33), Ia-02cx (12), Ia-csm (10) Ia-02es (7), Ia-03fg (7), Ia-99aa (2), Ia-Ca-rich (2), Ia-pec (2)	271	2864
II	IIP (49), IIb (31), IIn (16), II-flash (5), IIL (4), II-87A (1), II-pec (1), IIn-pec (1)	108	1216
Ib	Ib-norm (44), Ib-pec (6), Ib-Ca-rich (4)	54	509
Ic	Ic-norm (36), Ic-broad (30), Ic-05ek (2), Ic-Ca-rich (1)	69	632
Ibn	Ibn (12)	12	73
Icn	Icn (5)	5	29
SLSN	SLSN-I (48), SLSN-II (13)	61	297
TDE	TDE-H-He (11), TDE-He (5), TDE-H (3), TDE-Ftless (1)	20	105
AGN	AGN-Type1 (5), AGN-Type2 (4), AGN-Type1-BAL (1), AGN-Type1.5 (1)	11	15
Galaxy	Gal-SB (12), Gal-Sc (5), Gal-E (4), Gal-S0 (2), Gal-Sa (2), Gal-Sb (2)	27	27
CV	DN (4), Nova (4), NL (2), AM_CVn (1), IP (1), Polar (1)	13	68
GAP	LBV (5), ILRT (5), LRN (4)	14	72
LFBOT	20xnd (2), 18cow (1)	3	21
KN	17gfo (1)	1	10
Star	Varstar (11), Mdwarf (9), Symbiotic (3), WR-WC (3), WR-WN (3)	29	61
Total		698	5999

cross-validation, including residual and uncertainty calibration analyses.

At the end of this stage, the core SNID–SAGE analysis is complete. The pipeline returns a primary classification consisting of a best-fitting type and subtype, together with redshift and phase estimates and their associated internal-scatter measures. These results are accompanied by an absolute quality assessment for the winning solution and relative confidence measures that quantify how decisively the preferred type and subtype are favoured over competing alternatives. Lower-ranked clusters and subtype solutions are retained as explicit alternatives, enabling easy inspection and downstream decision-making by the user.

2.4 Additional GUI features

In addition to the core classification pipeline described above, the SNID–SAGE graphical interface provides a set of auxiliary analysis tools designed to support inspection, refinement, and interpretation of spectroscopic classifications. These tools operate on the results of the deterministic pipeline and do not modify the underlying classification logic.

2.4.1 Host redshift constraints

When a host-galaxy redshift is available from independent observations (e.g. narrow emission lines), it can be used to constrain the analysis. In the graphical interface, the host redshift may be either fixed to a single value, in which case the cross-correlation is evaluated only at that redshift, or specified as an interval, in which case the search is restricted to that redshift range (default ± 0.0005 about the chosen value). The interface supports both manual identification of common spectral features (e.g. Hα, Hβ, [O III]), and automatic estimation via galaxy template matching.

The scripted workflow provides the same functionality via the `-forced-redshift` option: If a fixed value is supplied, the templates are matched only at that redshift; if a bounded interval is supplied, the cross-correlation search is limited to that interval, with the same default width of ± 0.0005 when no custom range is specified. In both cases, constraining the redshift in this way reduces the search space and can improve classification robustness and phase estimation when the host redshift is well determined.

2.4.2 Line identification and velocity measurements

Given a consolidated redshift estimate, the SNID–SAGE Graphical User Interface (GUI) provides an interactive line-identification tool to aid inspection of both supernova and host-galaxy features. Users can overlay curated rest-frame line lists (selectable by transient type, approximate phase, and element group) and interactively toggle individual lines to visualize their expected observed wavelengths.

For any selected line, the GUI supports a lightweight measurement workflow: The user marks points around the feature and the tool returns basic line properties such as the peak wavelength and an FWHM estimate. When a rest wavelength is available, the measured width is converted into a characteristic velocity scale. These measurements are intended to support rapid, transparent interpretation of spectra (e.g. checking consistency with a proposed type/phase or identifying host emission lines), and do not feed back into the automated classification.

2.4.3 AI-powered analysis summaries

The GUI includes an optional *AI Assistant* button that generates a short, structured natural-language summary of the current SNID–SAGE results. When enabled, the tool formats the classification output (best type/subtype, redshift and phase estimates, match quality/confidence, and any user-selected line markers) together with optional observation metadata entered by the user, and submits this context to an externally configured large language model (LLM) service (via OpenRouter). The returned text is intended for rapid reporting and human-readable logging. An example of such a generated summary is provided in Appendix B.

The AI Assistant is strictly descriptive: It does not modify preprocessing, matching, clustering, or any numerical outputs, and its summary should not be interpreted as an independent classifier. Because this feature relies on an externally configured third-party service, it is disabled unless explicitly set up by the user, and it should be used with appropriate caution for proprietary or unpublished data.

3 TEMPLATE LIBRARY

The performance of SNID–SAGE is fundamentally determined by the quality, diversity, and accuracy of its spectral template library. The current default optical library comprises 698 rest-frame templates (5999 individual spectra) rebinned on to a common logarithmic wavelength grid (2500–10 000 Å; 1024 bins in ln λ).

The majority of templates originate from the public SuperSNID library (D. Magill et al. 2025), supplemented by stripped-envelope templates (Ib, Ic, Ibn, Icn) created by the Modjaz METAL group (Y. Liu & M. Modjaz 2014; Y.-Q. Liu et al. 2016; M. Modjaz et al. 2016; Y.-Q. Liu, M. Modjaz & F. B. Bianco 2017; M. Williamson, M. Modjaz & F. B. Bianco 2019; M. Williamson et al. 2023; N. Yesmin et al. 2025), together with additional templates curated and validated as part of this work. Table 1 summarizes the template types and subtypes represented in the default optical library.

Although SNID–SAGE is not a trained model in the supervised-learning sense, the template library still defines the applicability domain of the method. Classes with dense and diverse template coverage are expected to give more stable classifications, while rare or sparsely represented populations should be interpreted as the closest available spectral analogue within the current library rather than as definitive identifications. This also means that emerging transient populations can be incorporated by adding well-characterized templates spanning the relevant phase and spectral diversity, without retraining the pipeline.

Because template-based classification inherits any biases or inaccuracies present in the library, template quality control is essential to the reliability of the pipeline. We actively maintain and update the library and encourage community feedback. Researchers who identify errors in template metadata (e.g. redshift, type/subtype, or phase) or who wish to contribute additional well-characterized spectra are invited to contact the authors; proposed additions are reviewed and, when validated, incorporated into future releases of the default library. The associated Template Manager workflow is described in Appendix A.

In addition to the default optical library, an extended optical + near-infrared template library is currently under development and is already available in a preview form for testing. This extended library incorporates near-infrared spectral coverage, enabling cross-correlation over a broader wavelength baseline and therefore extending the effective redshift range over which reliable matches can be obtained.

4 VALIDATION AND PERFORMANCE ANALYSIS

We evaluate SNID–SAGE through two complementary validation approaches: (i) leave-one-out cross-validation on the template library to quantify type-classification accuracy and redshift and phase recovery; and (ii) large-scale application to the WISeREP data base (O. Yaron & A. Gal-Yam 2012), with comparison to independently measured host-galaxy redshifts, to assess redshift accuracy, precision, and robustness under realistic survey conditions.

Table 2. Leave-one-out validation results by transient type. Reported quantities are top-k type recovery (per cent), median absolute redshift error $|\Delta z|$, and median absolute phase error $|\Delta t|$ (d). Classes with too few templates for statistically meaningful leave-one-out evaluation (e.g. KN and LFBOT) are omitted.

Type	Top-1 (per cent)	Top-2 (per cent)	Top-3 (per cent)	$\lvert\Delta z\rvert$	$\lvert\Delta t\rvert$
Ia	99.3	99.8	99.9	0.001 46	1.9
II	88.6	97.9	99.3	0.001 58	10.9
Ibn	82.2	87.7	90.4	0.001 46	8.9
Ic	80.1	95.6	98.0	0.003 49	6.6
Ib	78.5	94.1	97.6	0.003 52	6.2
GAP	77.1	90.0	90.0	0.000 44	17.6
TDE	65.3	95.8	96.8	0.001 08	41.0
SLSN	56.1	74.8	81.7	0.003 36	11.5
Icn	37.0	48.1	48.1	0.001 11	3.4
AGN	100.0	100.0	100.0	0.000 08	–
Star	91.7	95.0	95.0	0.000 41	–
Galaxy	90.5	100.0	100.0	0.000 07	–
CV	86.5	100.0	100.0	0.000 22	–

4.1 Leave-one-out cross-validation

To quantify type-classification accuracy and the reliability of redshift and phase estimation within the template library, we perform leave-one-out cross-validation on the full set of templates. In this experiment, all spectra belonging to a template are removed in turn and analysed against the remaining library using the complete SNID–SAGE pipeline. This procedure tests whether a spectrum can be correctly classified and assigned consistent redshift and phase estimates when its exact template is excluded, providing a good assessment of template coverage and the robustness of the clustering-based consolidation. Table 2 summarizes per-type recovery for top-1, top-2, and top-3 predicted types together with the median absolute redshift error for correctly recovered spectra. Phase errors are reported only for transient classes with a well-defined maximum-light reference; for non-transient classes (e.g. galaxies, AGN, stars, CVs), phase is not defined and is therefore omitted. For example, the 'top-3' recovery column gives the fraction of spectra for which the correct type appears among the three highest-ranked final solutions returned by SNID–SAGE, with ranking determined by the cluster-level quality score Q. Here, we consider type recovery only; redshift and phase consistency are assessed separately in the analyses that follow.

Overall, the per-type breakdown in Table 2 shows that SNID–SAGE recovers the correct type reliably across the majority of classes. While the top-1 recovery varies depending on spectral diversity and template coverage, the top-2 and top-3 recovery rates are consistently much higher, indicating that the true type is very frequently contained within the first few candidate solutions returned by the pipeline. The median redshift errors remain small for the major classes, while the phase errors are generally modest but more sensitive to template coverage and intrinsic spectral variability.

Type Ia spectra are recovered with near-perfect top-1 accuracy (99.3 per cent), reflecting both their distinctive spectral morphology and the extensive phase coverage of the Ia template set. Type

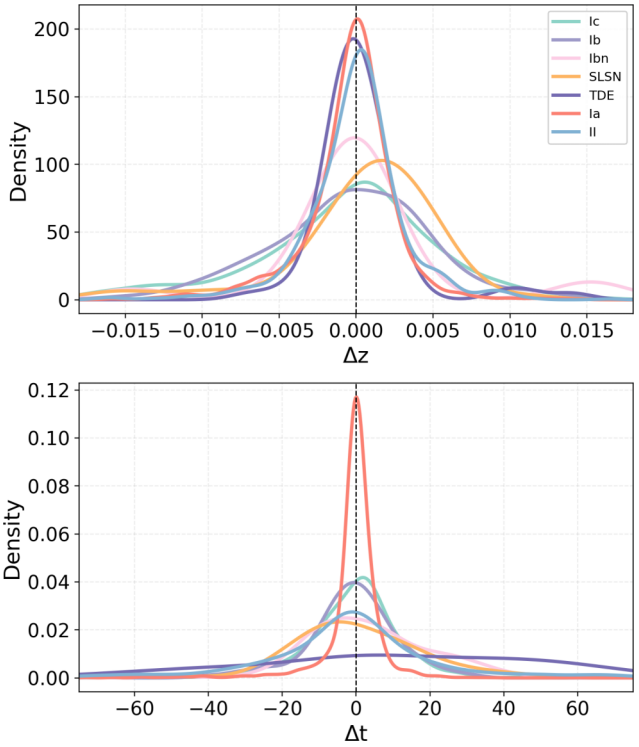

Figure 4. Leave-one-out residual distributions by transient type. Top: distribution of redshift residuals $\Delta z = z_{\rm pred} - z_{\rm true}$. Bottom: distribution of phase residuals $\Delta t = t_{\rm pred} - t_{\rm true}$. Vertical dashed lines mark zero residual. Redshift residuals are tightly centred for all major classes, while phase residuals show a stronger dependence on transient type, with broader distributions and extended tails for heterogeneous classes such as SLSNe and TDEs.

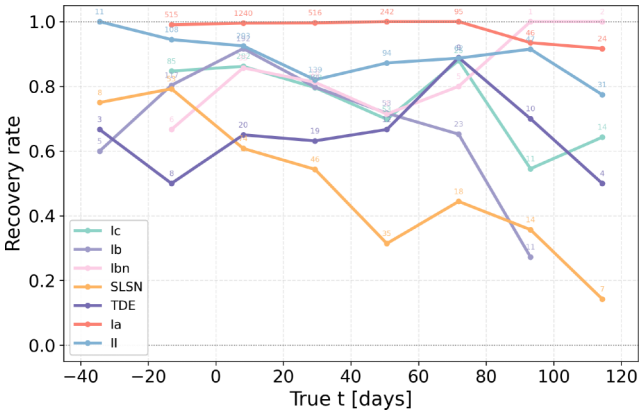

Figure 5. Leave-one-out recovery rate as a function of phase for selected transient classes. Classification performance is highest near peak light and degrades at late phases, with stronger phase dependence and variability for heterogeneous and sparsely sampled classes such as SLSNe and TDEs.

II supernovae also exhibit strong recovery performance (top-1 $\sim$ 89 per cent), while stripped-envelope subclasses (Ib and Ic) show moderately lower top-1 rates (78–80 per cent) but substantially improved top-2 and top-3 recall, consistent with their greater spectral diversity and overlap between neighbouring subtypes at certain phases. Classes such as SLSNe, TDEs, gap transients, and Icn exhibit reduced top-1 recovery, which reflects both intrinsic

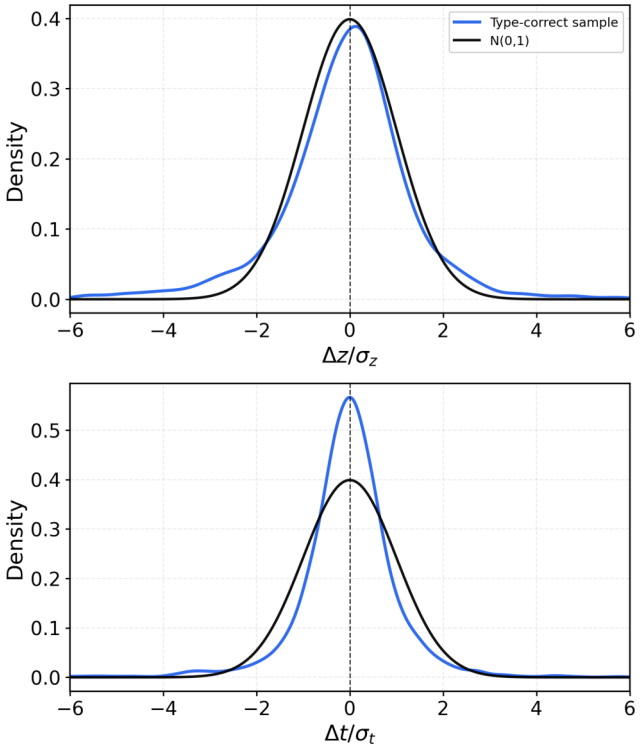

Figure 6. Calibration of internal uncertainty estimates from the leave-one-out experiment. Top: distribution of normalized redshift residuals $(z_{\rm pred} - z_{\rm true})/\sigma_z$ for spectra with correctly recovered type. Bottom: distribution of normalized phase residuals $(t_{\rm pred} - t_{\rm true})/\sigma_{\rm t}$. In both panels, the blue curve shows the kernel-density estimate of the empirical distribution, while the black curve shows the unit normal distribution $\mathcal{N}(0, 1)$ expected for well-calibrated uncertainties. The vertical dashed line marks zero residual.

heterogeneity and comparatively smaller template coverage; for many of these classes, the correct type none the less appears frequently within the top-2 or top-3 candidates. Very small classes (e.g. kilonova, LFBOT) are omitted from this table when leave-one-out evaluation is statistically unstable (removing a single template can eliminate the only representative of that class from the library), preventing meaningful recovery assessment.

In addition to overall recovery rates, we examine the distribution of redshift and phase residuals obtained in the leave-one-out experiment. Fig. 4 shows kernel-density estimates of Δz and Δt (predicted minus true values) for a representative subset of transient classes. The redshift residuals are tightly centred around zero for all major supernova classes, with Type Ia exhibiting the narrowest distribution. Core-collapse classes show broader wings, reflecting greater spectral diversity and increased degeneracy between neighbouring types. Phase residuals exhibit a stronger dependence on transient class, with Type Ia again showing the narrowest distribution, while classes such as SLSNe and TDEs display substantially broader and more extended residuals. This reflects both the more heterogeneous spectral evolution of these populations and the more limited template coverage. In particular, the presence of extended tails in the phase residuals indicates that phase estimates for these classes can be significantly less constrained.

To explore how classification performance depends on evolutionary phase, we compute the recovery rate as a function of phase for a representative subset of transient types. Fig. 5 shows

the fraction of correctly recovered spectra in phase bins. For Type Ia supernovae, recovery remains near unity across all sampled phases, reflecting both strong spectral homogeneity and dense template coverage. Type II supernovae also exhibit consistently high recovery, with only modest degradation at later phases as spectral features evolve. Stripped-envelope classes (Ib and Ic) show a stronger dependence on phase, with recovery declining at late times where line blending and signal-to-noise limitations become more pronounced. SLSNe and TDEs display both lower overall recovery and significantly larger variability with phase, including pronounced degradation at intermediate and late phases. Overall, these trends indicate that classification robustness is strongest near peak light and becomes increasingly sensitive to intrinsic diversity and template coverage at extreme early or late phases.

To assess whether the internal uncertainty estimates produced by SNID–SAGE are well calibrated, we examine the distribution of normalized residuals obtained in the leave-one-out experiment. For each correctly classified spectrum we compute $(z_{\rm pred} - z_{\rm true})/\sigma_z$ and $(t_{\rm pred} - t_{\rm true})/\sigma_{\rm t}$, where σ_z and $\sigma_{\rm t}$ are the internal uncertainty estimates reported by the pipeline. If the uncertainties are well calibrated and the residuals are approximately Gaussian, these normalized residuals are expected to follow a standard normal distribution $\mathcal{N}(0, 1)$.

Fig. 6 compares the empirical distributions of the normalized residuals with the unit normal expectation. The redshift residuals closely follow the $\mathcal{N}(0, 1)$ distribution, indicating that the redshift uncertainties derived from the cluster-weighted solution provide a realistic estimate of the true scatter. In contrast, the phase residuals are more narrowly distributed than a unit normal, implying that the reported phase uncertainties are somewhat conservative. In other words, the pipeline tends to slightly overestimate the uncertainty on phase estimates, such that the true phase errors are typically smaller than the nominal uncertainties reported by SNID–SAGE.

To further assess the reliability of the classification outputs, we examine how the internally defined match-quality and type-confidence categories relate to actual classification performance in the leave-one-out experiment. Here, the type confidence quantifies the relative strength of the preferred solution compared to the next-best competing type, as introduced in Section 2.3. Fig. 7 shows the fraction of correctly recovered types as a function of these two quantities.

As expected, the classification reliability increases strongly when both the match-quality score and the type-confidence class are high. The best performance is obtained in the upper-right region of the plot, where the correct type is recovered for the overwhelming majority of spectra. By contrast, the *Very Low* quality bins remain comparatively unreliable even when the confidence class is moderate or high, indicating that the quality score contains important information beyond the type-confidence flag alone. Taken together, these two quantities provide a compact and interpretable summary of classification reliability, enabling automated workflows to identify high-confidence results while flagging uncertain cases for further inspection.

Overall, the leave-one-out analysis demonstrates that SNID–SAGE provides robust and internally consistent estimates of type, redshift, and phase within the template library. Performance differences across transient classes are largely driven by template coverage: Well-represented and spectrally homogeneous classes (e.g. Type Ia) show the most stable behaviour, while rarer and more heterogeneous classes exhibit reduced performance. This suggests that further improvements can be achieved through the expansion and continued curation of the template library.

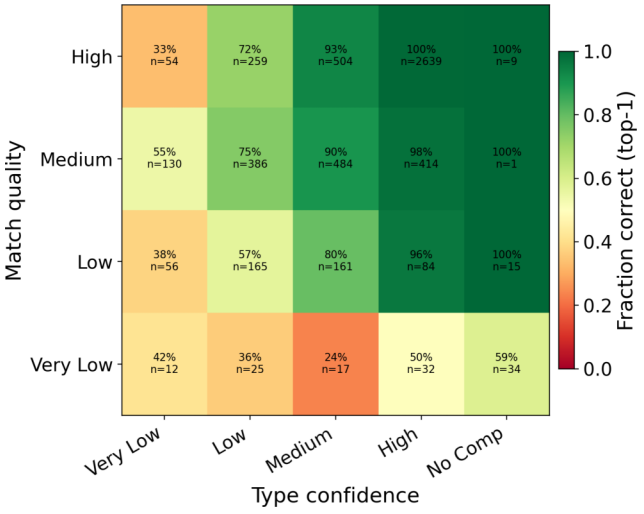

Figure 7. Type-recovery rate as a function of match quality and type confidence in the leave-one-out experiment. Each cell shows the fraction of spectra for which the correct type appears as the top-ranked final solution, together with the number of spectra contributing to that bin. Recovery improves strongly with increasing match quality, and the highest success rates occur when both the match quality and type-confidence class are high, indicating that the confidence metric provides a meaningful measure of classification reliability. The 'No Comp' category denotes cases where no competing secondary solution is identified. This occurs when a single solution is clearly preferred, or when only one marginally plausible match is present and all other candidates are rejected as too low quality.

4.2 WISeREP data base analysis

To evaluate performance under realistic observational conditions, we analyse a large sample of astrophysical transient spectra from the WISeREP data base (O. Yaron & A. Gal-Yam 2012) with spectral coverage over the 4000–7000 Å interval, irrespective of spectral type. This sample comprises 46 763 spectra spanning a wide range of object types, phases, and data quality, the vast majority of which are user-uploaded spectra. This data set serves two complementary purposes. First, it provides a large-scale public reference catalogue of SNID–SAGE classifications applied uniformly to a heterogeneous archival data set. Secondly, it enables quantitative assessment of redshift recovery through comparison with independently determined host-galaxy redshifts where available.

The WISeREP data base aggregates classifications from multiple surveys, instruments, and authors, and does not enforce a uniform taxonomy across all entries, therefore we do not perform a direct comparison between the predicted type/subtype labels and the WISeREP classifications. Instead, we provide the full SNID–SAGE set of classifications and derived quantities in a public catalogue[1] and leave detailed taxonomy comparisons to targeted user studies. The overall composition of inferred types and match-quality classes is summarized in Table 3, providing a baseline view of the full WISeREP sample as processed by SNID–SAGE.

[1] https://fiorenst.github.io/SNID-SAGE/table/

Table 3. Breakdown of main SNID–SAGE types and match-quality classes for the 45 551 spectra that yielded valid SNID–SAGE redshift solutions from the WISeREP sample.

Main type	Very low	Low	Medium	High
AGN	17	42	77	92
GAP	47	63	107	210
Galaxy	83	265	351	388
II	1071	2432	2992	4747
Ia	918	4802	8864	10 782
Ib	137	306	472	596
Ibn	64	75	123	154
Ic	303	848	836	528
Icn	23	5	7	19
KN	13	2	0	3
LFBOT	18	4	5	14
SLSN	203	274	229	259
Star	87	167	233	290
TDE	141	170	230	363

Redshift, by contrast, provides a well-defined quantity that can be validated against independent host-galaxy measurements, and we therefore focus on redshift recovery. Because transients are frequently offset from their hosts and may lie near multiple projected galaxies, secure host association is required before comparing inferred and catalogue redshifts. To perform this association, we adopt host galaxies from the REGALADE catalogue (H. Tranin et al. 2026), a unified compilation of multiple galaxy surveys with uniformly derived redshift and galaxy-property measurements. Host association is carried out using the directional light radius (DLR) framework (R. R. Gupta et al. 2016), in which the projected transient–galaxy separation is normalized by the galaxy's elliptical light profile. Candidate hosts are ranked by increasing DLR, with the minimum-DLR galaxy adopted as the reference host, and associations with DLR > 2.5 are rejected.

For this analysis, we restrict the sample to transients whose associated hosts have reliable spectroscopic redshifts. We further require consistency between the REGALADE and WISeREP redshift measurements, excluding objects with $|z_{\rm REGALADE} - z_{\rm WISeREP}| > 0.025$. This threshold is intentionally broader than the intrinsic redshift precision of the spectral pipeline and is designed to remove catastrophic host mismatches rather than small systematic offsets. Out of the initial 46 763 spectra, 45 551 yield a valid classification and redshift solution with SNID–SAGE. Of these, 25 931 satisfy the host-association and redshift-consistency criteria. We additionally exclude WISeREP entries flagged as CAUTION and a small set of telescope/instrument combinations for which most spectra were uploaded already de-redshifted by users and therefore do not provide a meaningful test of redshift recovery. After all cuts, the final validation sample comprises 24 223 spectra.

Before comparing SNID–SAGE redshifts with host-galaxy measurements, we first evaluate the consistency between host redshifts reported in WISeREP and those extracted from REGALADE. Fig. 8 shows the direct comparison between WISeREP-reported redshifts and the corresponding spectroscopic host redshifts from REGALADE for our sample. Although the overall agreement is good, a measurable intrinsic scatter is present. This is not unexpected: WISeREP redshifts are aggregated from multiple surveys and authors and are frequently user-reported, resulting in heterogeneous precision and occasional inconsistencies.

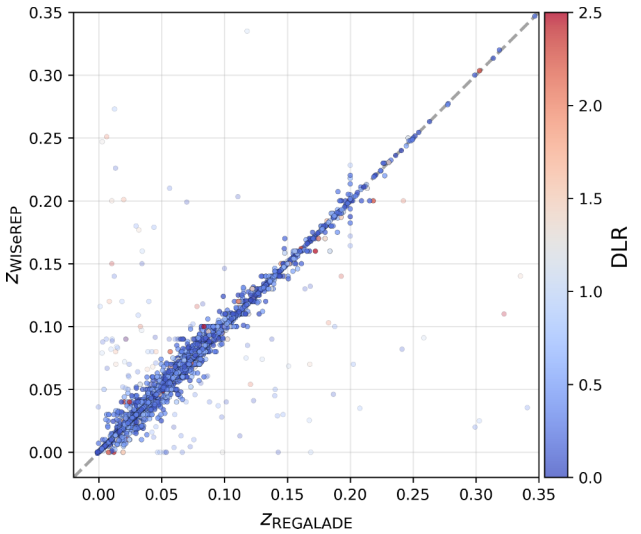

Figure 8. Comparison between WISeREP and REGALADE spectroscopic host-galaxy redshifts. Points are coloured by the host-association DLR. The dashed line shows the one-to-one relation. A small fraction of outliers reflects inconsistencies between catalogue redshifts and motivates the $|\Delta z_{\rm host}| < 0.025$ consistency cut used to define the final validation sample.

Table 4. Breakdown of main SNID–SAGE types and match-quality restricted to the validation sample after all cuts (REGALADE matching, DLR < 2.5, $|\Delta z_{\rm host}| \leq 0.025$, excluding WISeREP CAUTION spectra and specified problematic instruments). The total is 24 223 spectra, as described in Section 4.2.

Main type	Very Low	Low	Medium	High
AGN	10	25	43	55
GAP	36	47	53	120
Galaxy	23	143	203	278
II	466	1222	1476	2581
Ia	392	2239	4636	6457
Ib	63	199	339	408
Ibn	26	41	46	94
Ic	127	474	529	398
Icn	11	5	6	18
KN	5	0	0	0
LFBOT	5	4	4	13
SLSN	69	81	77	96
Star	4	25	52	11
TDE	53	83	148	204

The level of scatter between REGALADE and WISeREP is comparable to the residual dispersion measured later in this section for SNID–SAGE redshift recovery (see Fig. 9), indicating that part of the apparent redshift scatter in large-scale comparisons originates from catalogue-level differences rather than from the spectral inference itself. Having defined the final validation sample (24 223 spectra), we examine its composition in terms of inferred transient type and match quality. The resulting distribution is shown in Table 4.

Relative to the raw WISeREP sample, the validation subset is mildly enriched in higher-quality matches, particularly for Types Ia and II, reflecting the requirement of secure host associations and consistent catalogue redshifts. Nevertheless, the retained sample spans a broad range of transient classes,

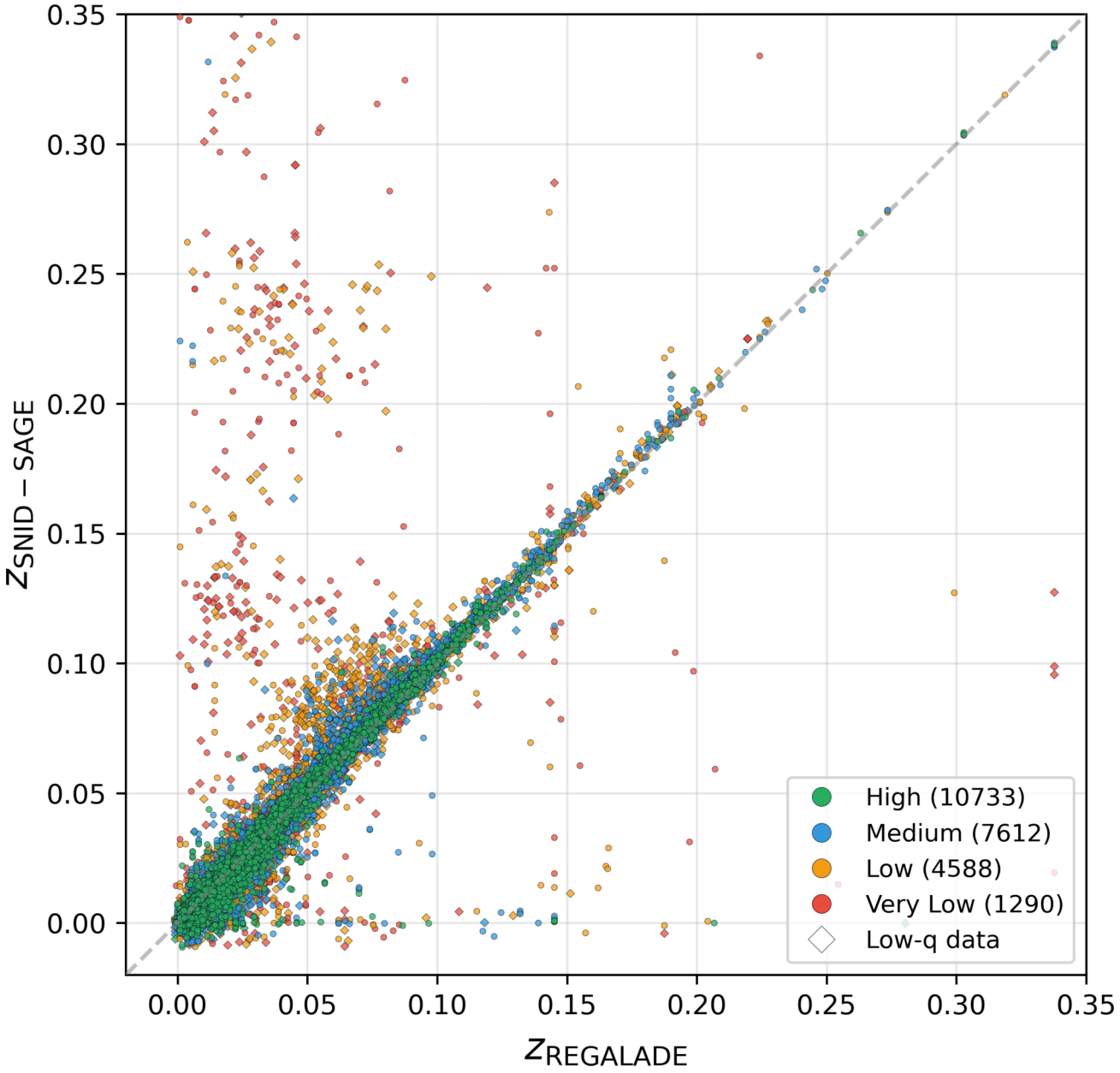

Figure 9. Comparison between SNID–SAGE redshifts and REGALADE host-galaxy reference redshifts for the full WISeREP validation sample. Points are colour-coded by SNID–SAGE match-quality class, and spectra from low-quality data instruments are shown as diamonds. The dashed line indicates the one-to-one relation. A small subset of spectra classified with *High* match quality but with predicted redshifts near zero likely corresponds to spectra that have been pre-corrected to the rest frame prior to upload.

Table 5. Catalogue-level redshift agreement between SNID–SAGE redshifts and REGALADE host-galaxy redshifts for the WISeREP validation sample of 24 223 spectra. We report $\Delta z_{\mathrm{SAGE}} \equiv z_{\mathrm{SNID-SAGE}} - z_{\mathrm{REGALADE}}$, giving the median and standard deviation within each SNID–SAGE match-quality class.

Match quality	Median Δz_{SAGE}	Std. Δz_{SAGE}
Very Low	0.001 613	0.130 564
Low	0.001 490	0.056 314
Medium	0.001 175	0.011 160
High	0.000 563	0.005 946

including stripped-envelope supernovae (Ib/Ic/Ibn), SLSN, TDEs, AGN, and other rarer populations. A substantial fraction of spectra in the major supernova categories achieve Medium or High match quality, indicating that the template library and clustering framework provide robust solutions across the dominant transient populations represented in the current library, while rarer events remain more sensitive to template coverage.

Fig. 9 presents the core redshift validation result, comparing SNID–SAGE redshifts to REGALADE host-galaxy reference redshifts for the full validation sample of 24 223 spectra. We also highlight a subset of spectra obtained with instruments that fre-

quently yield lower-quality matches. These telescope/instrument combinations are defined empirically as those for which more than 40 per cent of spectra are assigned *Very Low* or *Low* match quality by SNID–SAGE. This subset is dominated by lower-resolution classification instruments, in particular Palomar P60/SEDM, which was designed for rapid, coarse classification and plays an important role in transient filtering despite its limited spectral resolution. The Liverpool Telescope SPRAT spectrograph and the UH88/SNIFS combination also fall into this category. Spectra from these instruments contribute disproportionately to the *Very Low* and *Low* match-quality regimes and account for a significant fraction of the outliers seen in Fig. 9.

The majority of spectra cluster tightly around the one-to-one relation, with dispersion strongly correlated with match quality. High-quality matches exhibit minimal scatter and negligible systematic bias across the sampled redshift range. The Medium class remains well behaved, with only modest broadening, while the Low and Very Low classes show increasing dispersion and a higher incidence of catastrophic outliers. Spectra obtained with the instruments that we identify as producing lower quality data contribute disproportionately to the largest deviations. In particular, P60/SEDM displays a modest positive redshift bias that increases toward $z \simeq 0.06$–0.08, consistent with operation near the practical sensitivity limit where spectral features are less constraining and cross-correlation solutions become more susceptible to template degeneracies.

Table 5 quantifies the redshift residuals relative to REGALADE host-galaxy redshifts as a function of match quality. The median residual remains of order 10^{-3} across all match-quality classes, while the scatter decreases monotonically with increasing match quality. This behaviour confirms that the SNID–SAGE quality metric provides an effective empirical proxy for redshift reliability. Appendix C shows the same residuals as a function of host-galaxy redshift, separated by match-quality class, illustrating that the dominant trend is the reduction in scatter and catastrophic outliers for Medium and High quality solutions.

Overall, the WISeREP validation demonstrates that, for spectra with reliable host associations and at least *Medium* match quality, SNID–SAGE delivers robust and nearly unbiased redshift estimates across the sampled redshift range. At *Very Low* match quality, catastrophic outliers are possible – typically arising from extremely low-information spectra (weak or absent features, strong contamination, or limited wavelength coverage), as well as from objects that are rare or sparsely represented in the template library. The *Low* class remains usable in aggregate but carries a substantially higher outlier rate than *Medium* and *High*.

The comparison between WISeREP and REGALADE host-galaxy redshifts (Fig. 8) shows that a non-negligible fraction of the observed residual scatter arises from catalogue-level inconsistencies, implying that part of the dispersion in SNID–SAGE redshift recovery is driven by external reference uncertainties rather than limitations of the spectral inference itself.

We note that the WISeREP validation sample is not fully independent of the template library, because a subset of spectra and/or objects represented in public archives may also contribute to the construction of the default template set. The leave-one-out analysis on the template library therefore provides the primary controlled validation of classification and phase performance, while the WISeREP analysis should be interpreted as a large-scale deployment and redshift-recovery test on a heterogeneous archival sample rather than as a strictly independent classification benchmark.

5 CONCLUSIONS

We have presented the Python-based SNID–SAGE, a framework for supernova spectral classification that builds on the established SNID cross-correlation methodology with practical and methodological extensions for survey-scale use. The pipeline combines reproducible pre-processing, an optimized cross-correlation engine, a new match-quality metric, and redshift-space clustering to consolidate multiple template matches into stable type and subtype assignments.

Validation through leave-one-out testing of the template library yields high overall type-recovery performance for the main supernova types, with near-perfect recovery for Type Ia supernovae and stable redshift inference across the major classes represented in the library. Large-scale application to ∼46 000 WISeREP spectra, including a conservative validation subset of 24 223 spectra with secure host associations, demonstrates robust and nearly unbiased redshift recovery over the sampled redshift range.

Performance remains dependent on template coverage and spectral diversity. Recovery rates are reduced for intrinsically heterogeneous or sparsely represented classes, and very low-information spectra can produce catastrophic outliers at *Very Low* match quality. Continued expansion and curation of the template library will therefore be essential, particularly for rare or emerging transient populations. In that sense, SNID–SAGE can be naturally extended through the addition of validated templates spanning new classes, phases, and spectral diversity, allowing the framework to improve as the available spectroscopic coverage of rare and emerging transient populations grows.

SNID–SAGE is publicly available with full documentation and installation instructions, supporting both interactive analysis and scalable deployment in spectroscopic survey workflows.

ACKNOWLEDGEMENTS

FS acknowledges support from the Royal Society Newton International Fellowship NIFR1241769. SJS acknowledges funding from STFC Grants ST/Y001605/1, ST/X006506/1, ST/T000198/1, a Royal Society Research Professorship, and the Hintze Charitable Foundation. We thank Stephane Blondin for helpful discussions, and Hugo Tranin for assistance with the REGALADE catalogue.

DATA AVAILABILITY

The SNID–SAGE source code is publicly available at https://github.com/FiorenSt/SNID-SAGE, together with installation instructions and full user documentation.

The WISeREP classification catalogue produced in this work, including inferred types, redshifts, phases, match-quality metrics, and confidence estimates, is available in interactive form at https://fiorenst.github.io/SNID-SAGE/table/

The external data sets analysed in this study (WISeREP spectra and REGALADE host-galaxy redshifts) are publicly accessible through their respective archives, subject to their data access policies.

REFERENCES

Appenzeller I. et al., 1998, The Messenger, 94, 1
Bellm E. C., Kulkarni S. R., Graham M. J., Dekany R., Smith R. M., Riddle e. a., 2019, PASP, 131, 018002

Blagorodnova N. et al., 2018, PASP, 130, 035003
Blondin S., Tonry J. L., 2007, ApJ, 666, 1024
Bouchy F., Pepe F., Queloz D., 2001, A&A, 374, 733
Brown T. M. et al., 2013, PASP, 125, 1031
Cepa J., 1998, Ap&SS, 263, 369
Djupvik A. A., Andersen J., 2010, in Diego J. M., Goicoechea L. J., González-Serrano J. I., Gorgas J. eds, Astrophysics and Space Science Proceedings Vol. 14, Highlights of Spanish Astrophysics V. Springer-Verlag, Berlin, p. 211
Flewelling H. A. et al., 2020, ApJS, 251 7
Frohmaier C. et al., 2025, ApJ, 992, 158
Groot P. J., Bloemen S., Vreeswijk P. M., van Roestel J. C. J., Jonker P. G., Nelemans G. e. a., 2024, PASP, 136, 115003
Gupta R. R. et al., 2016, AJ, 152, 154
Harris F. J., 1978, Proc. IEEE, 66, 51
Harutyunyan A. H. et al., 2008, A&A, 488, 383
Heavens A. F., 1993, MNRAS, 263, 735
Hook I. M., Jørgensen I., Allington-Smith J. R., Davies R. L., Metcalfe N., Murowinski R. G., Crampton D., 2004, PASP, 116, 425
Howell D. A. et al., 2005, ApJ, 634, 1190
Ivezić Ž. et al., 2019, ApJ, 873, 111
Jin S. et al., 2024, MNRAS, 530, 2688
de Jong R. S. et al., 2019, The Messenger, 175, 3
Lin L. I.-K., 1989, Biometrics, 45, 255
Liu Y., Modjaz M., 2014, preprint (arXiv:1405.1437)
Liu Y.-Q., Modjaz M., Bianco F. B., Graur O., 2016, ApJ, 827, 90
Liu Y.-Q., Modjaz M., Bianco F. B., 2017, ApJ, 845, 85
Magill D., Fulton M. D., Nicholl M., Smartt S. J., Angus C. R., Srivastav S., Smith K. W., 2025, Res. Notes AAS, 9, 78
McLachlan G. J., Lee S. X., Rathnayake S. I., 2019, Annual Review of Statistics and Its Application, 6, 355
Modjaz M., Liu Y. Q., Bianco F. B., Graur O., 2016, ApJ, 832, 108
Muthukrishna D., Parkinson D., Tucker B. E., 2019, ApJ, 885, 85
Oke J. B. et al., 1995, PASP, 107, 375
Piascik A. S., Steele I. A., Bates S. D., Mottram C. J., Smith R. J., Barnsley R. M., Bolton B., 2014, in Ramsay S. K., McLean I. S., Takami H. eds, Proc. SPIE Conf. Ser. Vol. 9147, Ground-based and Airborne Instrumentation for Astronomy V. SPIE, Bellingham, p. 91478H
Savitzky A., Golay M. J. E., 1964, Anal. Chem., 36, 1627
Schwarz G., 1978, Ann. Stat., 6, 461
Smartt S. J. et al., 2015, A&A, 579, A40
Steeghs D. et al., 2022, MNRAS, 511, 2405
Tamura N. et al., 2016, in Evans C. J., Simard L., Takami H. eds, Proc. SPIE Conf. Ser. Vol. 9908, Ground-based and Airborne Instrumentation for Astronomy VI. SPIE, Bellingham, p. 99081M
Tonry J., Davis M., 1979, AJ, 84, 1511
Tonry J. L. et al., 2018, PASP, 130 064505
Tranin H. et al., 2026, A&A, 706, A284
Tucker M. A. et al., 2022, PASP, 134, 124502
Vernet J. et al., 2011, A&A, 536, A105
Williamson M., Modjaz M., Bianco F. B., 2019, ApJ, 880, L22
Williamson M. et al., 2023, ApJ, 944, L49
Yaron O., Gal-Yam A., 2012, PASP, 124, 668
Yesmin N. et al., 2025, A&A, 693, A307
Zamora S., Díaz A. I., 2023, preprint (arXiv:2310.04133)
Zucker S., 2003, MNRAS, 342, 1291

APPENDIX A: TEMPLATE MANAGER

SNID–SAGE includes a dedicated Template Manager, accessible via the `snid-sage-templates` command, for inspecting, curating, and extending the spectral template library. The tool serves both as a quality-control environment for the distributed templates and as a workflow for incorporating user-contributed spectra in a controlled and reproducible manner.

The Template Manager allows interactive inspection of template spectra across multiple epochs, including visual comparison, metadata editing, and verification of subtype, phase, and redshift assignments (Fig. A1). Users can create new templates from input spectra using the same preprocessing procedures employed in the main classification pipeline, ensuring full consistency between template construction and analysis.

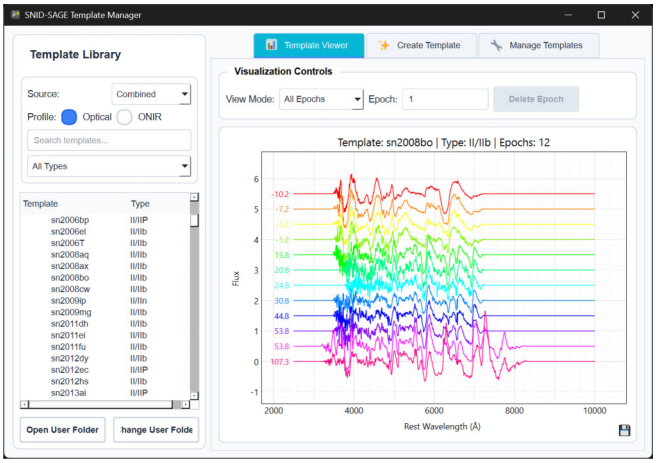

Figure A1. Template Manager interface. Example view of a multi-epoch template (SN 2008bo, Type IIb) with phase-coded spectra and interactive inspection tools.

User-created templates are maintained separately from the distributed base library and are merged at runtime, allowing corrections or extensions without modifying the original data set.

APPENDIX B: EXAMPLE OF AI-ASSISTED SUMMARY

As described in Section 2.4.3, the SNID–SAGE graphical interface includes an optional feature that generates a structured natural-language summary of the classification results using an externally configured LLM. This functionality is intended purely for reporting and human-readable interpretation; it does not influence any part of the numerical analysis.

The LLM is provided with a formatted summary of the SNID–SAGE output, including the best-fitting type and subtype, redshift and phase estimates with uncertainties, match-quality and confidence metrics, and a subset of the highest-ranked template matches. An example output, representative of that produced using a publicly available model via OpenRouter, is shown below.

SNID AI ANALYSIS SUMMARY

The spectrum is classified as a Type II supernova of subtype IIP, best matched by the template `sn1999em` (II/IIP), with High match quality and High confidence relative to the next-best alternative. The cluster-weighted redshift is $z = 0.0180 \pm 0.0021$, and the inferred phase is approximately 2–3 d relative to maximum light, with an uncertainty of about 11 d. The consistency of the top-ranked template matches supports the robustness of the classification, although the phase estimate remains moderately uncertain.

The exact wording of the generated summary will vary depending on the chosen model and prompt configuration. The summaries are intended to facilitate rapid reporting and interpretation, particularly in time-domain follow-up settings, and should be regarded as descriptive outputs derived from the de-

terministic SNID–SAGE analysis rather than as independent classifications.

APPENDIX C: ADDITIONAL REDSHIFT VALIDATION DIAGNOSTICS

To complement the summary statistics in Section 4.2, we show the redshift residuals as a function of host-galaxy redshift, separated by match-quality class. This figure is intended as a descriptive diagnostic rather than a full uncertainty decomposition, since the observed residuals include contributions from both the SNID–SAGE redshift estimate and the external host-galaxy redshift reference.

As shown in Fig. C1, the dominant trend is the clear reduction in scatter and catastrophic outliers towards higher match quality. The Very Low and Low classes show the broadest residual distributions, while the Medium and High classes remain much more tightly clustered around zero. Over the sampled redshift range, there is no strong evidence for a large systematic drift in the residuals with redshift itself. Some bins at higher redshift contain relatively few spectra, and their apparent scatter should therefore be interpreted with caution. In addition, a small number of points follow a diagonal sequence toward negative residuals; these are likely cases where spectra were uploaded to WISeREP already corrected to, or close to, the rest frame, producing SNID–SAGE redshifts near zero when compared with non-zero host-galaxy redshifts.

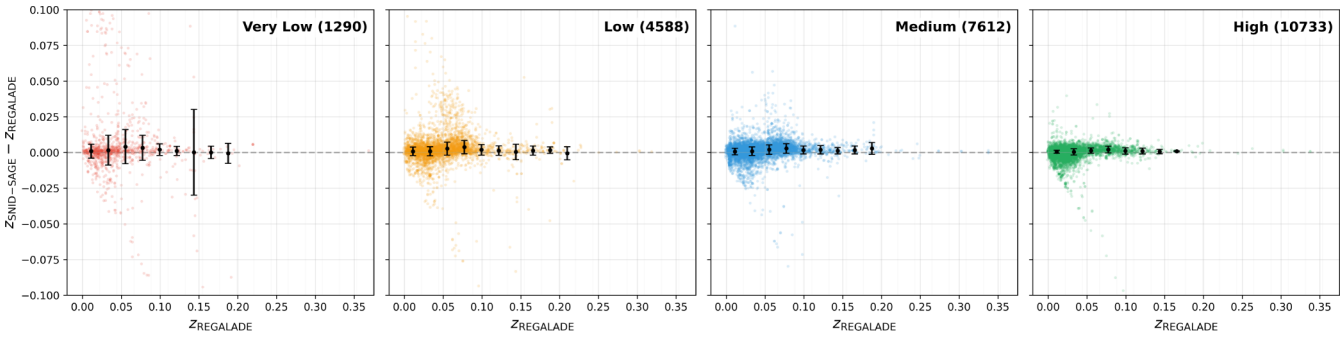

Figure C1. Redshift residuals as a function of REGALADE host-galaxy redshift, separated by SNID–SAGE match-quality class. The vertical axis shows $\Delta z = z_{\rm SNID-SAGE} - z_{\rm REGALADE}$. The dashed horizontal line marks zero residual. Coloured points show individual spectra. Black points show the median residual in bins of REGALADE redshift, with error bars corresponding to 1.4826 times the median absolute deviation. Higher match quality classes exhibit substantially reduced scatter and fewer catastrophic outliers. Some high-redshift bins contain relatively few spectra, so their scatter should be interpreted cautiously. A small number of points with approximately diagonal negative residuals likely correspond to spectra uploaded in, or close to, the rest frame. The observed dispersion includes contributions from both the SNID–SAGE redshift estimate and the external host-galaxy redshift reference.

This paper has been typeset from a TEX/LATEX file prepared by the author.