

# SUGAR: An improved empirical model of Type Ia supernovae based on spectral features<sup>★</sup>

P.-F. Léget<sup>1,2,3</sup>, E. Gangler<sup>1</sup>, F. Mondon<sup>1</sup>, G. Aldering<sup>4</sup>, P. Antilogus<sup>3</sup>, C. Aragon<sup>4</sup>, S. Bailey<sup>4</sup>, C. Baltay<sup>5</sup>, K. Barbary<sup>4</sup>, S. Bongard<sup>3</sup>, K. Boone<sup>4,6</sup>, C. Buton<sup>7</sup>, N. Chotard<sup>7</sup>, Y. Copin<sup>7</sup>, S. Dixon<sup>4</sup>, P. Fagrelis<sup>4,6</sup>, U. Feindt<sup>8</sup>, D. Fouchez<sup>9</sup>, B. Hayden<sup>4</sup>, W. Hillebrandt<sup>10</sup>, A. Kim<sup>4</sup>, M. Kowalski<sup>11,12</sup>, D. Kuesters<sup>11</sup>, S. Lombardo<sup>11</sup>, Q. Lin<sup>13</sup>, J. Nordin<sup>11</sup>, R. Pain<sup>3</sup>, E. Pecontal<sup>14</sup>, R. Pereira<sup>7</sup>, S. Perlmutter<sup>4,6</sup>, K. A. Ponder<sup>15</sup>, M. V. Pruzhinskaya<sup>1,16</sup>, D. Rabinowitz<sup>5</sup>, M. Rigault<sup>1</sup>, K. Runge<sup>4</sup>, D. Rubin<sup>4,17</sup>, C. Saunders<sup>3</sup>, L.-P. SAYS<sup>1</sup>, G. Smadja<sup>7</sup>, C. Sofiatti<sup>4,6</sup>, N. Suzuki<sup>4,20</sup>, S. Taubenberger<sup>10,18</sup>, C. Tao<sup>9,13</sup>, and R. C. Thomas<sup>19</sup> (The Nearby Supernova Factory)

<sup>1</sup> Université Clermont Auvergne, CNRS/IN2P3, Laboratoire de Physique de Clermont, 63000 Clermont-Ferrand, France  
e-mail: pierrefrancois.leget@gmail.com

<sup>2</sup> Kavli Institute for Particle Astrophysics and Cosmology, Department of Physics, Stanford University, Stanford, CA 94305, USA

<sup>3</sup> LPNHE, CNRS/IN2P3, Sorbonne Université, Paris Diderot, Laboratoire de Physique Nucléaire et de Hautes Énergies, 75005 Paris, France

<sup>4</sup> Physics Division, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley CA 94720, USA

<sup>5</sup> Department of Physics, Yale University, New Haven CT 06250-8121, USA

<sup>6</sup> Department of Physics, University of California Berkeley, 366 LeConte Hall MC 7300, Berkeley, CA 94720-7300, USA

<sup>7</sup> Université de Lyon, 69622, Lyon, France; Université de Lyon 1, Villeurbanne; CNRS/IN2P3, Institut de Physique des Deux Infinis, Lyon, France

<sup>8</sup> The Oskar Klein Centre, Department of Physics, AlbaNova, Stockholm University, 106 91 Stockholm, Sweden

<sup>9</sup> Aix Marseille Université, CNRS/IN2P3, CPPM, UMR 7346, 13288 Marseille, France

<sup>10</sup> Max-Planck Institut für Astrophysik, Karl-Schwarzschild-Str. 1, 85748 Garching, Germany

<sup>11</sup> Institut für Physik, Humboldt-Universität zu Berlin, Newtonstr. 15 12489, Berlin

<sup>12</sup> Deutsches Elektronen-Synchrotron, 15735 Zeuthen, Germany

<sup>13</sup> Tsinghua Center for Astrophysics, Tsinghua University, Beijing 100084, PR China

<sup>14</sup> Centre de Recherche Astronomique de Lyon, Université Lyon 1, 9 avenue Charles André, 69561 Saint Genis Laval, France

<sup>15</sup> Berkeley Center for Cosmological Physics, University of California Berkeley, 341 Campbell Hall, Berkeley, CA 94720, USA

<sup>16</sup> Lomonosov Moscow State University, Sternberg Astronomical Institute, Universitetsky pr. 13, Moscow 119234, Russia

<sup>17</sup> Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218, USA

<sup>18</sup> European Southern Observatory, Karl-Schwarzschild-Str. 2, 85748 Garching, Germany

<sup>19</sup> Computational Cosmology Center, Computational Research Division, Lawrence Berkeley National Laboratory, 1 Cyclotron Road MS 50B-4206, Berkeley, CA 94720, USA

<sup>20</sup> Kavli Institute for the Physics and Mathematics of the Universe, University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa, Chiba 277-8583, Japan

Received 21 December 2018 / Accepted 16 September 2019

## ABSTRACT

**Context.** Type Ia supernovae (SNe Ia) are widely used to measure the expansion of the Universe. Improving distance measurements of SNe Ia is one technique to better constrain the acceleration of expansion and determine its physical nature.

**Aims.** This document develops a new SNe Ia spectral energy distribution (SED) model, called the SUPernova Generator And Reconstructor (SUGAR), which improves the spectral description of SNe Ia, and consequently could improve the distance measurements.

**Methods.** This model was constructed from SNe Ia spectral properties and spectrophotometric data from the Nearby Supernova Factory collaboration. In a first step, a principal component analysis-like method was used on spectral features measured at maximum light, which allowed us to extract the intrinsic properties of SNe Ia. Next, the intrinsic properties were used to extract the average extinction curve. Third, an interpolation using Gaussian processes facilitated using data taken at different epochs during the lifetime of an SN Ia and then projecting the data on a fixed time grid. Finally, the three steps were combined to build the SED model as a function of time and wavelength. This is the SUGAR model.

**Results.** The main advancement in SUGAR is the addition of two additional parameters to characterize SNe Ia variability. The first is tied to the properties of SNe Ia ejecta velocity and the second correlates with their calcium lines. The addition of these parameters, as well as the high quality of the Nearby Supernova Factory data, makes SUGAR an accurate and efficient model for describing the spectra of normal SNe Ia as they brighten and fade.

**Conclusions.** The performance of this model makes it an excellent SED model for experiments like the Zwicky Transient Facility, the Large Synoptic Survey Telescope, or the Wide Field Infrared Survey Telescope.

**Key words.** supernovae: general – cosmology: observations

\* The data are only available at the CDS via anonymous ftp to [cdsarc.u-strasbg.fr](https://cdsarc.u-strasbg.fr) (130.79.128.5) or via <http://cdsarc.u-strasbg.fr/viz-bin/cat/J/A+A/636/A46>

## 1. Introduction

Type Ia supernovae (SNe Ia) are excellent cosmological probes: they are very luminous objects that are visible up to a redshift of  $z \sim 2$  (Guillochon et al. 2017), and their luminosity dispersion is naturally low and can be further reduced by an appropriate standardization process. As precise distance indicators, comparing their luminosity and their redshift allowed Perlmutter & Aldering (1998), Perlmutter et al. (1999), Riess et al. (1998), and Schmidt et al. (1998) to demonstrate that the expansion of the Universe is accelerating; a feature that gave rise to the dark energy paradigm. This result, obtained with a small sample of SNe Ia, has since been repeatedly confirmed with larger samples (Astier et al. 2006; Guy et al. 2010; Suzuki et al. 2012; Rest et al. 2014; Betoule et al. 2014; Scolnic et al. 2018), and in combination with other cosmological probes like the cosmological microwave background (CMB; Planck Collaboration XIII 2016), baryon acoustic oscillations (BAO; Delubac et al. 2015), and cosmic shear (Troxel et al. 2018) led to the so-called concordance model, the flat- $\Lambda$  cold dark matter model.

Improving our description of SNe Ia as a cosmological probe is needed, not only to discriminate among alternate dark energy models (Copeland et al. 2006), but also in order to address existing tensions, such as the  $3.4\sigma$  difference between the values of the Hubble constant  $H_0$  from the local measurement by Riess et al. (2016) and the cosmological fit from the Planck Collaboration XIII (2016). The current uncertainty budget due to limited SN Ia statistics will be greatly improved by current surveys like the Zwicky Transient Facility (ZTF; Bellm 2014) or next generation surveys like the Large Synoptic Survey Telescope (LSST) or the Wide Field Infrared Survey Telescope (WFIRST) (LSST Dark Energy Science Collaboration 2012; Spergel et al. 2015). Since, systematic and statistical uncertainties are already of the same order of magnitude (Betoule et al. 2014; Scolnic et al. 2018), a better understanding of systematics will be needed to improve the accuracy of SNe Ia as a cosmology probe.

Part of the systematic uncertainty comes from the standardization process; the observed flux of the SN Ia has to be corrected for variations observed from one object to another. Two main contributions to this variation have been observed. The first one, seen by Rust (1974) and Pskovskii (1977, 1984), is the correlation between the peak luminosity of a SNe Ia and the light curve decrease time, the so-called brighter-slower effect. Various parametrizations have been proposed for this effect, the most commonly used being  $\Delta m_{15}$  (Phillips 1993), stretch (Perlmutter et al. 1997), or  $X_1$  (Guy et al. 2007). The second contribution, observed first by Hamuy et al. (1995) and Riess et al. (1996), is that the peak luminosity depends on color, the so-called brighter-bluer effect. This effect can be explained by the presence of dust in varying quantities along the line of sight that is possibly combined with an intrinsic color-brightness correlation after the stretch effect has been taken into account. Most standardization techniques in photometry thus rely on a stretch and color relation known as the Tripp (1998) relation. The SALT2 model (Guy et al. 2007) is one such standardization method, and has over the years become the reference in cosmological analysis. Despite many attempts, no consensus has yet emerged on how to go beyond a two-component model to describe SNe Ia light curves. However, the observation of a step in the standardized luminosity with respect to the host mass by Kelly et al. (2010) and Sullivan et al. (2010) has led to the inclusion of a corrective term in the subsequent cosmological anal-

ysis. This corrective term linked to the environment hints that more fundamental properties of SNe Ia physics are not captured by the stretch-color standardization scheme, and that there is room for improvement.

However, the most obvious indication that the parameterization used is insufficient comes from the observation of a residual dispersion of SNe Ia luminosity around the Hubble diagram after standardization. In order to obtain statistically coherent results, Perlmutter et al. (1999) and Astier et al. (2006) introduced an intrinsic dispersion in luminosity as an additional uncertainty in the fit. Betoule et al. (2014) estimate the dispersion value at 0.11 mag using a SALT2 standardization, for an observed total dispersion of 0.16 mag. Since the intrinsic dispersion is due to unmodeled SN Ia variations that may depend on the redshift, this may result in a significant error on the extraction of cosmological parameters (Rigault et al. 2018). The precision of SNe Ia as cosmological probes therefore depends on this intrinsic dispersion in luminosity. However, the intrinsic dispersion strongly depends on the assumptions about measurement uncertainties. A way to characterize the overall accuracy obtained by a given standardization method is to use the weighted root mean square (wRMS) metric on the standardized magnitude (e.g., Blondin et al. 2011). Thus, in order to reduce the effects of unmodeled variations of SNe Ia, an improved standardization procedure must also involve reducing the wRMS.

Most efforts to improve the standardization of SNe Ia involve the search for a new parameter correlated with the intrinsic luminosity. After the discovery of the mass-step, many efforts focused on describing the environmental effects of the SNe Ia (Kelly et al. 2010; Sullivan et al. 2010; Rigault et al. 2013, 2015, 2018; Roman et al. 2018). A complementary approach consists of directly looking for this new parameter from the analysis of the light curves or observed spectra. For example, Mandel et al. (2017) proposed that the color of SNe Ia is a mixture of intrinsic color and extinction by dust, and proposed a Bayesian model to separate these two components. For their part, Chotard et al. (2011) showed that once the Ca II H&K variability is taken into account, the color law is compatible with a Cardelli et al. (1989) extinction curve. Moreover, Mandel et al. (2014) highlighted the dependence of the extinction curve on the minima of P Cygni profiles. Thus the use of additional variables related to spectral indicators seems to be a promising avenue to describe the variability of SNe Ia.

The use of spectral indicators to standardize SNe Ia has a long history (Nugent et al. 1995; Arsenijevic et al. 2008; Bailey et al. 2009; Wang et al. 2009; Foley & Kasen 2011; Chotard et al. 2011). As example, using only the flux ratio  $\mathcal{R}_{642/443}$ , Bailey et al. (2009) is able to get a significantly better standardization in  $B$ -band in comparison to the classical Tripp (1998) relation. The method of using spectral information is also suggested by the analysis of Fakhouri et al. (2015), which shows that the best method to measure distance with SNe Ia is with spectroscopic twins SNe Ia. Another recent study by Nordin et al. (2018) has shown that the use of spectral information from the UV part of the spectra improves distance measurement compared to the Tripp (1998) relation. However, those methods do not currently lead to a spectral energy distribution (SED) model, which is necessary for cosmological analyses on purely photometric data, such as LSST (LSST Dark Energy Science Collaboration 2012). Thus we propose here a full SED model which will be based on spectral indicators generalizing the procedure originally developed in

Chotard et al. (2011). This method uses spectral features, which allows for the addition of more than one intrinsic parameter and they offer a possible way to separate intrinsic properties from extrinsic properties. From it, we build a full SED model which may be used for purely photometric surveys.

The aim of this study is to revisit the parametrization of SNe Ia SED in light of the SNfactory spectrophotometric dataset (Aldering et al. 2002), and to seek new sources of variability by statistical analysis. Our model is trained using spectral indicators, derived around maximum light in *B*-band, as features to describe the model. They provide both a reduced dimensionality description of spectra, and a description which is linked to the physics of the explosion. In addition, we select indicators insensitive to reddening in order to decouple the characterization of the reddening from effects purely linked to the intrinsic part of the explosion. This new SNe Ia SED model is named the SUPERnova Generator And Reconstructor (SUGAR) model.

Another approach to develop a new SED model was undertaken in parallel by Saunders et al. (2018), using the same dataset and based on a “SALT2 like” framework. Saunders et al. (2018) generalized a strategy that was originally proposed by Guy et al. (2007) and which was first performed on broad band photometry from the SNfactory dataset in Kim et al. (2013). In brief, Saunders et al. (2018) did a principal component analysis (PCA)-like study on interpolated spectral time series in order to go beyond the classical Tripp (1998) relation. Saunders et al. (2018) is able to significantly improve the SED description with respect to the SALT2 model. The SNEMO model developed in Saunders et al. (2018) differs from the SUGAR model in that SNEMO attempts to find principal components purely based on spectral time series variability in the relative luminosities at each wavelength and as a function of time, whereas SUGAR uses spectral features to try to find a compact description of such spectral time series. However, both models share some technical details in the model training and were developed in common.

This paper is organized as follows: Sect. 2 presents how the spectrophotometric time series of the SNfactory were obtained, Sect. 3 focuses on the intermediary data used in building the SUGAR model (spectral indicators at maximum light, dimensionality reduction through factor analysis, derivation of extinction parameters and time interpolation). Section 4 describes the SUGAR model: its formalism, training, and components. Section 5 presents the performance of the model for fitting spectral time series using SUGAR model. These results are compared to the performance achieved by SALT2 for the same data. Finally, Sect. 6 discusses adding additional components to the SUGAR model and some technical choices that were made in the training of the SUGAR model. The appendices describe details of the mathematical implementation of the SUGAR model. The SUGAR template is available online<sup>1</sup>, while the data used to train SUGAR are available online<sup>2</sup>.

## 2. Spectrophotometric time series from SNfactory

This analysis is based on 171 SNe Ia obtained by the SNfactory collaboration beginning in 2004 (Aldering et al. 2002) with the SuperNova Integral Field Spectrograph (SNIFS, Lantz et al. 2004) installed on the University of Hawaii 2.2-m telescope (Mauna Kea). SNIFS is a fully integrated instrument optimized for semi-automated observations of point sources on a structured

background over an extended optical window at moderate spectral resolution. SNIFS has a fully-filled  $6.4'' \times 6.4''$  spectroscopic field-of-view subdivided into a grid of  $15 \times 15$  contiguous square spatial elements (spaxels). The dual-channel spectrograph simultaneously covers 3200–5200 Å (*B*-channel) and 5100–10 000 Å (*R*-channel) with 2.8 and 3.2 Å resolution, respectively. The data reduction of the  $x, y, \lambda$  data cubes is summarized by Aldering et al. (2006) and updated in Sect. 2.1 of Scalzo et al. (2010). The flux calibration is developed in Sect. 2.2 of Pereira et al. (2013) based on the atmospheric extinction derived in Buton et al. (2013). In addition, observations are obtained at the SN Ia location at least one year after the explosion to serve as a final reference to enable subtraction of the underlying host and the host subtraction, as described in Bongard et al. (2011). For every SN Ia followed, the SNfactory creates a spectrophotometric time series, typically composed of  $\sim 14$  epochs, with the first spectrum taken on average three days before maximum light in *B*-band (Bailey et al. 2009; Chotard et al. 2011). The sample of 171 SNe Ia contains the objects with good final references, objects that passed quality cuts suggested by Guy et al. (2010), and is restricted to objects with at least one observation in a time window of  $\pm 2.5$  days around maximum light in *B*-band. The size of this window is kept identical with respect to the study of Chotard et al. (2011) and is discussed in Léget (2016). After flux calibration, host-galaxy subtraction and correction for Milky Way extinction, the flux of the 171 spectra is integrated in synthetic top-hat filters defined in Pereira et al. (2013) and for reference a SALT2 fit is applied with the model from Betoule et al. (2014) in order to obtain the  $X_1, C$ , and  $m_B$  parameters. The spectra of the 171 SNe Ia are transformed to the rest frame with a fiducial cosmology and an arbitrary Hubble constant. It is allowing to respect blinding of future cosmological analysis using this dataset. For the spectral analysis here, the spectra are rebinned at  $1500 \text{ km s}^{-1}$  between 3254 and 8649 Å (197 bins per spectra) for computational efficiency while still resolving spectral features and converted to the absolute AB magnitude system. The training is done on 113 SNe Ia observed before 2010 and the validation is done on 58 SNe Ia observed after 2010.

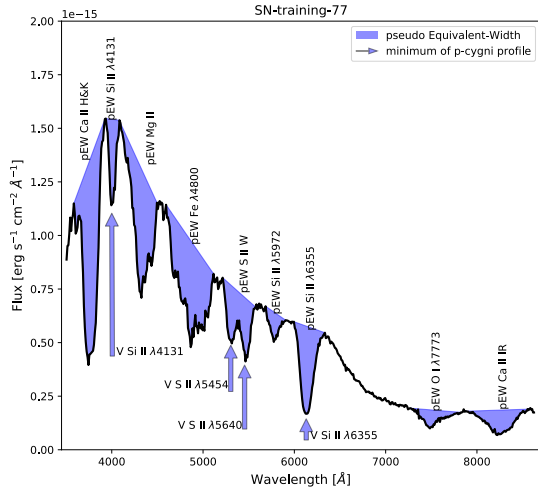
## 3. Derived data

Any empirical SN Ia modeling must solve three problems<sup>3</sup>: Choose what features to model, account for color, and deal with the data sampling. For the features modeling of SUGAR, we used spectral features at maximum light to describe the intrinsic part of the SED. This is described successively in Sects. 3.1 and 3.2. Section 3.1 describes how the spectral features are selected and measured. Section 3.2 explains how the spectral features are projected onto a new basis that allows us to work in an orthogonal basis for the SUGAR training. To estimate the average color curve of SNe Ia, we generalize the method of Chotard et al. (2011). This is described in Sect. 3.3. To deal with the data sampling, we project the observed spectra onto a common time grid so that the SED model can be calculated on the same time grid. This is done in Sect. 3.4 using the Gaussian process method. The following sections describe those intermediate steps that determine the full SUGAR SED model in Sect. 4. We use the spectral features derived and color curve parameters

<sup>3</sup> In general, empirical SNIa modeling must also handle spectroscopic and photometric data that are not observed on the same day or by the same instruments. However, this is not a problem with the SNfactory dataset.

<sup>1</sup> <http://supernovae.in2p3.fr/sugar>

<sup>2</sup> <https://snfactory.lbl.gov/sugar>



**Fig. 1.** Pseudo-equivalent widths of absorption lines at maximum light as well as minima of P-Cygni profiles at maximum of light which are used in our analysis. They are represented on the spectrum of SN-training-77 at a phase of 1 day after maximum brightness.

derived at maximum light combined with the interpolated spectra to infer the full SED at all epochs.

### 3.1. Spectral indicators at maximum light

Spectral indicators are metrics of empirical features of the input spectrum such as equivalent widths or line velocities. They offer an efficient characterization of spectral variability by representing the underlying spectral complexity with a few key numbers, and play a role in nonlinear dimensionality reduction of the original data. These two characteristics, interpretability and simplicity, make them ideal for describing the intrinsic part of SNe Ia SED, and consequently derive the extrinsic part in the same spirit as Chotard et al. (2011). This decoupling restricts the set of spectral indicators to pseudo-equivalent widths and the wavelengths of P-Cygni profile minima. As the spectral indicators evolve with phase, we select the spectrum closest to maximum light in *B*-band, if it is within a time window of  $\pm 2.5$  days. This window size was chosen to optimize the trade-off between the total number of SNe Ia in the sample and the potential loss in precision due to time evolution.

Near maximum light and within the spectral range covered by the SNfactory, it is possible to systematically obtain the 13 spectral indicators represented in Fig. 1. They are the nine pseudo-equivalent widths of Ca II H&K, Si II  $\lambda 4131$ , Mg II, Fe  $\lambda 4800$ , Si II W, Si II  $\lambda 5972$ , Si II  $\lambda 6355$ , O I  $\lambda 7773$ <sup>4</sup>, and Ca II IR features, as well as four minima of P Cygni profiles (Si II  $\lambda 4131$ , Si II  $\lambda 5454$ , Si II  $\lambda 5640$ , and Si II  $\lambda 6355$ ). For the pseudo-equivalent widths, we rely only on well-defined troughs that are present in all SNe Ia in the sample, and consider line blends as a whole. Some of the possible minima were also discarded either because the corresponding lines form a complex mixture or because the corresponding trough is too shallow to accurately define a minimum for some SNe Ia. This is why the Si II  $\lambda 5972$  feature velocity is rejected. The spectral indicators and their uncertainties were automatically derived from the spectra at maximum light following the procedure of Chotard et al. (2011), which is described in detail in Appendix A.

<sup>4</sup> We refer to this as O I  $\lambda 7773$  but note it overlaps with Mg II  $\lambda 7812$ .

## 3.2. Factor analysis on spectral features

### 3.2.1. Factor analysis model

The space defined by the 13 selected spectral indicators has too high dimensionality to efficiently train a model. Additionally, some of the spectral indicators are correlated, and therefore contain redundant information. Most of the model variation can be captured in a reduced number of dimensions. Principal component analysis (Pearson 1901) is one of the methods to implement such a dimensionality reduction. It consists of diagonalizing the covariance matrix of the sample and projecting the data into the resulting eigenvectors basis. In this new basis, the variables are uncorrelated, and an approximation of the input data is found by neglecting the dimensions corresponding to the smallest eigenvalues. This method has been employed in the case of SNe Ia by Guy et al. (2007), Kim et al. (2013), and Sasdelli et al. (2015). However, in the case considered here, some directions are dominated by noise, so their eigenvectors would align along the direction of measurement errors rather than the intrinsic sample variance. To solve this problem, we employ a variant of PCA, Factor Analysis (Spearman 1904, 1927), which has the advantage of taking into account the variance caused by measurement uncertainties. This technique decomposes the observables into two terms, one representing the explanatory factors and one representing the noise affecting each variable. This is expressed by the following relation (Ghahramani & Hinton 1997):

$$\mathbf{x}_i = \mathbf{\Lambda} \mathbf{q}_i + \mathbf{u}_i, \quad (1)$$

where  $\mathbf{x}_i$  is the spectral feature vector with  $s$  components for the  $i$ th SNe Ia.  $\mathbf{q}_i$  is the explanatory factor vector of dimension  $m \leq s$ , and is linearly related to  $\mathbf{x}_i$  by the matrix  $\mathbf{\Lambda}$  of dimensions  $m \times s$ .  $\mathbf{u}_i$  is the noise with variance  $\mathbf{\Psi}$  of dimensions  $s \times s$ . In order to fix the normalization of  $\mathbf{\Lambda}$  and  $\mathbf{q}_i$ , we assume that the  $\mathbf{q}_i$  are drawn from a centered normal distribution:

$$P(\mathbf{q}_i) \sim \mathcal{N}(0, \mathbf{I}). \quad (2)$$

In the framework of the PCA, and up to a normalization, the matrix  $\mathbf{\Lambda}$  and the  $\mathbf{q}_i$  are respectively equivalent to the eigenvector matrix and the projections into the new basis. Factor analysis thus consists of determining the matrices  $\mathbf{\Lambda}$  and  $\mathbf{\Psi}$  that maximize the likelihood, under the assumption that the  $\mathbf{x}_i$  are distributed according to a normal distribution with both intrinsic scatter and measurement noise:

$$P(\mathbf{x}_i) \sim \mathcal{N}(0, \mathbf{\Lambda} \mathbf{\Lambda}^T + \mathbf{\Psi}). \quad (3)$$

To estimate  $\mathbf{\Lambda}$ ,  $\mathbf{\Psi}$ , and the explanatory factors  $\mathbf{q}_i$ , Ghahramani & Hinton (1997) propose a solution based on an expectation-maximization algorithm where  $\mathbf{q}_i$  and  $\mathbf{\Lambda}$  are estimated iteratively. Here, unlike in conventional factor analysis, a reliable estimate of the spectral indicator measurement error is provided. For each SNe Ia  $i$ ,  $\mathbf{\Psi}_i$  is the known diagonal matrix that contains the squared errors of the spectral indicators, hence, it is not necessary to fit for a global  $\mathbf{\Psi}$ . We adapted the aforementioned expectation-maximization algorithm to take this change into account. Additional details can be found in Appendix B.

Finally, a prescription is needed to normalize each individual variable. The amplitude of the spectral indicator variation, expressed in Å, is not a good indicator of their impact on the spectral shape. As an example, Branch et al. (2006) show that the weak Si II  $\lambda 5972$  line can play a significant role in subclassifying SNe Ia. As a consequence, the input data are normalized to

unit variance prior to the factor analysis, so that no spectral indicator is favored a priori. Within this framework, the eigenvalues of  $\Lambda\Lambda^T$  represent the variance explained by each factor, and  $s - \text{Tr}(\Lambda\Lambda^T)$  is the variance coming from the noise, where here  $s = 13$  is the number of spectral indicators and  $\text{Tr}$  is the trace operator.

### 3.2.2. Outlier rejection

Outliers affect the sample variance of any population. As we care most about a correct description of the bulk of SNe Ia, it is desirable to identify and remove these outliers. Amongst them, SNe Ia of type SN1991T (Filippenko et al. 1992a) or SN1991bg (Filippenko et al. 1992b) are known to have different spectral and photometric behavior from other SNe Ia. However, basing an outlier rejection on this empirical identification has two issues. The first is that within the spectral indicator space, some of those subtypes may not appear as distinct subclasses, and do not offer objective grounds for rejection. The second issue is that the attribution of a given SN Ia to one of these subclasses by SNID (Blondin & Tonry 2007) may provide inconsistent results depending on the epoch considered for the identification. In order to apply a self-contained criterion for defining an outlier, we thus resorted to a  $\chi^2$ -based definition for identifying outliers: this quantity can be interpreted as the squared distance to the center of the distribution normalized by the natural dispersion. For SN Ia  $i$ , it is given by

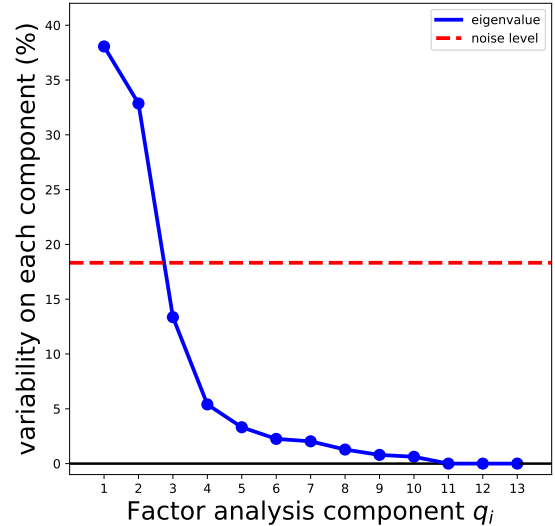
$$\chi_i^2 = \mathbf{x}_i^T (\Lambda\Lambda^T + \Psi_i) \mathbf{x}_i. \quad (4)$$

In the case of a Gaussian distribution,  $\chi_i^2$  should follow a  $\chi^2$ -distribution with 13 degrees of freedom. An iterative cut at  $3\sigma$  on the value of  $\chi_i^2$  rejects 8 SNe Ia from the sample. A visual inspection carried out on each of these 8 SNe Ia, as well as a sub-classification made using SNID, show that four of those objects exhibit shallow silicon features as defined by Branch et al. (2006), two of them have high-velocity silicon features as defined by Wang et al. (2009), one is within the broad line subcategory, and the remaining one has a failed estimation of one of the spectral indicators.

We studied the influence of the  $3\sigma$  cut by carrying out the analysis successively with and without the cut. Even though this had no major effect on the direction of the vectors, the cut is applied to the rest of the analysis to avoid training the model on outliers.

### 3.2.3. Factor analysis results on spectral features

The expectation–maximization factor analysis (EM-FA) algorithm described in the previous sections is applied to the 13 spectral features measured at maximum light. The relative weight of the eigenvalues on the total variance is shown in Fig. 2, and the correlations between the eigenvectors, and the spectral indicators are presented in Fig. 3. The significance of the correlation is expressed in units of  $\sigma$ , meaning that the null hypothesis, that is 0 correlation is rejected at  $\sigma$  confidence level. Figure 2 shows that two vectors dominate the variability in the space of spectral indicators. The first vector is linked with coherent variations of the pseudo-equivalent widths, except the widths of the Ca II H&K and Si II W lines as seen in Fig. 3. The second vector is anticorrelated with these same two features, and describes a coherent variation of the velocities. Since the first vector is very similar to the pseudo-equivalent width of Si II  $\lambda 4131$ , which is shown by Arsenijevic et al. (2008) to be highly correlated with



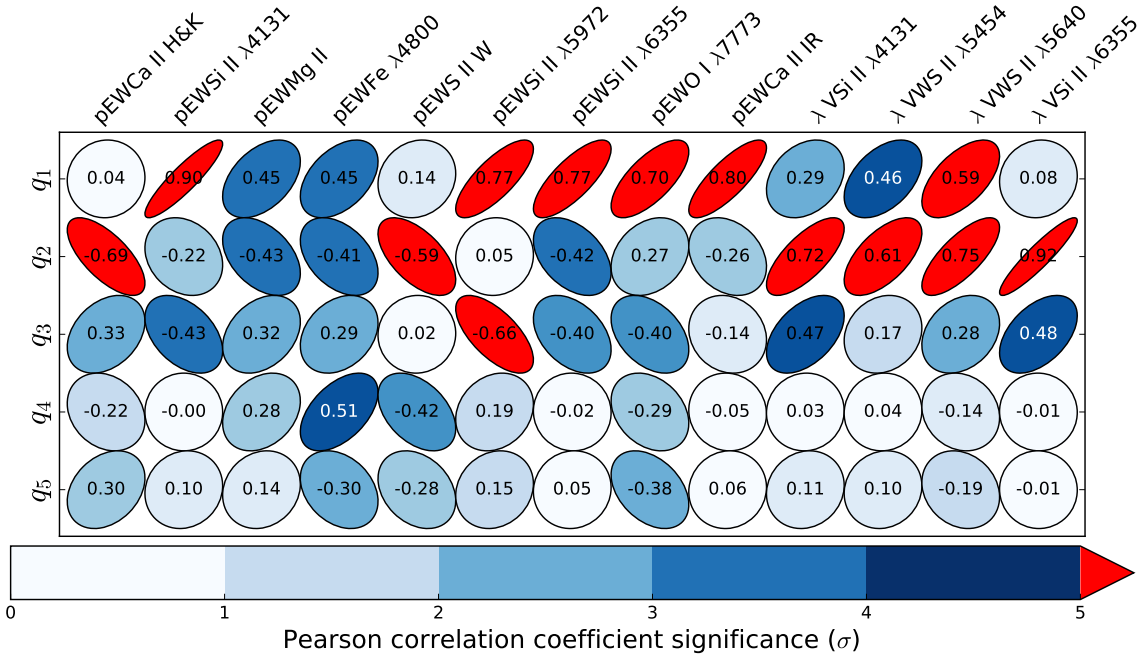
**Fig. 2.** Relative importance of the eigenvalues associated with the eigenvectors. They are ordered in decreasing order of variance, and the total amounts to 100% once the contribution of the noise (indicated by a red line) is taken into account. We can see that the first three vectors dominate the variability of this space.

the stretch parameter, it is therefore natural to assume that this vector represents the main source of intrinsic variability already known for SNe Ia. The second vector is mostly driven by the line velocities, pEW(Ca II H&K), and pEW(Si II W) lines, with smaller dependencies from the other pseudo-equivalent widths. Interestingly, while investigating the spectral diversity of SNe Ia beyond stretch and color, Chotard et al. (2011) demonstrated the role of pEW(Ca II H&K), while Wang et al. (2009) focused on the role of the speed of Si II  $\lambda 6355$ . Our second vector unifies these two approaches. The interpretation of the next vectors is less straightforward. As their rank increases, they describe less of the sample variance, and thus exhibit lower correlations with the original data. The third vector is the last of the eigenvectors to show strong correlation with one of the spectral indicators, pEW Si II  $\lambda 5972$ . The correlations with the SALT2 parameters are presented in Fig. 4. As expected from the correlations with the pseudo-equivalent widths,  $q_1$  is strongly correlated with  $X_1$ . Also,  $q_3$  exhibits a significant correlation with  $X_1$ , which may be an indication that the stretch is driven by two different parameters. Remaining factors are only weakly correlated with  $X_1$ . The correlations between SALT2 color and any of the factors are close to zero. The SALT2 absolute magnitude cosmology residual,  $\Delta\mu_B$ , are mainly correlated with  $q_1$  and  $q_3$ , presumably due to the correlation of both with  $X_1$ .

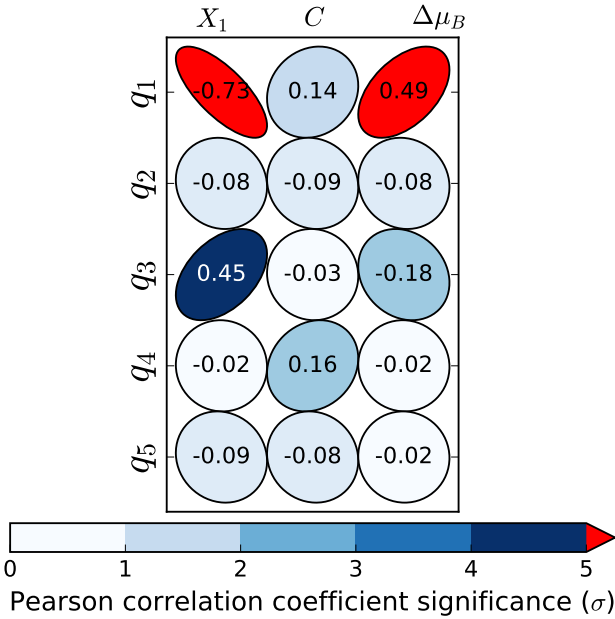
Within this framework, the eigenvalues of  $\Lambda\Lambda^T$  represent the variance explained by each factor, and  $s - \text{Tr}(\Lambda\Lambda^T)$  is the variance coming from the noise, where here  $s = 13$  is the number of spectral indicators and  $\text{Tr}$  is the trace operator. Noise, in this situation, represents 18% of the variability observed, so only the first two vectors clearly outweigh the noise and it is legitimate to ask whether or not the other vectors are sensitive to statistical fluctuations. For the rest of the analysis, we keep the first three factors for the final training and discuss this choice in the Sect. 6.2, by looking at the impact on the final model.

### 3.3. Extinction curve estimation

In this section, we estimate the average extinction curve by generalizing the procedure described in Chotard et al. (2011). For this,



**Fig. 3.** Pearson correlation coefficients between the spectral indicators and the first five factors. The eccentricities of the ellipses indicate the magnitude of the correlation, and the color the importance of the correlation in units of  $\sigma$ . The vectors one and two correspond respectively to a global correlation of the pseudo-equivalent widths and of the line velocities. Only the first three vectors display correlations with a significance higher than  $5\sigma$ .



**Fig. 4.** Pearson correlation coefficients between the SALT2 parameters and the first five vectors. The eccentricities of the ellipses indicate the magnitude of the correlation, and the color the importance of the correlation in units of  $\sigma$ .

a model of the SED at maximum light is derived. This model allows us to deduce the average extinction curve and the procedure to estimate it: the results are described in the following.

### 3.3.1. Empirical description

The SED model at maximum light presented here is a generalization of the model presented in Chotard et al. (2011), in which the

authors removed intrinsic variability tied to spectral features and measured the remaining spectral variation, which was found to be consistent with an extinction curve like that for the Galaxy. The major difference is that the intrinsic description is here motivated by the factors derived in Sect. 3.2 rather than being described only by pEW(Si II λ4131) and pEW(Ca II H&K) as was done in Chotard et al. (2011). To model the SED we define for the SN Ia  $i$  the vector  $\mathbf{x}_i \equiv \{h_{i,1}, h_{i,2}, h_{i,3}\}$ , which is the true value of the measured factor  $\mathbf{q}_i = \{q_{i,1}, q_{i,2}, q_{i,3}\}$ . We propose that these are related to the intrinsic absolute magnitude by:

$$M_i(\lambda) = M_0(\lambda) + \sum_{j=1}^{j=3} h_{i,j} \alpha_j(\lambda) + A_{\lambda_0,i} \gamma(\lambda), \quad (5)$$

where  $M_0(\lambda)$  is the average spectrum in absolute magnitude,  $\alpha_j(\lambda)$  is the spectrum related to the factor  $h_{i,j}$ , where  $j$  is the factor index running from one to three (the choice of the number of components is discussed in Sect. 6.2).  $A_{\lambda_0,i}$  is a free parameter that can be interpreted as the absorption due to extinction at a reference wavelength  $\lambda_0$  (set here at the median wavelength, 5305 Å, without loss of generality).  $\gamma(\lambda)$  is an arbitrary function that represents the effect of the extinction, and we have not set any prior on its shape. Once the function  $\gamma(\lambda)$  is fixed, it is possible to fit the total-to-selective extinction ratio of the Cardelli et al. (1989) law,  $R_V$ , and the absorption in V-band,  $A_V$  to it; this is described in the next sub-section. The model parameters here are  $M_0(\lambda)$ , the  $\alpha_j(\lambda)$ , the  $\gamma(\lambda)$ , the  $h_{i,j}$ , and the  $A_{\lambda_0,i}$ . It is understood that we are modeling the SED at specific wavelengths where we have measurements. For notational efficiency we rewrite Eq. (5) as:

$$\mathbf{M}_i = \mathbf{A} \mathbf{h}_i, \quad \text{where} \quad (6)$$

$$\mathbf{A} = (\mathbf{M}_0, \boldsymbol{\gamma}, \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3), \quad (7)$$

$$\mathbf{h}_i^T = (1, A_{\lambda_0,i}, h_{i,1}, h_{i,2}, h_{i,3}). \quad (8)$$

In Chotard et al. (2011), the parameters  $\alpha_j(\lambda)$  were computed in sequential order, and the reddening curve  $\gamma(\lambda)$  as well as the extinction  $A_{\lambda_0,i}$  were determined in a second step. We improve on this procedure by applying a global fit for all parameters at once. This is described below, and allows us to derive a global average  $R_V$  for all SNe Ia, and an  $A_V$  for each SN Ia.

### 3.3.2. Fitting the model at maximum light and $R_V$

In this section we explain the framework for fitting the free parameters of Eq. (5) and how we then determine the  $A_V$  and  $R_V$  parameters. Fitting the parameters of the absolute SED of Eq. (5) is done using an orthogonal distance regression. Estimating these parameters within the framework of an orthogonal distance regression amounts to minimizing the following  $\chi^2$ :

$$\chi^2 = \sum_i (\mathbf{M}_i^{\text{obs}} - \mathbf{A}\mathbf{h}_i)^T \mathbf{W}_{M_i} (\mathbf{M}_i^{\text{obs}} - \mathbf{A}\mathbf{h}_i) + (\mathbf{q}_i - \mathbf{x}_i)^T \mathbf{W}_{q_i} (\mathbf{q}_i - \mathbf{x}_i), \quad (9)$$

where  $\mathbf{M}_i^{\text{obs}}$  is the observed spectrum of the SN Ia  $i$  in absolute AB magnitude and  $\mathbf{W}_{M_i}$  is the weight matrix of  $\mathbf{M}_i^{\text{obs}}$ , defined as:

$$\mathbf{W}_{M_i} = \left[ \begin{pmatrix} \ddots & & 0 \\ & \sigma_{\lambda i}^2 & \\ 0 & & \ddots \end{pmatrix} + (\sigma_{\text{cal}}^2 + \sigma_z^2) \begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{pmatrix} + \mathbf{D} \right]^{-1}, \quad (10)$$

where  $\sigma_{\lambda i}$  is the uncertainty of the SN Ia  $i$  at the wavelength  $\lambda$  derived from spectral error,  $\sigma_{\text{cal}}$  the per-spectrum calibration uncertainty, taken to be 0.03 mag,  $\sigma_z$  the combination of the redshift error and the uncertainty of 300 km s<sup>-1</sup> due to peculiar velocities, and  $\mathbf{D}$  is the dispersion matrix. The dispersion matrix is fit to deal with the remaining variability which is not described by the three factors and the extinction. It includes an estimate of the remaining chromatic and achromatic (gray) dispersions. The estimation of  $\mathbf{D}$  is described in Appendix D.  $\mathbf{W}_{q_i}$  is the weight matrix of  $\mathbf{q}_i$  that comes from projecting the spectral feature uncertainties into the factor sub-space, and is defined as:

$$\mathbf{W}_{q_i} = \mathbf{\Lambda}^T \mathbf{\Psi}_i^{-1} \mathbf{\Lambda}, \quad (11)$$

where  $\mathbf{\Lambda}$  and  $\mathbf{\Psi}_i$  are the same as in Sect. 3.2. The unknown parameters of the model are the matrix  $\mathbf{A}$  and the vector  $\mathbf{h}_i$ . These are found by minimizing Eq. (9) using an expectation-minimization algorithm described in Appendix C.

Once the parameters that minimize Eq. (9) have been found, we can compare  $\gamma_\lambda$  with an extinction curve to see if it is compatible. It is straightforward to directly estimate the global mean  $R_V$  from  $\gamma_\lambda$  by minimizing

$$\chi_{R_V}^2 = \left[ \gamma - \left( \mathbf{a} + \frac{1}{R_V} \mathbf{b} \right) \right]^T \left[ \gamma - \left( \mathbf{a} + \frac{1}{R_V} \mathbf{b} \right) \right], \quad (12)$$

where the  $\mathbf{a}$  and  $\mathbf{b}$  vectors are the extinction curve coefficients defined in Cardelli et al. (1989). Minimizing  $\chi_{R_V}^2$  as a function of  $R_V$  then gives:

$$R_V = \frac{\mathbf{b}^T \mathbf{b}}{\gamma^T \mathbf{b} - \mathbf{a}^T \mathbf{b}}. \quad (13)$$

We note that the value we quote for  $R_V$  within the context of our model depends on the assumptions adopted in the evaluation of the dispersion matrix (discussed in detail in Sect. 6.1) and therefore does not necessarily correspond to the true mean properties of dust. This uncertainty associated with separating SN Ia color behavior from dust behavior is common to all SN Ia fitting methods. Once the value of the global mean  $R_V$  is fixed, we determine the absorption in the  $V$ -band,  $A_V$  for each SN Ia. This amounts to minimizing

$$\chi_{A_V}^2 = \sum_i \left( \delta \tilde{\mathbf{M}}_i - A_{V,i} \left( \mathbf{a} + \frac{1}{R_V} \mathbf{b} \right) \right)^T \mathbf{W}_{M_i} \times \left( \delta \tilde{\mathbf{M}}_i - A_{V,i} \left( \mathbf{a} + \frac{1}{R_V} \mathbf{b} \right) \right), \quad (14)$$

where  $A_{V,i}$  is the absorption in the  $V$ -band for the SN Ia  $i$  and  $\delta \tilde{\mathbf{M}}_i$  are the residuals corrected for the intrinsic variabilities

$$\delta \tilde{\mathbf{M}}_i = \mathbf{M}_i^{\text{obs}} - \mathbf{M}_0 - \sum_j q_{i,j} \alpha_j. \quad (15)$$

Taking the derivative of  $\chi_{A_V}^2$  by  $A_V$  we find:

$$A_{V,i} = \frac{\left( \mathbf{a} + \frac{1}{R_V} \mathbf{b} \right)^T \mathbf{W}_{M_i} \delta \tilde{\mathbf{M}}_i}{\left( \mathbf{a} + \frac{1}{R_V} \mathbf{b} \right)^T \mathbf{W}_{M_i} \left( \mathbf{a} + \frac{1}{R_V} \mathbf{b} \right)}. \quad (16)$$

### 3.3.3. Results

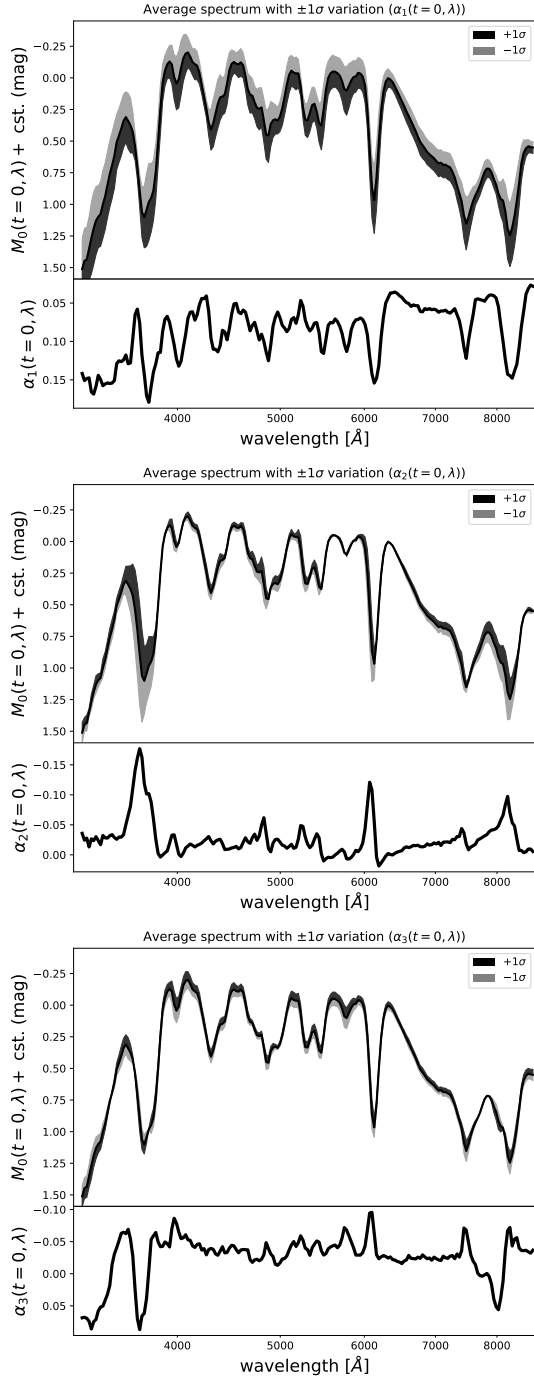
The model above is trained on the 105 SNe Ia that remain in the training sample. The vectors  $\alpha_j$  for  $j = 1$ ,  $j = 2$ , and  $j = 3$ , as well as their effect on the average spectrum  $\mathbf{M}_0$ , are presented in Fig. 5. To quantify the overall impact of each vector in units of magnitude, we compute the average rms of  $h_j \alpha_j$ , the deviation from the average spectrum  $\mathbf{M}_0$ .

The vector  $\alpha_1$  presented in Fig. 5, has an average impact of 0.17 mag when multiplied by the standard deviation of  $q_1$ . This is the expected amplitude for a stretch effect. The structure of this vector is associated with variations of the line depths, consistent with the correlation of the factor  $q_1$  with both the pseudo-equivalent widths and the stretch (cf. Figs. 3 and 4).

The vector  $\alpha_2$  shown in Fig. 5, which is correlated with velocities, has a much weaker impact of 0.05 mag on the average spectrum  $\mathbf{M}_0$ . This is consistent with the fact that this effect has not yet been detected on purely photometric data. Moreover, the associated variability is centered in localized structures such as the regions of the Ca II H&K, Si II  $\lambda$ 6355, and Ca II IR lines. A closer scrutiny shows that the variability especially affects the bluer edge of the features, leading to an overall effect on the line velocities. This is as expected from the correlations of  $q_2$  and the minima of P-Cygni profile shown in Fig. 3.

The vector  $\alpha_3$  shown in Fig. 5, has an impact of 0.05 mag on the average spectrum  $\mathbf{M}_0$ . Like the first component  $q_1$ ,  $q_3$  is correlated with the stretch (cf. Fig. 4), but the corresponding vector  $\alpha_3$  is less structured than  $\alpha_1$  in the optical band. However, the expected correlations with the spectral features are consistent with the behavior of  $q_3$  (cf. Fig. 3). Moreover, the most prominent features affect the extreme UV part of the spectrum as well as the Ca II H&K and Ca II IR regions. For the latter, brighter SNe Ia exhibit a stronger trough in the higher velocity part of the blends, which could link this vector to the presence of high-velocity calcium structure in ejectas.

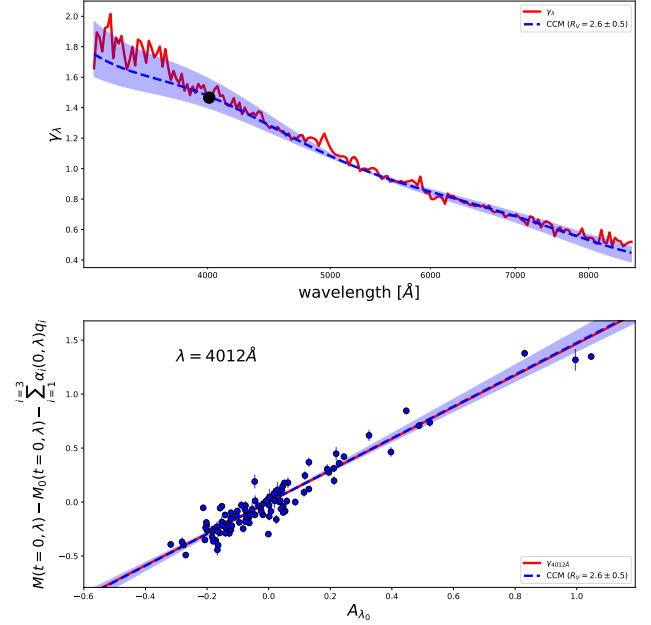
In summary, the analysis of the SED components at maximum light confirms that the stretch has the dominant effect on



**Fig. 5.** Impact of each component on the average spectrum. *Top panel*, top: average spectrum  $\mathbf{M}_0$  and the predicted effect of a variation of  $q_1$  by  $\pm 1\sigma$ . Bottom: corresponding  $\alpha_1$  vector. *Middle panel*, top: average spectrum  $\mathbf{M}_0$  and the predicted effect of a variation of  $q_2$  by  $\pm 1\sigma$ . Bottom: corresponding  $\alpha_2$  vector. *Bottom panel*, top: average spectrum  $\mathbf{M}_0$  and the predicted effect of a variation of  $q_2$  by  $\pm 1\sigma$ . Bottom: corresponding  $\alpha_2$  vector.

magnitudes as expected, but one also has to take into account other variabilities which are difficult to detect in photometric bands since they are linked to localized features such as velocities.

The  $\gamma(\lambda)$  curve is shown in Fig. 6 and is consistent with a dust extinction curve. This shows that three eigenvectors provide a sufficient description of the intrinsic variability from the purpose of deriving a color law. The fit of  $\gamma(\lambda)$  to a Cardelli et al. (1989)



**Fig. 6.** Extinction curve fit result. *Top panel*: empirical average extinction curve,  $\gamma(\lambda)$ , compared to the best fit by a Cardelli et al. (1989) law, which gives an  $R_V = 2.6$ . The shaded blue area represent a  $\pm 0.5$  range on  $R_V$ , in order to illustrate the typical range of value that are measured for SN Ia in the literature. The black dot on this curve indicates the wavelength used on the lower panel graph. *Lower panel*: residuals after correction for the intrinsic behavior at  $4012 \text{ \AA}$  as a function of  $A_{\lambda_0}$ . Each dot corresponds to a single SN Ia. The  $\gamma_{4012 \text{ \AA}}$  slope between  $M$  and  $A_{\lambda_0}$  is indicated by the red continuous line. Dashed represent the Cardelli et al. (1989) law for  $R_V = 2.6$ . The shaded blue area represent a  $\pm 0.5$  range on  $R_V$ , in order to illustrate the typical range of value that are measured for SN Ia in the literature.

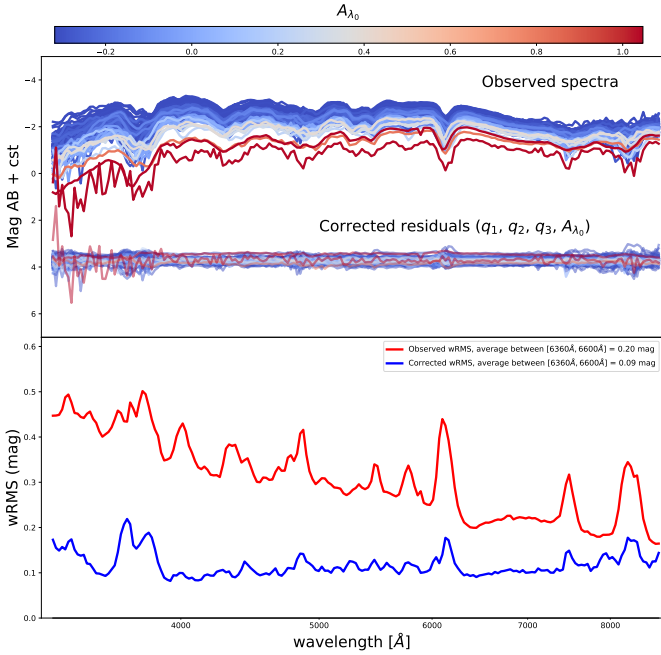
law gives a value of  $R_V = 2.6$ . The Cardelli et al. (1989) law fit coincides remarkably well with  $\gamma(\lambda)$ , except in the UV, where it differs slightly. The  $R_V$  value is mainly driven by the reddest SNe Ia. Indeed, as can be seen in the lower panel of Fig. 6, three SNe Ia have the dominant contribution on the final value of  $R_V$  and removing any of these from the sample would significantly alter the result, hence the rather large value of the uncertainty.

We made the choice of fitting an average extinction curve for the whole sample, but there are indications from observations that extinction curves exhibit some diversity (Amanullah et al. 2015). Moreover, any such extinction variation, or intrinsic color variation (e.g. Foley & Kasen 2011; Polin et al. 2019), that the model at maximum is unable to capture goes into the dispersion matrix. However, in order to keep the model simple, we keep an average value of  $R_V$  and we leave  $R_V$  variation analysis for the future.

Figure 7 presents the set of spectra before any correction and the residuals resulting from the difference between the observed spectrum and Eq. (5). In the top panel, the effect of extinction is clearly visible for spectra before any correction. After correction, the residuals between the models and data are mainly shifted in magnitude with respect to one another and no effect on color appears to remain. This is the signature expected for a mostly gray residual offset.

The amplitude of the correction made by the three factors  $q_1$ ,  $q_2$ ,  $q_3$ , and the extinction curve, given by the wRMS as a function of wavelength, is presented in the lower panel of Fig. 7. Before any correction, the dispersion in the  $B$ -band is around 0.4 mag, as expected. Spectra before any correction exhibit both



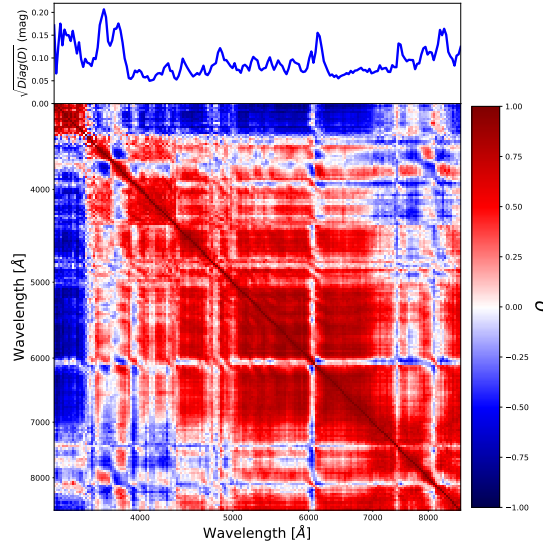


**Fig. 7.** Spectra before and after applying the model at maximum light. *Top panel:* 105 spectra in the absolute AB magnitude before correction (upper set of spectra) and after correction for intrinsic properties and color (lower set of spectra). An arbitrary offset has been applied for each set. The color code indicates the value of  $A_{\lambda_0}$ . *Lower panel:* wRMS of the residuals for the uncorrected (in red) and corrected (in blue) spectra.

localized structure coming from intrinsic variabilities and a slowly increasing dispersion in the bluer part of the spectrum, which is the signature of extinction. Once spectra are corrected by the three factors  $q_1$ ,  $q_2$ ,  $q_3$ , and the extinction curve, the wRMS of the residuals drops to a floor of 0.1 mag. However, localized structures remain, concentrated in UV, Ca II H&K, Si II  $\lambda 6355$ , O I  $\lambda 7773$ , and Ca II IR regions. The amplitude of the baseline is compatible with the value of the gray intrinsic dispersion added in cosmological fits for nearby supernovae (Betoule et al. 2014). The fact that a part of the residual is gray is also visible in the dispersion matrix as can be seen in Fig. 8. Indeed, the high correlation across all wavelengths is consistent with gray fluctuation. However, some effects observed in the dispersion matrix are not due to the gray effect, but to the unmodeled variabilities not captured by SUGAR in certain spectral zones, which explains some features observed in the dispersion matrix and in the Fig. 7.

### 3.4. Time interpolation using Gaussian process

Now, we want to extend our model from maximum light to the full spectral time series. Spectra from the SNfactory are taken at different epochs in the development of each SN Ia. The treatment of these observations taken at different times is an important step to handle before constructing the full SED model. In order to work with a fixed time grid, it is necessary to use a method of interpolation. One solution is to use Gaussian processes, as was done for SNe Ia in works like Kim et al. (2013), Fakhouri et al. (2015), and Saunders et al. (2018). A complete review of Gaussian processes can be found in Rasmussen & Williams (2006). The Gaussian process implementation presented here is based on this review. However, we have developed a specific implementation in order to take into account wavelength dependencies



**Fig. 8.** Square root of the diagonal of the dispersion matrix  $\mathbf{D}$  (top) and the correlation matrix corresponding to the dispersion matrix  $\mathbf{D}$  (bottom). The shading represents the degree of correlation, as given by the Pearson correlation coefficient,  $\rho$ .

and to accelerate computation of the interpolation. These are described below.

A Gaussian process is a generalization of the Gaussian probability distribution, and is a non-parametric way to interpolate data. The main assumption of Gaussian processes is that observed data are the realization of Gaussian random fields that are characterized by an average function and a correlation function. Therefore, the distribution of SNe Ia magnitude  $m$  at a given phase  $t$  and a given wavelength  $\lambda$  follows this relation:

$$m(t, \lambda) \sim \mathcal{N}(m_0(t, \lambda), \mathbf{K}_\lambda), \quad (17)$$

where  $m_0(t, \lambda)$  is the average function of the SNe Ia SED, and  $K_\lambda$  corresponds to the correlation matrix (commonly called the kernel) that describes time correlations between different epochs. In the absence of an explosion model that would yield the analytical form of  $K_\lambda$ , we choose a squared exponential kernel, to which the measurements and calibration uncertainties are added:

$$K_\lambda(t_i, t_j) = \phi_\lambda^2 \exp\left[-\frac{1}{2}\left(\frac{t_i - t_j}{l_\lambda}\right)^2\right] + \sigma_{\text{gray}}^2 \delta'_{ij}, \quad (18)$$

where  $\phi_\lambda^2$  represent the spectral variance around the average function for the wavelength  $\lambda$ ,  $l_\lambda$  corresponds to the temporal correlation length for the wavelength  $\lambda$ ,  $t_i$  and  $t_j$  correspond to the phase of observation  $i$   $j$ , and  $\sigma_{\text{gray}}$  is the gray error taken to be 0.03 mag. All wavelengths are treated independently of each other, that is the interpolation is performed on  $\mathbf{y}_{\lambda n}$ , the  $n$ th SN Ia light curve for each wavelength  $\lambda$  from which the average function has been subtracted. For a given set of global hyperparameters  $(\phi_\lambda, l_\lambda)$ , the interpolation  $\mathbf{y}'_{\lambda n}$  and the covariance matrix of uncertainties on the interpolation,  $\text{cov}(\mathbf{y}'_{\lambda n})$ , are given by:

$$\mathbf{y}'_{\lambda n} = \mathbf{K}_{\lambda n}(\mathbf{t}', \mathbf{t})^T (\mathbf{K}_{\lambda n}(\mathbf{t}, \mathbf{t}) + \sigma_\lambda^2 \mathbf{I})^{-1} \mathbf{y}_{\lambda n} \quad (19)$$

$$\text{cov}(\mathbf{y}'_{\lambda n}) = \mathbf{K}_{\lambda n}(\mathbf{t}', \mathbf{t}') - \mathbf{K}_{\lambda n}(\mathbf{t}', \mathbf{t})^T (\mathbf{K}_{\lambda n}(\mathbf{t}, \mathbf{t}) + \sigma_\lambda^2 \mathbf{I})^{-1} \mathbf{K}_{\lambda n}(\mathbf{t}, \mathbf{t}), \quad (20)$$

where the vector  $\sigma_\lambda^2$  is the error on the magnitude for the wavelength  $\lambda$  and for all observed phases.  $\mathbf{t}'$  is the new time grid and

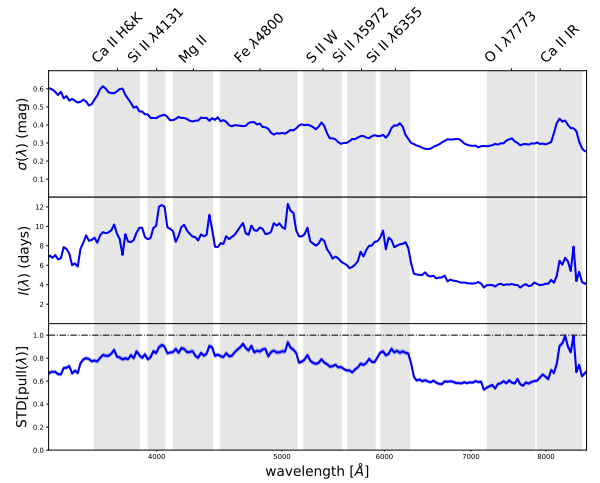
$\mathbf{t}$  the observed phases of the SN Ia. Gaussian processes assume that the observed data are distributed around an average function. This function is a necessary input, because in a region with no data points, the Gaussian process interpolation converges to this average function (after  $\sim 3$  temporal correlation lengths). The calculation of this average function is performed as follows. The training spectra are binned according to phase from maximum light in  $B$ -band, ranging from  $-12$  to  $+48$  days. Each phase bin spans two days and contains at least three spectra. The weighted average spectrum is computed for each phase bin and given as a first approximation of the average function. However, due to some lower signal-to-noise spectra, this version of the average function is not smooth. A Savitsky–Golay filter is used to smooth the average function, which is then used for the Gaussian process interpolation.

What remains is to determine the pair of hyperparameters  $\theta_\lambda = (\phi_\lambda, l_\lambda)$  that represent, respectively, the amplitude of the fluctuation around the average function and the temporal correlation length. To estimate the hyperparameters for each wavelength, we maximize the product of all individual likelihoods:

$$\mathcal{L}_\lambda = \prod_n \frac{1}{2\pi^{\frac{N}{2}}} \frac{1}{|\mathbf{K}_{\lambda n}(\mathbf{t}, \mathbf{t}) + \sigma_\lambda^2 \mathbf{I}|^{\frac{1}{2}}} \times \exp\left(-\frac{1}{2} \mathbf{y}_{\lambda n}^T (\mathbf{K}_{\lambda n}(\mathbf{t}, \mathbf{t}) + \sigma_\lambda^2 \mathbf{I})^{-1} \mathbf{y}_{\lambda n}\right). \quad (21)$$

This differs from previous work for two main reasons: the wavelength dependence of the hyperparameters, and their estimation using all the SNe Ia. These two new features are justified mainly by SNe Ia physics and computation time. Since SNe Ia are standardizable candles, we treat them as realizations of the same Gaussian random field, that is we assume that they share the same set of hyperparameters. In this case, we can use all available SNe Ia to estimate these hyperparameters. The wavelength dependence of the hyperparameters is due to as example dust extinction, intrinsic variability, the second maximum in infrared, etc. It follows that the standard deviation around the average function and the temporal correlation length should vary across wavelength. Moreover, the wavelength dependence allows us to be very efficient in terms of computation time. Indeed, it sped up matrix inversion from the classical  $\mathcal{O}(N_\lambda^3 \times N_t^3)$  to  $\mathcal{O}(N_\lambda \times N_t^3)$ , where  $N_\lambda$  is the number of bin in wavelength (197 here), and  $N_t$  the number of observed phases ( $\sim 14$  in average). On a Mac Book Pro with a 2.9 GHz Intel Core i5 processor and 8 GB of RAM, finding all hyperparameters ( $2 \times 197$ ), and computing the interpolation, and the pull distribution of residuals took under 6 min.

Results of hyperparameter adjustment in terms of wavelength are shown in Fig. 9. First, it can be seen that the structure of hyperparameters as a function of wavelength is not random. Indeed, for the parameter  $\phi_\lambda$ , its value varies from 0.3 mag to more than 0.6 mag: this number is difficult to interpret directly because it is the amplitude around an average value at all phases of the data. Nevertheless, the increase in the ultraviolet and blue wavelengths can be explained by the dispersion caused by dust extinction of the host-galaxies and a larger intrinsic dispersion in this wavelength range (Maguire et al. 2012; Nordin et al. 2018). There are also structures in the peaks of this intrinsic dispersion that correspond to the regions of Ca II H&K, Si II  $\lambda 6355$ , and Ca II IR lines. The correlation length,  $l_\lambda$ , has structure in the silicon and calcium regions and remains globally stable with no features between 6000 Å and 8000 Å (average value of seven days). Once the hyperparameters have been estimated, the interpolations for each SN Ia are performed onto a new phase grid. The



**Fig. 9.** Results of the Gaussian process interpolation. *From top to bottom:* amplitude of the kernel, correlation length and standard deviation of the pull distribution in terms of wavelength. The gray areas represent the presence of the absorption line that were used in our original factor analysis.

phase grid chosen covers a range between  $-12$  and  $+48$  days around maximum brightness in  $B$ -band, divided into three day bins. Finally, to check if both the predictions (interpolation) and associated errors are in agreement with what is expected from a Gaussian process, we compute the pull distribution of the residuals. Because we assume that our data are realizations of a Gaussian random fields, the pull distribution should follow a centered unit normal distribution. For each wavelength we fit a Gaussian to the pull distribution and calculate its standard deviation. The results are shown in the bottom of Fig. 9. The pull standard deviation is on average lower than 1 with an average value of 0.8. These interpolation results are used to establish the SED model that is described in the following sections.

## 4. The SUGAR model

### 4.1. The model

The SUGAR model assumes that SED variation at any epoch may be described by spectral features measured at maximum. The SUGAR SED model is based on a linear combination of factors derived in Sect. 3.2.3 and an extinction curve taken as a Cardelli et al. (1989) law derived in Sect. 3.3. To model the SUGAR SED, we propose a model similar to the one we constructed for maximum light. This SUGAR model is based on the three factors for each SN Ia  $i$ , combined into the vector  $\mathbf{x}_i \equiv \{h_{i,1}, h_{i,2}, h_{i,3}\}$ , which in the model is the true value of the measured factor  $\mathbf{q}_i = \{q_{i,1}, q_{i,2}, q_{i,3}\}$ , and are related to the intrinsic magnitude by:

$$M_i(t, \lambda) = M_0(t, \lambda) + \sum_{j=1}^{j=3} h_{i,j} \alpha_j(t, \lambda) + A_{V_i} f(\lambda, R_V) + \Delta M_{\text{gray } i}, \quad (22)$$

where  $M_0(t, \lambda)$  is the magnitude of the average spectral time series,  $\alpha_j(t, \lambda)$  is the intrinsic variation related to the factor  $h_{i,j}$ ; as discussed in Sect. 3.2.3, the numbers of factors is set to three. As at maximum,  $h_{i,j}$  represents the true value of the measured factor  $q_{i,j}$  derived in Sect. 3.2.  $A_{V_i}$  is the extinction in the  $V$ -band and  $f(\lambda, R_V)$  is the Cardelli et al. (1989) law where  $R_V$  is the extinction ratio that is derived in Sect. 3.3. Finally, the term

$\Delta M_{\text{gray } i}$  is a gray and time independent term comparable to the parameter  $X_0$  of SALT2, which allows us to work independently of the distance. Indeed, the effect of the distance on magnitude is equivalent to an additive constant independent of wavelength. Moreover, the parameter  $\Delta M_{\text{gray } i}$  also contains the average of the spectral time series residuals with the SUGAR model. The addition of  $\Delta M_{\text{gray } i}$  differs from what was done when determining the extinction curve in Sect. 3.3 and this choice is discussed in Sect. 6.1. We model the SED at the same wavelengths and epochs where the Gaussian process interpolation is done. For notational efficiency we rewrite Eq. (22) as:

$$\mathbf{M}_i = \mathbf{A} \mathbf{h}_i, \text{ where} \quad (23)$$

$$\mathbf{A} = (\mathbf{M}_0, \mathbf{f}(R_V), 1, \alpha_1, \alpha_2, \alpha_3), \quad (24)$$

$$\mathbf{h}_i^T = (1, A_{V,i}, \Delta M_{\text{gray } i}, h_{i,1}, h_{i,2}, h_{i,3}). \quad (25)$$

The unknown parameters of the SUGAR model are therefore the average spectrum  $M_0(t, \lambda)$ , the intrinsic coefficients  $\alpha_j(t, \lambda)$ , and the gray offset  $\Delta M_{\text{gray } i}$ . Each SN Ia is parametrized during the training by the  $q_{i,j}$  determined in Sect. 3.2.3 and  $A_{V,i}$  derived in Sect. 3.3. The average spectral time series  $M_0(t, \lambda)$ , the intrinsic coefficients  $\alpha_j(t, \lambda)$ , and the gray offset  $\Delta M_{\text{gray } i}$  are determined in the same way as the modeled described at maximum light in the Sect. 3.3. The main differences with Sect. 3.3 are that the model now depends explicitly on the phases of observation, the extinction curve is fixed at the maximum light value (same  $R_V$ ), and the model is trained on the Gaussian process interpolation of spectra derived in Sect. 3.4 instead of measured spectra. The procedure is described below.

#### 4.2. Fitting the SUGAR model

We use an orthogonal distance regression method to estimate the parameters of the SUGAR model. This is a version of the method described in Sect. 3.3, modified to include the Gaussian process interpolation, the known extinction parameters, and the gray offset. To estimate the SUGAR parameters, we minimize the following  $\chi^2$ :

$$\chi^2 = \sum_i (\mathbf{M}_i^{\text{gp}} - \mathbf{A} \mathbf{h}_i)^T \mathbf{W}_{M_i} (\mathbf{M}_i^{\text{gp}} - \mathbf{A} \mathbf{h}_i) + (\mathbf{q}_i - \mathbf{x}_i)^T \mathbf{W}_{q_i} (\mathbf{q}_i - \mathbf{x}_i), \quad (26)$$

where the vector  $\mathbf{M}_i^{\text{gp}}$  are the Gaussian process interpolations of the observed spectra (Sect. 3.4) which is in the following vector format:

$$\mathbf{M}_i^{\text{gp}} = (\mathbf{y}'_{3340 \text{ \AA } i} \cdots \mathbf{y}'_{\lambda i} \cdots \mathbf{y}'_{8580 \text{ \AA } i}), \quad (27)$$

where  $\mathbf{y}'_{\lambda i}$  is the light curve at wavelength  $\lambda$  interpolated by Gaussian processes onto new time phase as described in Sect. 3.4. As in Sect. 3.3, the uncertainties affecting the  $\mathbf{x}_i$  are propagated through the weight matrix  $\mathbf{W}_{x_i}$ . The  $\mathbf{h}_i$  plays a role analogous to the  $\mathbf{q}_i$  of Sect. 3.2 and are estimated like the other free parameters of the model. Finally, the matrix  $\mathbf{W}_{M_i}$  is the weight matrix of the SN Ia  $i$  and follows this matrix format:

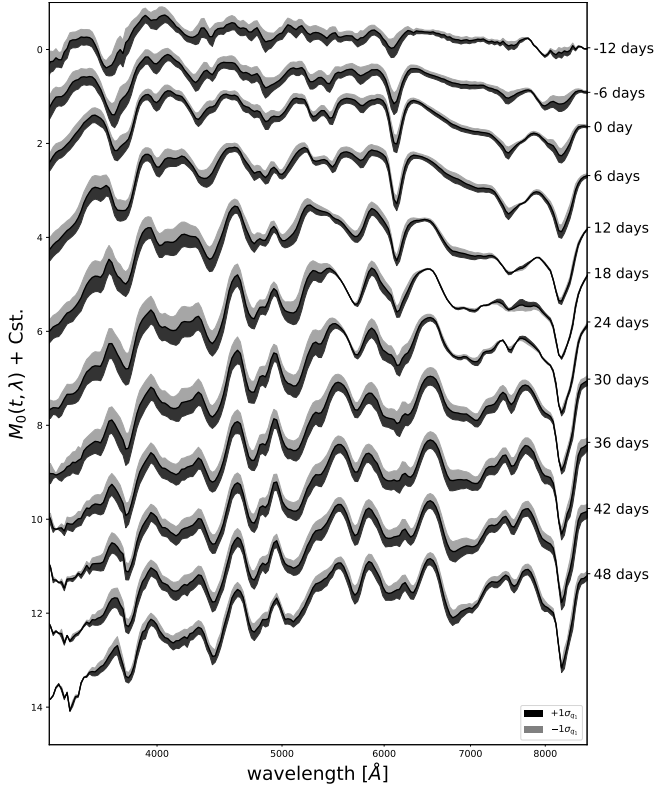
$$\mathbf{W}_{M_i} = \begin{pmatrix} \text{cov}(\mathbf{y}'_{3340 \text{ \AA } i}) & & & & 0 \\ & \ddots & & & \\ & & \text{cov}(\mathbf{y}'_{\lambda i}) & & \\ & & & \ddots & \\ 0 & & & & \text{cov}(\mathbf{y}'_{8580 \text{ \AA } i}) \end{pmatrix}^{-1}, \quad (28)$$

where  $\text{cov}(\mathbf{y}'_{\lambda i})$  is the covariance of interpolation uncertainties from the Gaussian process interpolations determined in Sect. 3.4. The interpolation is done only in terms of the phase relative to  $B$ -band maximum, this makes the weight matrix block diagonal for each SN Ia. In this analysis, we do not add any covariance in terms of wavelength coming from the calibration uncertainties, for reasons of computation time and because they are expected to be small. These block diagonal matrices allowed us to accelerate the minimization algorithm. To estimate the 16 653 free parameters of the SUGAR model, it took just less than 1 h on a Mac Book Pro with a 2.9 GHz Intel Core i5 processor and 8 GB of RAM, thanks to sparse linear algebra. We note that while the covariance terms were neglected, the uncertainty at each wavelength was taken into account during the training of the Gaussian processes (these are the blocks  $\text{cov}(\mathbf{y}'_{\lambda i})$ ). The unknown parameters of the model (the matrix  $\mathbf{A}$  and vectors  $\mathbf{h}_i$ ) are computed by minimizing Eq. (26) using an expectation and minimization algorithm. This algorithm, described in Appendix C, avoids degeneracies between parameters.

#### 4.3. The SUGAR components

The SUGAR model is described by a set of spectral time series components ( $M_0, \alpha_1, \alpha_2, \alpha_3$ ) at three day intervals. The SED of a given SN Ia is then obtained by a linear combination of those components employing the  $(q_1, q_2, q_3)$  factors describing the supernova. For illustrative purposes, we present a spectral time series in six day intervals as well as their integration in synthetic top-hat filter system comprised of five bands with the following wavelength ranges:  $\hat{U}$  [3300.0–3978.0]  $\text{\AA}$ ;  $\hat{B}$  [3978.0–4795.3]  $\text{\AA}$ ;  $\hat{V}$  [4795.3–5780.6]  $\text{\AA}$ ;  $\hat{R}$  [5780.6–6968.3]  $\text{\AA}$ ;  $\hat{I}$  [6968.3–8400.0]  $\text{\AA}$ . The diacritic hat serves as a reminder that these are not standard Johnson-Cousins filters. The effect of the vectors  $\alpha_j$  on the average spectrum  $M_0$ , on the average light curves (obtained by integration of  $M_0$  in the  $\hat{U}$ ,  $\hat{B}$ ,  $\hat{V}$ ,  $\hat{R}$ ,  $\hat{I}$  bands) and on the average colors  $\hat{U} - \hat{B}$ ,  $\hat{B} - \hat{V}$ ,  $\hat{V} - \hat{R}$ ,  $\hat{R} - \hat{I}$  is presented in Figs. 10 and 11 for  $\alpha_1$ , Figs. 12 and 13 for  $\alpha_2$ , and Figs. 14 and 15 for  $\alpha_3$ . The upper and lower contours correspond to a variation of  $\pm 1\sigma$  of the associated  $q_j$  parameter. The  $\alpha_j(t = 0)$  components obtained when training the color law with only data at maximum or with the full time series are almost identical: all conclusions from Sect. 3.3.3 remain valid and we focus here on the temporal behavior.

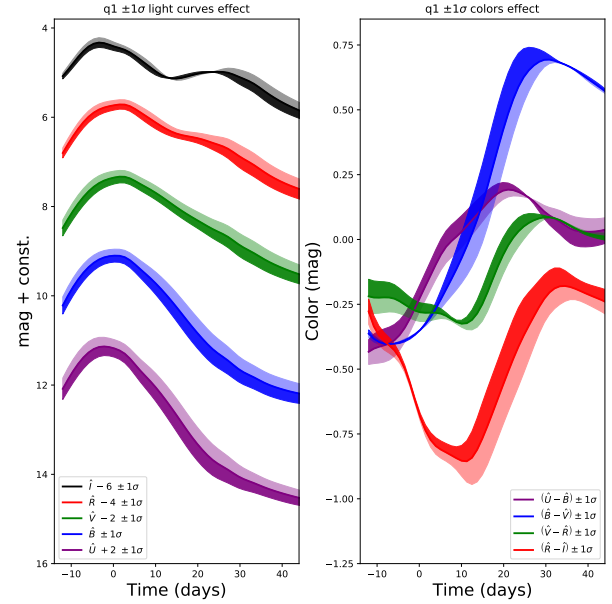
Of the three factors,  $q_1$  has the strongest impact: this can be seen from the time series presented in Fig. 10 and on the broad band light curves presented in Fig. 11. This is associated with a stretch effect, visible in  $\hat{U}$ ,  $\hat{B}$ , and  $\hat{V}$  bands as an enlargement of the variation band as we move away from maximum. Between  $-12$  days and  $+30$  days, the  $q_1$  factor has a strong influence on spectroscopic details. The brightest SNe Ia exhibit shallower troughs in their spectral features: this is especially visible for Si II  $\lambda 4131$ , Si II  $\lambda 5972$ , and Ca II IR, but also valid for most of the lines. This effect is more pronounced for SNe Ia at earlier phases, and fades over time. After  $+30$  days, the factor  $q_1$  is less sensitive to localized features, and shows a relative enhancement of the optical  $\hat{B}$  to  $\hat{R}$  bands with respect to  $\hat{U}$  and  $\hat{I}$ . Interestingly, between  $+12$  days and  $+24$  days, the brightest SNe Ia are bluer than at maximum or at later phases. This is true in absolute value, but also relative to the average SN Ia. This shows that the SNe Ia color space is driven by a variable intrinsic component in addition to reddening by dust. This intrinsic color variation is linked with the position of the second peak in  $\hat{I}$ , which appears later for brighter SNe Ia.



**Fig. 10.** Average spectral time series and the effect of a variation of  $q_1$  by  $\pm 1\sigma$ . Each phase is separated by a constant magnitude offset.

As at maximum, the effect of the  $q_2$  factor for the whole time series (Fig. 12) is mainly localized around specific spectral features, with little impact on broad band light curves (Fig. 13), with the exception of  $\hat{I}$ -band. This factor associates higher line velocities around maximum with deeper absorption troughs, strongly visible before maximum and up to +18 days in the Ca II H&K region and at all phases in the Ca II IR region. The net effect described by  $q_2$  is that higher velocities are associated with slightly dimmer SNe Ia. At later phases velocity effects are still observable and at all phases after  $-6$  days, high-velocity SNe Ia have localized differences in the  $4700 - 5100 \text{ \AA}$  spectral region. In addition, they are bluer in  $\hat{V}$ ,  $\hat{R}$ , and  $\hat{I}$  after maximum, with a maximal effect on the  $\hat{I}$ -band at around +12 days. A slight stretch effect is visible in  $\hat{U}$ , and  $\hat{B}$ , and also results in a later phase for the second maximum in  $\hat{I}$  for the brightest SNe Ia in  $\hat{B}$ , as was mentioned for  $q_1$ . The shape of variations of the light curve in  $\hat{I}$  are very different for each factor, indicating that this band can help reconstruct the  $q_2$  factor when using photometric data only. Furthermore, while the  $q_2$  factor has a small impact on the variations of individual light curves, it has a sizable influence on colors, especially  $\hat{R} - \hat{I}$ ,  $\hat{U} - \hat{B}$ , and after +25 days, on  $\hat{B} - \hat{V}$ . In addition to the  $\hat{I}$ -band, the different color variation pattern in  $\hat{B} - \hat{V}$  at late phases offers a way of disentangling the  $q_1$  and  $q_2$  factors when dealing with photometric data.

As seen in Fig. 14, the influence of  $q_3$  is minimal around maximum light, similar to the results from Sect. 3.3. It grows at other phases and appears as a stretch effect on the light curves presented in Fig. 15. Unlike  $q_1$ , the stretch described by  $q_3$  shows little correlation with the magnitude at maximum. While  $q_3$  has a larger impact on the light curves than  $q_2$ , the reverse is true for individual colors. At late phases,  $q_3$  exhibits a brighter-redder correlation that might be employed to distinguish this vector from  $q_1$  when using photometric data only.



**Fig. 11.** Average light curves and color evolution with phase in synthetic bands, and the predicted effect of a variation of  $q_1$  by  $\pm 1\sigma$ .

The influence of  $q_3$  on spectral structures is mainly visible in calcium regions, although variations can also be observed in other regions. Around Ca II IR, between  $-6$  days and maximum, brighter SNe Ia exhibit a deeper absorption trough at high calcium velocity. This high-velocity calcium feature is visible for all  $q_3$  values at  $-12$  days, but fades faster for dimmer SNe Ia. Meanwhile the lower velocity counterpart is only visible in  $q_3$  for the dimmest supernovae at  $-12$  days and appears later for brighter SN Ia.

## 5. SUGAR performance

### 5.1. Fitting spectra with SUGAR

The measurement of spectral indicators is very sensitive to the signal-to-noise ratio of the spectrum, and at high redshift, measuring the indicators in the observer's reference frame is not possible. However, once the model has been trained, the parameters  $q_1$ ,  $q_2$ ,  $q_3$ ,  $A_V$ , and  $\Delta M_{\text{gray}}$  for a given SN Ia can be directly estimated from its spectral time series. Estimating the parameters  $q_1$ ,  $q_2$ ,  $q_3$ ,  $A_V$ , and  $\Delta M_{\text{gray}}$  is done directly on spectra by minimizing the following  $\chi_i^2$ :

$$\chi_i^2 = (\mathbf{M}'_i - \mathbf{A}'\mathbf{x}_i)^T \mathbf{W}_{\mathbf{M}'_i} (\mathbf{M}'_i - \mathbf{A}'\mathbf{x}_i), \quad (29)$$

where  $\mathbf{M}'_i$  is the vector containing the entire spectral time series of SN Ia  $i$ , ordered in the same way as for the SUGAR training:

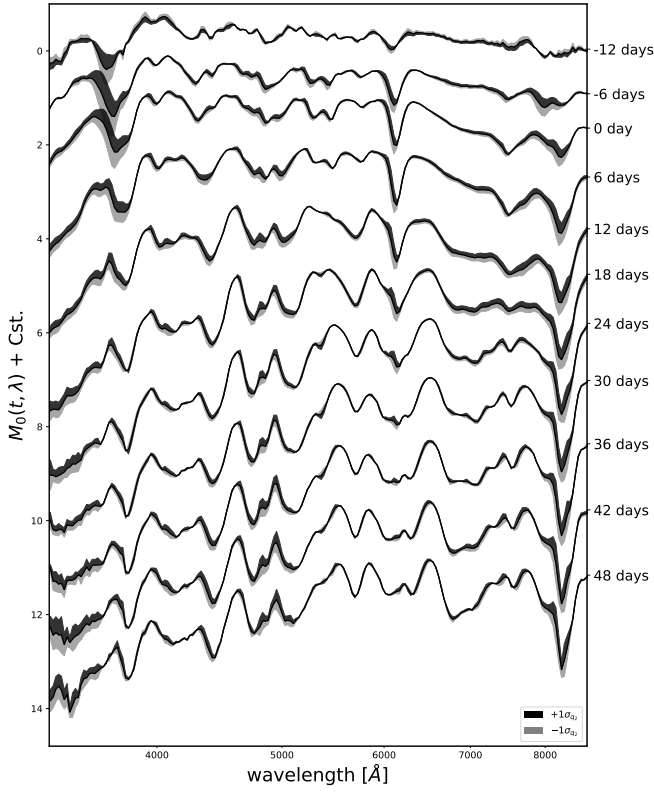
$$\mathbf{M}'_i = (\mathbf{y}_{3340 \text{ \AA } i} \cdots \mathbf{y}_{\lambda i} \cdots \mathbf{y}_{8580 \text{ \AA } i}). \quad (30)$$

where  $\mathbf{y}_{\lambda i}$  corresponds to the light curve at wavelength  $\lambda$  for the observed phases. The model is projected onto observation phases using cubic spline interpolation. For notation efficiency, it is thus ordered as:

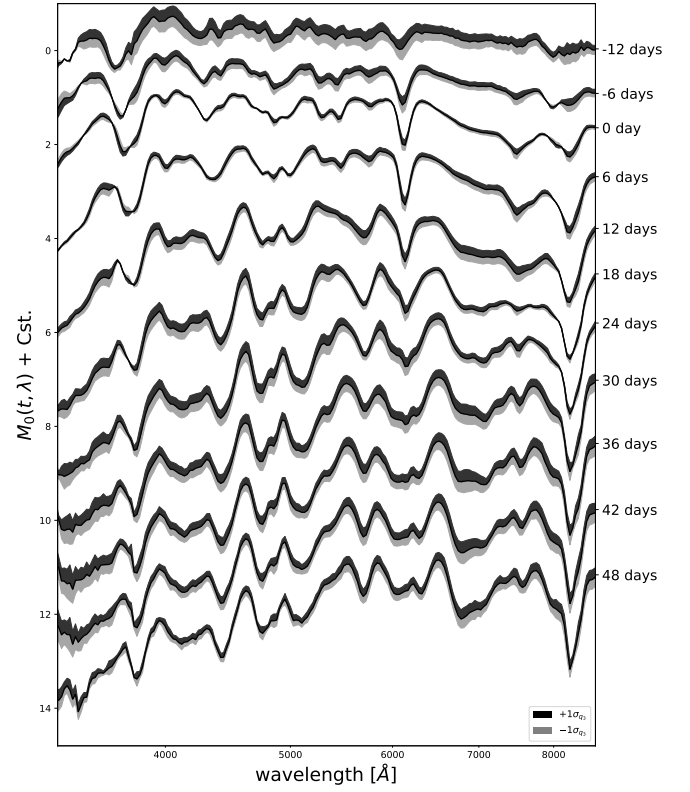
$$\mathbf{A}' = (\mathbf{M}'_0, 1, f(\lambda, R_V), \alpha_{1'}, \alpha_{2'}, \alpha_{3'}). \quad (31)$$

The vector  $\mathbf{x}_i$  contains the parameters that describe the SN Ia  $i$  according to the SUGAR model:

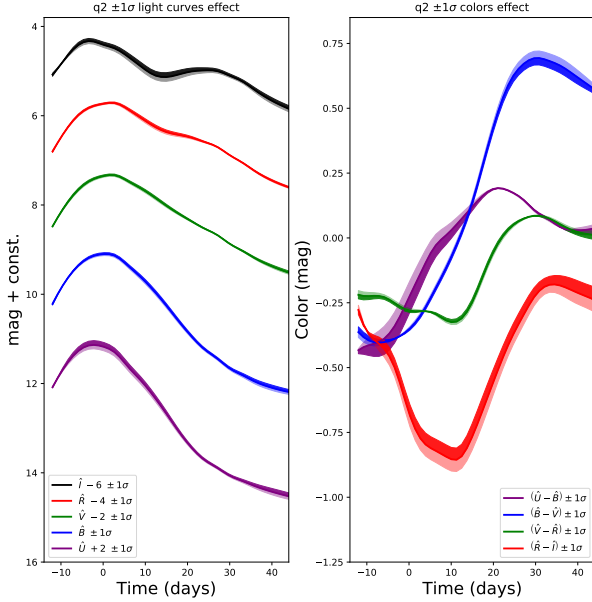
$$\mathbf{x}_i^T = (1, \Delta M_{\text{gray } i}, A_V i, q_{i,1}, q_{i,2}, q_{i,3}). \quad (32)$$



**Fig. 12.** Average spectral time series and the effect of a variation of  $q_2$  by  $\pm 1\sigma$ . Each phase is separated by a constant magnitude offset.



**Fig. 14.** Average spectral time series and the effect of a variation of  $q_3$  by  $\pm 1\sigma$ . Each phase is separated by a constant magnitude offset.



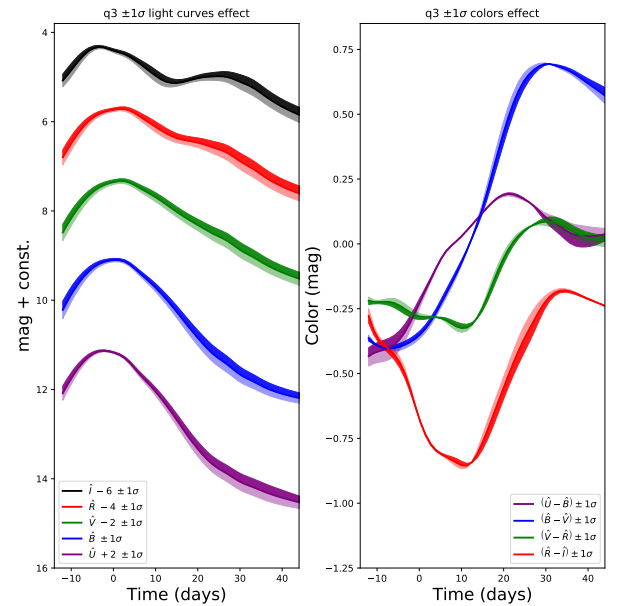
**Fig. 13.** Average light curves and color evolution with phase in synthetic bands, and the predicted effect of a variation of  $q_2$  by  $\pm 1\sigma$ .

Finally,  $\mathbf{W}_{M'_i}$  is the weight matrix that comes from the  $M'_i$  errors. Estimating  $\mathbf{x}_i$  amounts to minimizing the  $\chi^2_i$  with respect to  $\mathbf{x}_i$ , giving the solution:

$$\mathbf{x}_i = (\mathbf{A}'^T \mathbf{W}_{M'_i} \mathbf{A}')^{-1} (\mathbf{A}'^T \mathbf{W}_{M'_i} M'_i) \quad (33)$$

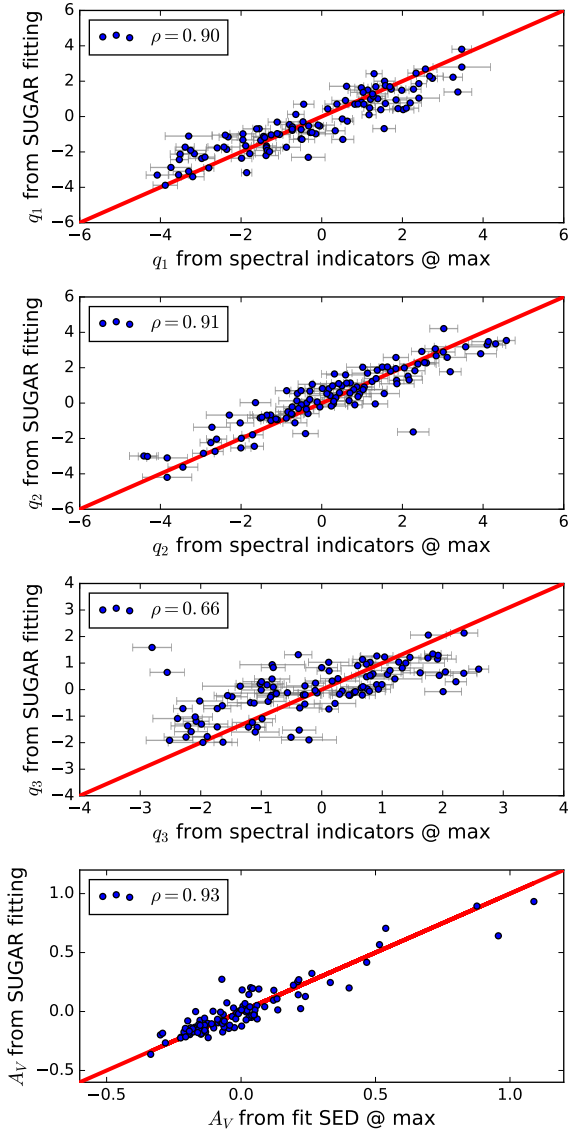
Therefore the covariance on the  $\mathbf{x}_i$  is given by:

$$\text{cov}(\mathbf{x}_i) = (\mathbf{A}'^T \mathbf{W}_{M'_i} \mathbf{A}')^{-1}. \quad (34)$$



**Fig. 15.** Average light curves and color evolution with phase in synthetic bands, and the predicted effect of a variation of  $q_3$  by  $\pm 1\sigma$ .

The term  $\mathbf{x}_i$  is computed for the 105 SNe Ia that were used to train SUGAR. The values obtained are compared with the factors and  $A_V$  measured at maximum light in Fig. 16. Both quantities are highly correlated. However, this correlation decreases with decreasing order of the factor index. This is because the factor uncertainties increase with index due to the noise being greater for higher-order components. The increase in noise therefore reduces the correlation. The extinction term is quite



**Fig. 16.** Factors estimated directly from the spectral indicators, as a function of the factors estimated only from the SUGAR model and the  $A_V$  estimated at maximum as a function of those estimated with SUGAR. The red lines represent the equation  $x = y$ .

compatible between its maximum estimate and its estimation with SUGAR.

The term  $x_i$  is also computed for the 58 SNe Ia kept for validation. The distribution of the  $x_i$  parameters from the validation sample is compared to the  $x_i$  parameters from the training in Fig. 17. From this, one can see that the training and validation samples are compatible, as it is confirmed in following sections.

## 5.2. Comparison with SALT2

### 5.2.1. Difference between SUGAR and SALT2 models

Throughout this article, we use the SALT2 model as a benchmark to evaluate the properties of the SUGAR model, because it has become a reference for cosmological analyses. SALT2 describes the SED of SNe Ia in the following functional form (Guy et al. 2007):

$$F(t, \lambda) = X_0[F_0(t, \lambda) + X_1F_1(t, \lambda)] \exp[C C_L(\lambda)], \quad (35)$$

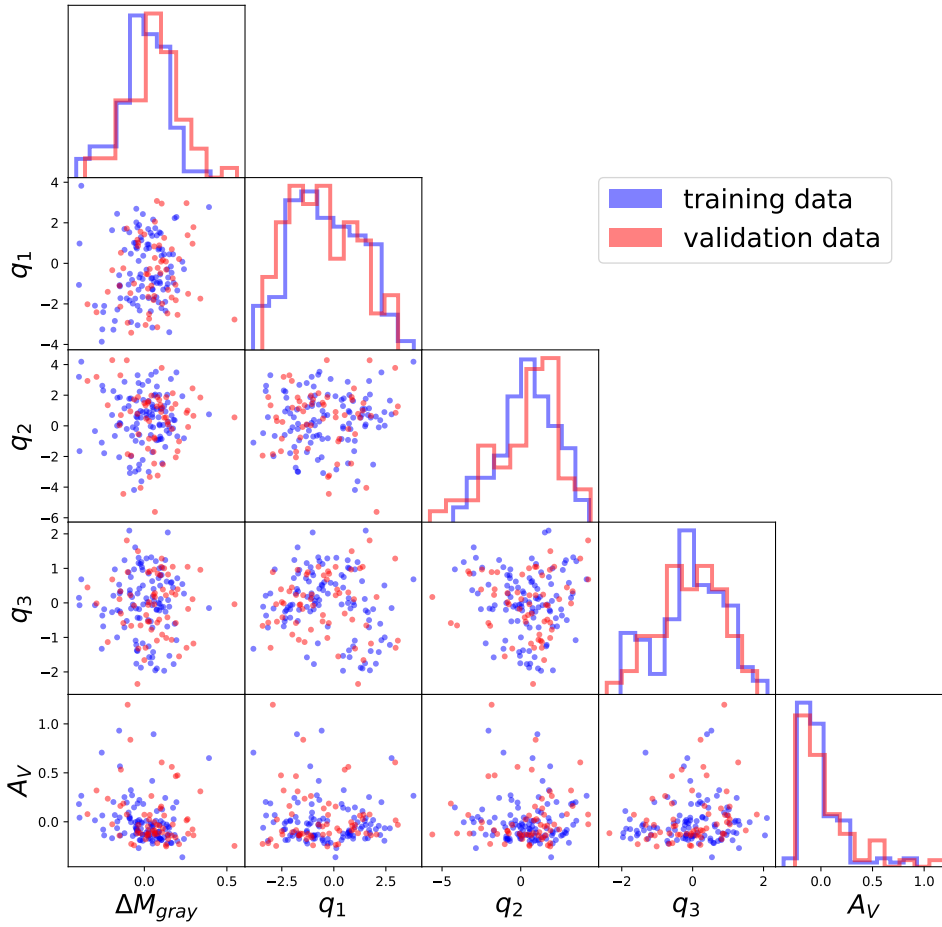
where the observed flux  $F$  depends on the individual SN Ia parameters ( $X_0, X_1, C$ ) and on the model average flux  $F_0$  (defined up to a multiplicative factor), the model flux variation  $F_1$ , and an empirical color law  $C_L$ . In the regime where  $X_1F_1$  is small with respect to  $F_0$ , this model, once expressed in magnitudes, can be identified with the SUGAR model: ( $F_0, F_1, C_L$ ) and ( $X_1, C$ ) take the roles of ( $M_0, \alpha_1, f(R_V)$ ), and ( $q_1, A_V$ ), respectively.  $\Delta M_{\text{gray}}$  is the equivalent of a linear combination of ( $\log X_0, X_1, C$ ).

Although the major differences between the models come from the inclusion of two additional intrinsic components in SUGAR, its color law, and the data on which they were trained, there are many other differing characteristics, which are summarized Table 1. The inclusion of two additional intrinsic components is enabled by the level of detail available in our spectrophotometric time series. We therefore expect the SUGAR model to provide a more faithful representation of the spectral details than SALT2. Though analogous in spirit,  $X_1$  and  $q_1$  are derived employing a quite different paradigm:  $F_1$  comes from a PCA analysis performed in flux space, while the  $\alpha_i$  are trained in magnitude space and driven to reproduce the prominent spectral indicators. A direct generalization of the SALT2 approach with more components is presented in Saunders et al. (2018). While the empirical SALT2  $C_L$  function models the average color variation, our color curve was shown to be accurately described by a Cardelli et al. (1989) extinction law. The SALT2 color parameter  $c$  thus mixes information about reddening by interstellar media (described by  $A_V$ ) with intrinsic SNe Ia properties that were not directly matched with one of the  $q_i$  parameters. These differences in the treatment of colors would manifest in long-range spectral variation (because the extinction curve is smooth). In regions where the phase coverage is sufficient, the different phase interpolation methods used are not expected to contribute significantly to differences between models. Otherwise, and for early phases in particular, we expect the differences to be driven by the training samples rather than interpolation technique.

The wavelength and phase coverage of SUGAR are restricted compared to SALT2. This is directly related to the training data: SUGAR is trained on low redshift spectrophotometric time series, while SALT2 is trained on a wide redshift range of photometric data, employing a handful of spectra to help refine the SED details. As SALT2 uses low and high redshift data it offers a better coverage of the UV domain. It also has a larger phase coverage thanks to the rolling cadence strategy, while SNfactory SNe Ia have to be specifically targeted. While SALT2 could be easily trained with SNfactory data added to their sample, the reverse is not true: major adaptation would be needed to allow SUGAR to incorporate photometric data in its training. However, this does not prevent fitting photometric light curves with SUGAR in the same way as with SALT2, provided they fall within the phase and spectral coverage of SUGAR. In contrast, SALT2 was not designed to fit spectrophotometric data.

### 5.2.2. Spectral residuals

The SUGAR model is designed to improve the spectral description of SNe Ia by adding components beyond stretch and color. While the ideal model would perfectly match the data, the level of mismatch can be quantified by the residuals, that is the difference between the left and right term of Eq. (22). The improvement with respect to SALT2 is then studied by comparing the data to the spectral reconstruction predicted by both models. One example of these comparisons is presented Fig. 18, which shows the spectra and residuals for SN-training-4. For this supernova, SUGAR gives a better description than SALT2. The



**Fig. 17.** SUGAR parameters fit directly from spectral data for the 105 SNe Ia used to train SUGAR (blue) and for the 58 SNe Ia from the validation sample (red).

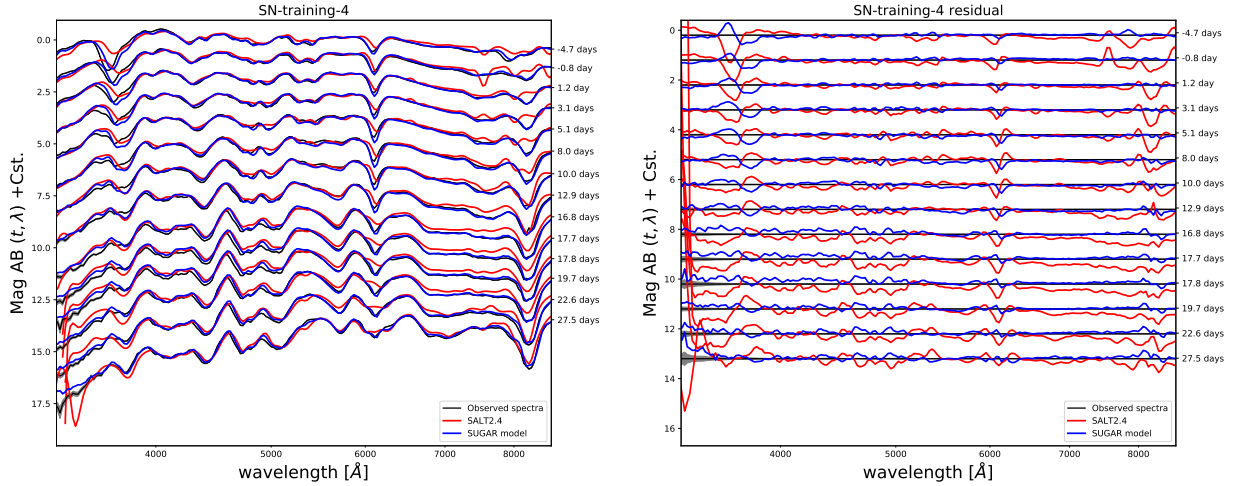
**Table 1.** Main differences between the SALT2.4 model and the SUGAR model.

Model properties	SALT2	SUGAR
Rest frame wavelength coverage	2000–9200 Å	3254–8649 Å
Rest frame phase coverage	–20 to +50 days	–12 to +48 days
Training data	Spectroscopy and photometry	Spectrophotometric time series
Redshift range	Low and high redshift	Low redshift
Interpolation	Cubic spline	Gaussian process with squared exponential kernel
Intrinsic properties from	Weighted PCA like on SED	EM-FA on spectral features
Number of intrinsic component	1	3
Color	Color law (third order polynomial)	Cardelli et al. (1989) law
Error on the model	Yes	No

improvement is especially clear in the UV after +17 days and in the infrared at any phase, and particularly in the O I  $\lambda 7773$  and Ca II IR regions. This is not surprising insofar as SALT2 is essentially trained to reproduce the bluer part of the spectrum. The SUGAR description is also better in the regions around lines, even if this description is not totally satisfactory in some areas: Ca II H&K, Si II  $\lambda 6355$ , and the bluer part of the spectrum at late phases. In the  $B$ -band, the reconstruction of the spectral details is quite similar between SUGAR and SALT2, although SUGAR seems to follow more faithfully the spectral details, as can be seen in the Mg II area.

SN-training-4 is only one example among the 105 SNe Ia on which SUGAR was trained. To have a statistical description of the precision of the models, we derive the dispersion of the residuals obtained for all SNe Ia as a function of wavelength and

phase. In Fig. 19, we present the wRMS of this dispersion: for each wavelength, we compute the wRMS across all phases and weight it by the spectral variance. To show how the accuracy of the model evolves when going from a classic Tripp (1998) relation (when only  $q_1$  and  $A_V$  are included) to the full model, the wRMS for SUGAR is computed three times with progressive inclusion of  $q_1$ ,  $q_2$ , and  $q_3$ . We compute the residual dispersion for both the training and the validation sample. What was observed for SN-training-4 is confirmed in this statistical analysis across all SNe Ia: at any wavelength, SUGAR gives a better description than SALT2. The strongest improvement is in the IR, but is also significant in the UV. All the spectral details are significantly improved, such as the Ca II H&K, Si II  $\lambda 6355$ , the O I  $\lambda 7773$ , and the Ca II IR. The best performance of the SUGAR model is obtained in the region around 6500 Å, where



**Fig. 18.** SUGAR and SALT2 fit of SN-training-4. *Left:* observed spectral series of SN-training-4 (in black), compared to SALT2 (in red) and SUGAR (blue). *Right:* residuals of SALT2 (red) and SUGAR (blue) models to the observations. The black line represents the zero value of the residuals.

the dispersion gets as low as 0.05 mag. It should be kept in mind that the spectral dispersion includes noise variance that may be high in some spectra and is therefore not representative of the dispersion expected in photometry after integration over broad band filters. The phase evolution of the wRMS is presented in Fig. 20: the dispersion in each phase bin is calculated across all wavelengths. This includes wavelengths for which SALT2 is not adapted: this explains relatively high values of wRMS for this model compared to SUGAR.

The successive inclusion of  $q_1$ ,  $q_2$ , and  $q_3$  can help us understand the origin of SUGAR's improvement over SALT2. As the training samples for both models are different, one may wonder what SALT2 would have given if trained on SNfactory data. This situation can be studied when including only  $q_1$ : SUGAR $_{|q_1}$  is slightly worse than SALT2 between Si II  $\lambda 4131$  and Si II  $\lambda 6355$  features, and between Si II  $\lambda 6355$  and O I  $\lambda 7773$ . These regions cover the wavelength range where SALT2 was trained and optimized, while SUGAR $_{|q_1}$  is only the by-product of the three component model and is not by itself optimal for a single component. Earlier phase are more difficult to reproduce in both models: this could be an effect of higher variability beyond stretch at those phases. For later phases, SUGAR $_{|q_1}$  does not exhibit the same increase in wRMS as SALT2. This is also an effect of the lack of coverage for IR wavelength for SALT2. One could therefore expect SALT2 to behave at least as well as SUGAR $_{|q_1}$  in UV and IR when trained on SNfactory data. As expected, the inclusion of additional components improves the accuracy of SUGAR:  $q_2$  improves the description of the spectral features because it is correlated with (Ca II H&K, Si II  $\lambda 6355$ , and Ca II IR) while one has to wait for the inclusion of  $q_3$  to reach the full SUGAR precision.

The improvement brought by SUGAR is thus enabled by two factors: the spectral coverage and the spectral resolution that allows the inclusion of the new components,  $q_2$ , which improves the color description, and  $q_3$ , which helps reduce the dispersion away from maximum light. Moreover, the improvement added by  $q_2$  and  $q_3$  are also seen in the validation sample, which confirms that these components are not due to overtraining.

### 5.2.3. Relation between SALT2 and SUGAR parameters

In Sect. 3.2 we presented the correlations between the SALT2 parameters and the first five factors found using factor analysis.

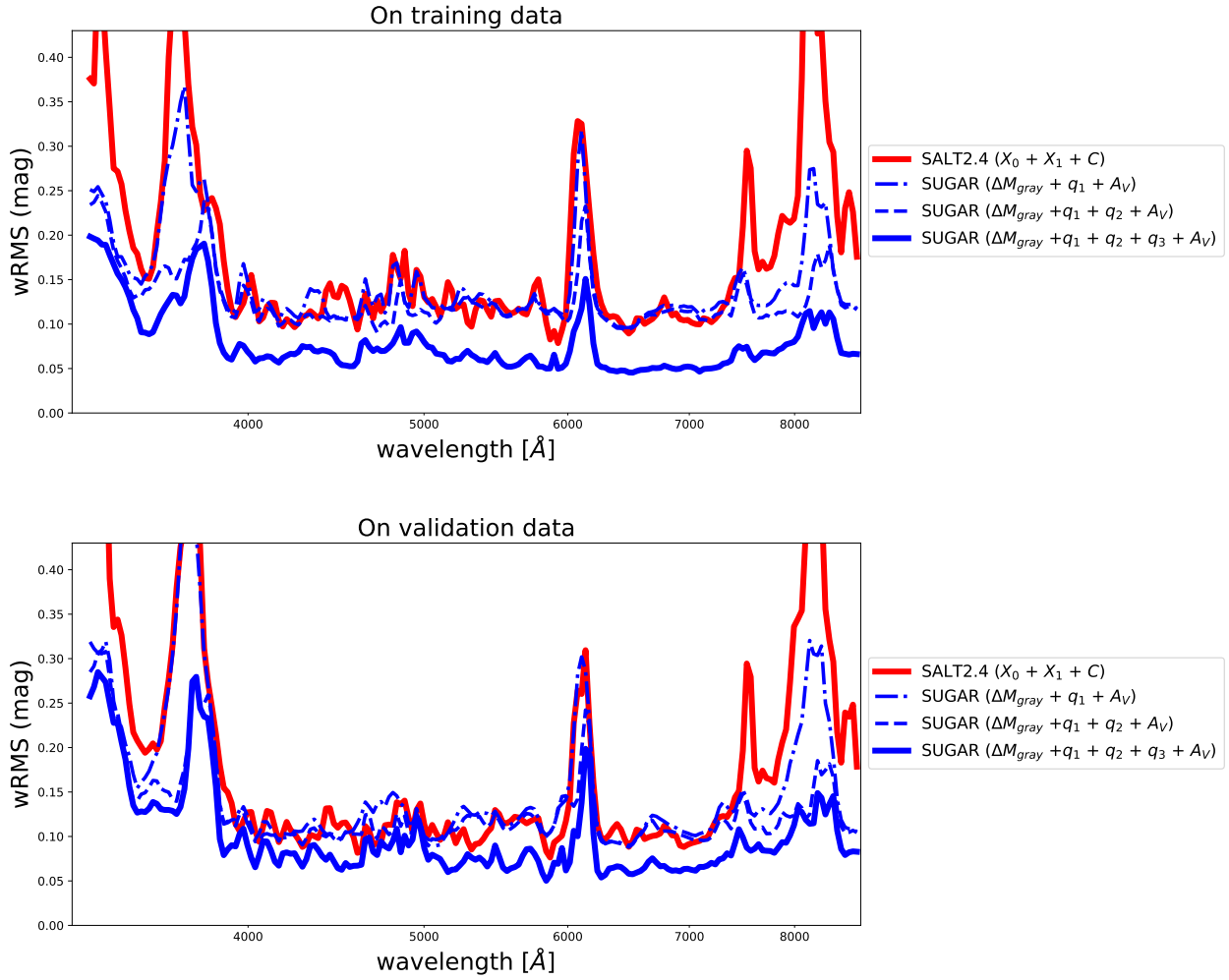
After training the full SUGAR model and reconstructing the parameters of individual supernovae as described in Sect. 5.1, we can study the correlations between the SALT2 parameters ( $X_0$ ,  $X_1$ ,  $C$ ) and the SUGAR parameters ( $\Delta M_{\text{gray}}$ ,  $q_1$ ,  $q_2$ ,  $q_3$ ,  $A_V$ ). The direct comparison of  $X_0$  and  $\Delta M_{\text{gray}}$  is irrelevant due to the different prescription for lifting the degeneracies of both models. We recognize  $\Delta M_{\text{gray}}$  as the Hubble diagram residuals, a quantity that we can also compute from SALT2 parameters (and denote as  $\Delta\mu^{\text{corr}}$ ). The results of this comparison between SALT2 and SUGAR parameters are presented in Fig. 21. The lessons from the correlations of SALT2 parameters and the original factors still hold:  $X_1$  is strongly correlated with  $q_1$  and  $q_3$ ;  $q_2$  has no significant correlation with any of SALT2 parameters, except with  $\Delta\mu$ . The small correlation of  $q_2$  and  $q_3$  with  $\Delta\mu$  shows that the inclusion of this factor can help improve the standardization. The SALT2 color  $C$  is strongly correlated with  $A_V$ , as would be expected since the purpose of these two parameters is to take into account the reddening of SNe Ia. Contrary to the results from Sect. 3.2 with the factors alone,  $C$  now exhibits a  $4.5\sigma$  correlation with  $q_3$ . This correlation might come from the slight fluctuation in color with respect to phase visible in Fig. 15 and the inclusion of the validation sample. The slight correlation of  $C$  and  $\Delta M_{\text{gray}}$  comes from the correlation of  $\Delta M_{\text{gray}}$  and  $A_V$ . Indeed,  $\Delta M_{\text{gray}}$  and  $A_V$  were determined separately during the training, without fixing potential degeneracy between them. This explains why they are correlated. Finally,  $\Delta\mu$  has a strong correlation with  $\Delta M_{\text{gray}}$ . This indicates that the inclusion of the additional factors in SUGAR is unable to catch the major source of magnitude variability of SNe Ia, even if it provides a much improved description of spectral details.

## 6. Discussion

### 6.1. Gray dispersion and dispersion matrix fitting

For the model at maximum light, we did not fit a gray term but did include a dispersion matrix while estimating the extinction curve. However, for the full spectral time series, we fit for a gray offset and used the extinction curve determined at maximum light. One might wonder why we did not do a simultaneous fit of the full spectral time series, the gray offset and the extinction curve. Our decision was motivated by the following logic.





**Fig. 19.** wRMS of the residuals as a function of the wavelength for the full SUGAR model (blue line), for the SUGAR model only corrected by  $q_1$  and  $A_V$  (blue dashed dotted line), for the SUGAR model only correct by  $q_1$ ,  $q_2$  and  $A_V$  (blue dashed lines), and for the SALT2 model (red line). *Top:* training data. *Bottom:* validation data.

In order to perform a simultaneous fit, one important thing is to have a good estimate of the dispersion matrix, which is essential for the estimation of the extinction curve, as discussed in Scolnic et al. (2014). In addition, this matrix would have the advantage of taking into account temporal correlations of residuals, in addition to the color correlations of the residuals, which would improve the model developed at maximum light.

However, from a numerical point of view, this would break the sparse algebra approximations that significantly speed up the training. Indeed, the sparse algebra (due to Eq. (28)) makes it possible to improve speed efficiency of minimization of Eq. (26) from  $\mathcal{O}((N_\lambda N_t N_{\text{param}})^3)$  to  $\mathcal{O}(N_\lambda \times (N_t N_{\text{param}})^3)$ , where  $N_\lambda$  is the number of wavelength bin,  $N_t$  the number of bin in terms of time, and  $N_{\text{param}}$  is the number of parameters to fit for a given wavelength and epoch. In SUGAR training,  $N_\lambda \sim 190$ ,  $N_t \sim 30$ , and  $N_{\text{param}} = 5$  (average spectrum + three factors + extinction curve), consequently, not doing sparse algebra significantly slow down the speed of the training algorithm by  $\mathcal{O}(N_\lambda^2)$ .

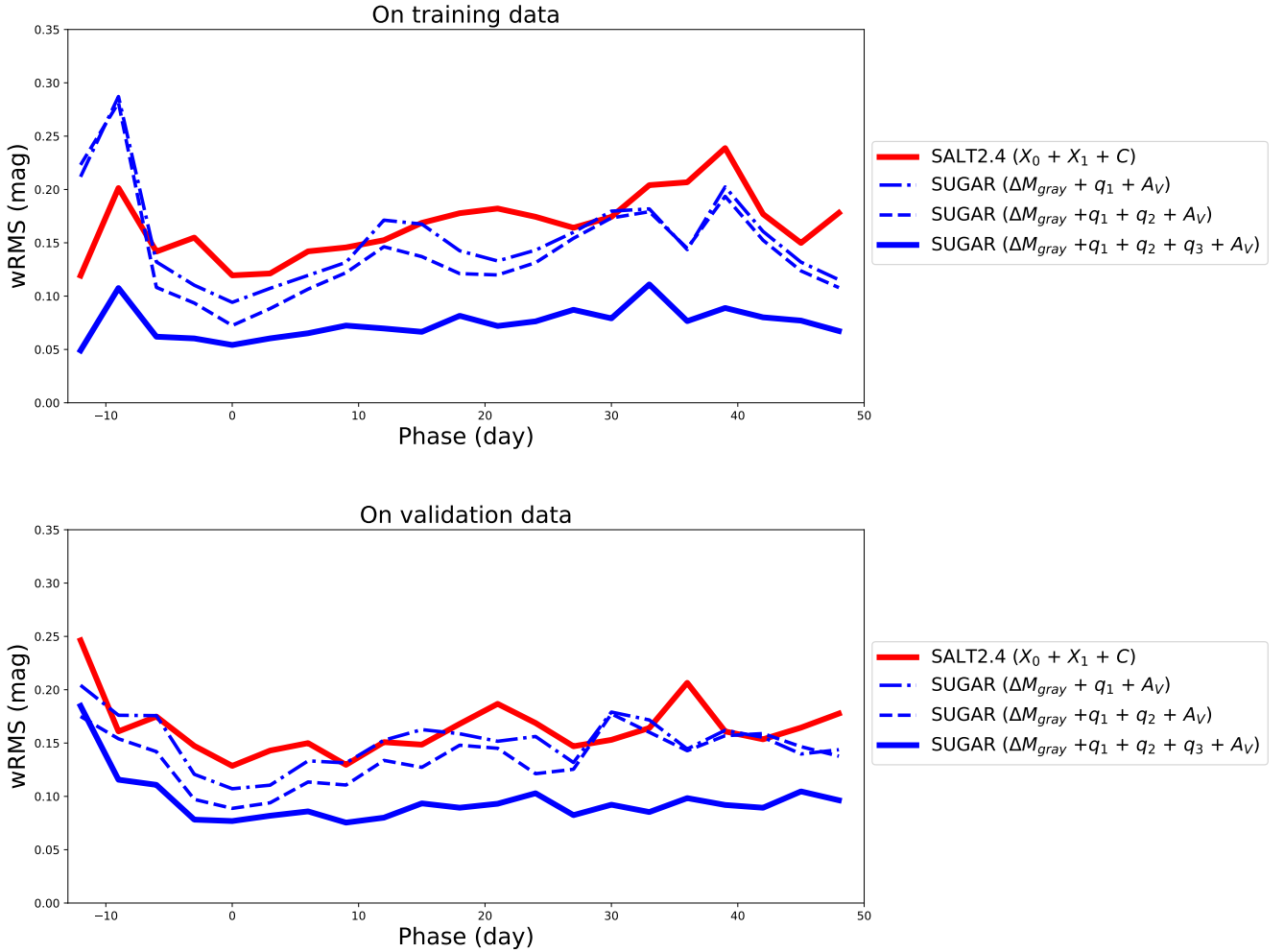
Moreover, there would be problems with numerical stability for the estimation of the dispersion matrix due to its size; in the case of the simultaneous fit: it would be  $\sim 20\,000 \times 20\,000$ . Indeed, it would be necessary to make sure that the dispersion matrix is positive definite, which would involve doing a Singular Value Decomposition that would again slow down the speed

of the training algorithm. In addition, it is evident that the gray offset is degenerate with the dispersion matrix because the dispersion matrix can fully capture a gray dispersion. Therefore we deliberately did not include it explicitly in the fit of the extinction curve within SUGAR in Sect. 3.3. There is however a degeneracy between  $A_{\lambda_0}$  and the gray offset, and part of the latter is captured by this parameter. The SUGAR model presented here is already a significant improvement over the SALT2 model and provides insights into understanding SN Ia variability.

## 6.2. Test of adding an additional component

In the Sect. 3.2, we discussed the number of factors needed to describe the final SUGAR SED, and we concluded that this cannot be determined only from the factor description of spectral features at maximum light. Indeed, even if they are strongly related, the main goal is to know the number of components needed to describe the full SED and not the number of components needed to describe the spectral features space at maximum light. One way to check if the choice of three factors used here is optimal is to retrain the SUGAR model with more than three components and observe how the spectral residuals evolve with this change.

In the following, we ran the training of SUGAR twice, each time adding an additional component, that is we reproduced



**Fig. 20.** wRMS of the residuals as a function of phase for the full SUGAR model (blue line), for the SUGAR model only corrected by  $q_1$  and  $A_V$  (blue dashed dotted line), for the SUGAR model only correct by  $q_1$ ,  $q_2$  and  $A_V$  (blue dashed lines), and for the SALT2 model (red line). *Top:* training data. *Bottom:* validation data.

Sec. 4.2 with  $q_1 - q_4$  and  $q_1 - q_5$  components. In both cases the value of  $R_V = 2.6$  found with three factors is fixed in order to focus only on intrinsic parameters. In Fig. 22 we compare the spectral residuals of the SUGAR model to those of SUGAR trained with the additional factors.

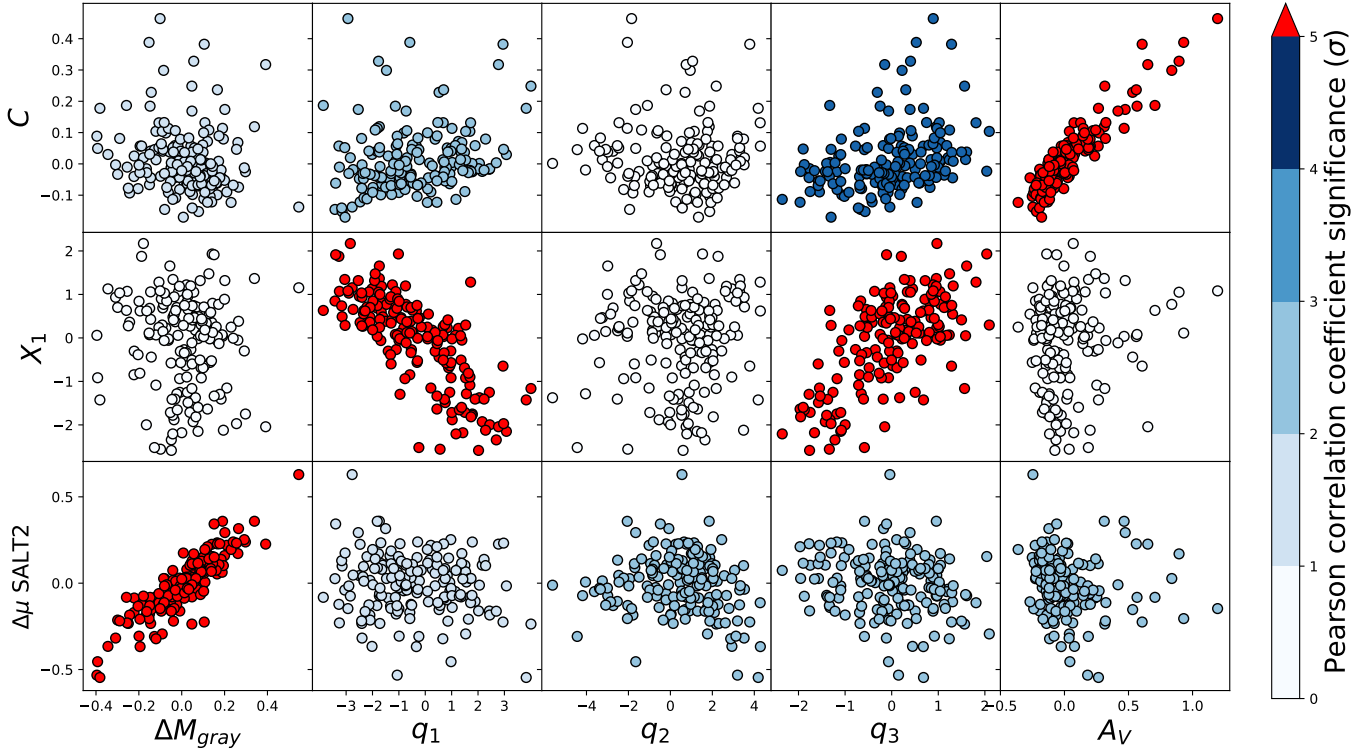
As expected, the addition of the two new components from factor analysis does not improve the description of SNe Ia as significantly as the addition of  $q_2$  or  $q_3$ . The factor  $q_4$  slightly improves the description within the Ca II H&K area (0.05 mag) and the Si II  $\lambda 6355$  (0.03 mag), but does not improve the SED outside these areas. The factor  $q_5$  does not significantly improve the description of the SED. This is confirmation that our choice of using three factors provides a good description of the SED.

## 7. Conclusion

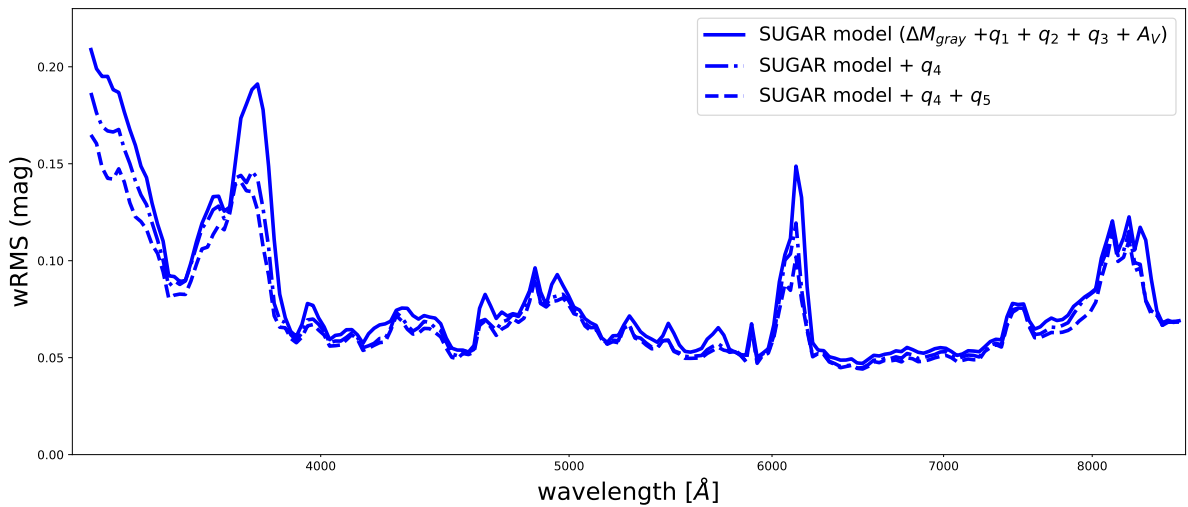
In this paper we have presented a new spectra-temporal empirical model of SNe Ia, named SUGAR. This model significantly improves the spectral description of SNe Ia compared to the current state of the art by going beyond the classical stretch and color parametrization.

In Sect. 2, we presented the SNfactory spectrophotometric dataset that was used to train the model. In Sect. 3, we presented

the intermediary data that were used to train the full SUGAR model. In a first step, we selected a set of 13 spectral indicators near maximum light in  $B$ -band that are composed of pseudo-equivalent widths and minima of P-Cygni profiles. Those spectral indicators were chosen to describe the intrinsic part of the SUGAR model because they are easy to define and are independent of host-galaxy dust extinction. Then, we defined a new basis where the 13 spectral indicators are uncorrelated. For this we developed a factor analysis algorithm that is more robust in the presence of errors than PCA algorithms. Three factors seems to be effective enough to describe the spectral indicators space at maximum light. The first factor describes the coherent variation of the pseudo-equivalent widths, mainly of the silicon and calcium lines. Like those lines, this factor is strongly correlated with stretch. The second factor is mainly correlated with the velocities and shows a very weak link with pseudo-equivalent widths, except for the Ca II H&K and Si II W. The third factor shows a slight correlation with the stretch parameter. Once the factors have been defined, we established a model of SED at maximum light based on these three first factors from spectral features, using the same underlying method as in Chotard et al. (2011). This allows us to separate the intrinsic behavior from color variation due to dust. Finally, we find that the color curve obtained is compatible with a Cardelli et al. (1989) extinction curve with  $R_V$  of 2.6. We then developed an interpolation method using



**Fig. 21.** SALT2 parameters estimated in photometry as a function of the SUGAR parameters estimated in spectroscopy (on training and validation sample). The color code estimates the correlation between the parameters SALT2 and SUGAR.



**Fig. 22.** wRMS of the residuals as a function of the wavelength for the full SUGAR model (blue line), for the SUGAR model with the addition of the factor  $q_4$  (blue dashed dotted line), and for the SUGAR model with the addition of the factor  $q_4$  and  $q_5$  (blue dashed lines).

Gaussian process in order to train the SUGAR model on a uniform and fixed time grid. The correlation length obtained varies between 5 and 12 days depending on spectral regions, which justifies the use of a 3-day time step for the grid. The uncertainty from the Gaussian processes is generally underestimated by a factor of 1.3 and more investigation is needed to understand why this occurs.

The training process for the SUGAR model is described in Sect. 4, and an interpretation of each new component is provided. Both factors  $q_1$  and  $q_3$  resemble a stretch effect, but  $q_3$  has less impact around maximum light and in color space than  $q_1$ . The effects of  $q_2$  are strongest in the areas of spectral features,

and mainly evident in the infrared as compared to broad band photometry.

After calculating the model, we showed that, instead of going through the calculation of spectral indicators, we can work directly with the spectral time series to recover the three factors and extinction parameter. By studying model residuals as a function of wavelength, it is shown that SUGAR improves the spectral description 0.1–0.4 mag with respect to SALT2. This is valid for both training and validation data sets, which confirms that there was no overtraining resulting from the addition of  $q_2$  and  $q_3$ . This shows that three parameters, defined at a given phase (i.e. maximum light) have predictive power at other phases.

Performance of the SUGAR model makes it an excellent candidate for use with surveys such as ZTF, LSST or WFIRST, and offers an alternative way of going beyond stretch and color to measure distance with SNIa.

*Acknowledgements.* We thank the technical staff of the University of Hawaii 2.2-m telescope, and Dan Birchall for observing assistance. We recognize the significant cultural role of Mauna Kea within the indigenous Hawaiian community, and we appreciate the opportunity to conduct observations from this revered site. This work was supported in part by the Director, Office of Science, Office of High Energy Physics of the U.S. Department of Energy under Contract No. DE-AC025CH11231. Support in France was provided by CNRS/IN2P3, CNRS/INSU, and PNC; LPNHE acknowledges support from LABEX ILP, supported by French state funds managed by the ANR within the Investissements d’Avenir programme under reference ANR-11-IDEX-0004-02. NC is grateful to the LABEX Lyon Institute of Origins (ANR-10-LABX-0066) of the University de Lyon for its financial support within the program “Investissements d’Avenir” (ANR-11-IDEX-0007) of the French government operated by the National Research Agency (ANR). Support in Germany was provided by DFG through TRR33 “The Dark Universe” and by DLR through grants FKZ 50OR1503 and FKZ 50OR1602. In China support was provided by Tsinghua University 985 grant and NSFC grant No. 11173017. Some results were obtained using resources and support from the National Energy Research Scientific Computing Center, supported by the Director, Office of Science, Office of Advanced Scientific Computing Research of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. We thank the Gordon and Betty Moore Foundation for their continuing support. Additional support was provided by NASA under the Astrophysics Data Analysis Program grant 15-ADAP15-0256 (PI:Aldering). We also thank the High Performance Research and Education Network (HPWREN), supported by National Science Foundation Grant Nos. 0087344 and 0426879. This project has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement No 759194 – USNAC). PFL acknowledges support from the National Science Foundation grant PHY-1404070. The work of MVP (participation in SUGAR implementation in sncoos) was supported by Russian Science Foundation grant 18-72-00159. We thank Claire-Alice Hébert for reviewing and giving helpful advice on this paper.

## References

- Aldering, G., Adam, G., Antilogus, P., et al. 2002, *SPIE Conf. Ser.*, 4836, 61
- Aldering, G., Antilogus, P., Bailey, S., et al. 2006, *ApJ*, 650, 510
- Amanullah, R., Johansson, J., Goobar, A., et al. 2015, *MNRAS*, 453, 3300
- Arsenijevic, V., Fabbro, S., Mourão, A. M., & Rica da Silva, A. J. 2008, *A&A*, 492, 535
- Astier, P., Guy, J., Regnault, N., et al. 2006, *A&A*, 447, 31
- Bailey, S., Aldering, G., Antilogus, P., et al. 2009, *A&A*, 500, L17
- Bellm, E. 2014, in *The Third Hot-wiring the Transient Universe Workshop*, eds. P. R. Wozniak, M. J. Graham, A. A. Mahabal, & R. Seaman, 27
- Betoule, M., Kessler, R., Guy, J., et al. 2014, *A&A*, 568, A32
- Blondin, S., & Tonry, J. L. 2007, *ApJ*, 666, 1024
- Blondin, S., Mandel, K. S., & Kirshner, R. P. 2011, *A&A*, 526, A81
- Bongard, S., Soulez, F., Thiébaud, É., & Pecontal, É. 2011, *MNRAS*, 418, 258
- Branch, D., Dang, L. C., Hall, N., et al. 2006, *PASP*, 118, 560
- Buton, C., Copin, Y., Aldering, G., et al. 2013, *A&A*, 549, A8
- Cardelli, J. A., Clayton, G. C., & Mathis, J. S. 1989, *ApJ*, 345, 245
- Chotard, N. 2011, PhD Thesis, Université Claude Bernard - Lyon I
- Chotard, N., Gangler, E., Aldering, G., et al. 2011, *A&A*, 529, L4
- Copeland, E. J., Sami, M., & Tsujikawa, S. 2006, *Int. J. Mod. Phys. D*, 15, 1753
- Delubac, T., Bautista, J. E., Busca, N. G., et al. 2015, *A&A*, 574, A59
- Dempster, A. P., Laird, N. M., & Rubin, D. B. 1977, *J. R. Stat. Soc. Ser. B (Methodological)*, 39, 1
- Fakhouri, H. K., Boone, K., Aldering, G., et al. 2015, *ApJ*, 815, 58
- Filippenko, A. V., Richmond, M. W., Matheson, T., et al. 1992a, *ApJ*, 384, L15
- Filippenko, A. V., Richmond, M. W., Branch, D., et al. 1992b, *AJ*, 104, 1543
- Folatelli, G. 2004, *New Astron. Rev.*, 48, 623
- Foley, R. J., & Kasen, D. 2011, *ApJ*, 729, 55
- Garavini, G., Folatelli, G., Nobili, S., et al. 2007, *A&A*, 470, 411
- Ghahramani, Z., & Hinton, G. E. 1997, *The EM Algorithm for Mixtures of Factor Analyzers*, Tech. rep.
- Guillochon, J., Parrent, J., Kelley, L. Z., & Margutti, R. 2017, *ApJ*, 835, 64
- Guy, J., Astier, P., Baumont, S., et al. 2007, *A&A*, 466, 11
- Guy, J., Sullivan, M., Conley, A., et al. 2010, *A&A*, 523, A7
- Hamuy, M., Phillips, M. M., Maza, J., et al. 1995, *AJ*, 109, 1
- Kelly, P. L., Hicken, M., Burke, D. L., Mandel, K. S., & Kirshner, R. P. 2010, *ApJ*, 715, 743
- Kim, A. G., Thomas, R. C., Aldering, G., et al. 2013, *ApJ*, 766, 84
- Lantz, B., Aldering, G., Antilogus, P., et al. 2004, *SPIE Conf. Ser.*, 5249, 146
- Léget, P. F. 2016, PhD Thesis, Université Blaise Pascal
- LSST Dark Energy Science Collaboration 2012, ArXiv e-prints [arXiv:1211.0310]
- Maguire, K., Sullivan, M., Ellis, R. S., et al. 2012, *MNRAS*, 426, 2359
- Mandel, K. S., Foley, R. J., & Kirshner, R. P. 2014, *ApJ*, 797, 75
- Mandel, K. S., Scolnic, D. M., Shariff, H., Foley, R. J., & Kirshner, R. P. 2017, *ApJ*, 842, 93
- Nordin, J., Östman, L., Goobar, A., et al. 2011, *A&A*, 526, A119
- Nordin, J., Aldering, G., Antilogus, P., et al. 2018, *A&A*, 614, A71
- Nugent, P., Phillips, M., Baron, E., Branch, D., & Hauschildt, P. 1995, *ApJ*, 455, L147
- Pearson, K. 1901, *Phil. Mag.*, 2, 559
- Pereira, R., Thomas, R. C., Aldering, G., et al. 2013, *A&A*, 554, A27
- Perlmutter, S., Gabi, S., Goldhaber, G., et al. 1997, *ApJ*, 483, 565
- Perlmutter, S., Aldering, G., della Valle, M., et al. 1998, *Nature*, 391, 51
- Perlmutter, S., Aldering, G., Goldhaber, G., et al. 1999, *ApJ*, 517, 565
- Phillips, M. M. 1993, *ApJ*, 413, L105
- Planck Collaboration XIII. 2016, *A&A*, 594, A13
- Polin, A., Nugent, P., & Kasen, D. 2019, *ApJ*, 873, 1
- Pskovskii, I. P. 1977, *Sov. Astron.*, 21, 675
- Pskovskii, Y. P. 1984, *Sov. Astron.*, 28, 658
- Rasmussen, C. E., & Williams, C. K. 2006, *Gaussian Processes for Machine Learning* (Cambridge: The MIT Press), 38, 715
- Rest, A., Scolnic, D., Foley, R. J., et al. 2014, *ApJ*, 795, 44
- Riess, A., Press, W., & Kirshner, R. 1996, *AJ*, 473, 88
- Riess, A. G., Filippenko, A. V., Challis, P., et al. 1998, *AJ*, 116, 1009
- Riess, A. G., Macri, L. M., Hoffmann, S. L., et al. 2016, *ApJ*, 826, 56
- Rigault, M., Copin, Y., Aldering, G., et al. 2013, *A&A*, 560, A66
- Rigault, M., Aldering, G., Kowalski, M., et al. 2015, *ApJ*, 802, 20
- Rigault, M., Brinnel, V., Aldering, G., et al. 2018, *A&A*, submitted
- Roman, M., Hardin, D., Betoule, M., et al. 2018, *A&A*, 615, A68
- Rust, B. W. 1974, PhD Thesis, Oak Ridge National Lab., TN
- Sasdelli, M., Hillebrandt, W., Aldering, G., et al. 2015, *MNRAS*, 447, 1247
- Saunders, C., Aldering, G., Antilogus, P., et al. 2018, *ApJ*, 869, 167
- Savitzky, A., & Golay, M. J. E. 1964, *Anal. Chem.*, 36, 1627
- Scalzo, R. A., Aldering, G., Antilogus, P., et al. 2010, *ApJ*, 713, 1073
- Schmidt, B. P., Suntzeff, N. B., Phillips, M. M., et al. 1998, *ApJ*, 507, 46
- Scolnic, D. M., Riess, A. G., Foley, R. J., et al. 2014, *ApJ*, 780, 37
- Scolnic, D. M., Jones, D. O., Rest, A., et al. 2018, *ApJ*, 859, 101
- Silverman, J. M., Kong, J. J., & Filippenko, A. V. 2012, *MNRAS*, 425, 1819
- Spearman, C. 1904, *Am. J. Psychol.*, 15, 201
- Spearman, C. 1927, *The Abilities of Man* (New York: Macmillan)
- Spergel, D., Gehrels, N., Baltay, C., et al. 2015, ArXiv e-prints [arXiv:1503.03757]
- Sullivan, M., Conley, A., Howell, D. A., et al. 2010, *MNRAS*, 406, 782
- Suzuki, N., Rubin, D., Lidman, C., et al. 2012, *ApJ*, 746, 85
- Tripp, R. 1998, *A&A*, 331, 815
- Troxel, M. A., MacCann, N., Zuntz, J., et al. 2018, *Phys. Rev. D*, 98, 043528
- Wang, X., Filippenko, A. V., Ganeshalingam, M., et al. 2009, *ApJ*, 699, L139

## Appendix A: Spectral-indicator measurements

As the spectral indicators are a key ingredient of our statistical analysis, we need a robust and automatic algorithm to derive them from the rest-frame spectra and estimate the associated statistical and systematic uncertainties. The main issue to be addressed is the automatic detection of feature boundaries – usually local extrema – as they shift both along the phase and from one supernova to another. Sometimes, the extremum itself cannot be uniquely defined, for example, when there is a mix of several local extrema at the same position or no significant extremum such as in the blue edge of the O I  $\lambda 7773$  feature. The photon noise also limit the accuracy when determining the wavelength of the maximum, and may induce a systematic shift due to the nonlinear process involved in finding an extremum. In this section, we give a brief overview of the general and automatic method developed and implemented by Chotard (2011) to measure these spectral indicators and evaluate their corresponding uncertainties. This method shares some similarities with previous analyses (Folatelli 2004; Garavini et al. 2007; Nordin et al. 2011; Blondin et al. 2011; Silverman et al. 2012): it is based on extrema searches in a fixed spectral domain after smoothing the data. Improving on previous studies, our smoothing procedure is based on an optimal determination of the regularization parameters, and more noteworthy, we developed a thorough determination of the uncertainties based on a Monte Carlo procedure taking as input the variance of our signal.

### A.1. Method

Most of the spectral indicators presented here are defined by the position in wavelength and flux of at least one local minimum or maximum, that is, peaks or troughs of the SN Ia spectral features. In order to compute a precise estimate of these local extrema, a Savitsky–Golay (Savitzky & Golay 1964, SG) smoothing is applied to the original spectra before the rebinning at  $1500 \text{ km s}^{-1}$  described in Sect. 2. Since the SG optimal window depends both on the underlying spectral shape and on the S/N ratio, which varies across wavelength, an independent smoothing is applied to each of the nine spectral zones of interest defined in Table A.1.

Each extremum is then selected on the smoothed spectrum as the local extremum inside a given wavelength range, which can be found in Tables A.2 and A.3 for the equivalent widths and the feature velocities respectively. These wavelength ranges have been trained on a set of  $\sim 50$  SNe Ia within a phase range of  $\pm 5$  days around  $B$ -band maximum light so that they match the observed diversity of our spectra. The positions in wavelength and flux of these extrema are then used in the spectral-indicator measurements, either directly, for example, for the velocities, or indirectly, for instance, for the equivalent widths through the definition of their pseudo-continuum.

### A.2. Optimal smoothing

The purpose of smoothing, also known as regularization, is to transform the original noisy data in order to get closer on average to the unknown original spectrum, based on some regularity hypothesis. A parameter describing how smooth the final function is has to be introduced, and has to be estimated, either based on physical consideration or deduced from the data themselves. Here we follow the latter approach and describe an optimal way of setting the smoothing parameter given a class of transformations.

**Table A.1.** Independently smoothed spectral regions, with minimal and maximal boundaries  $\lambda_{\min}$  and  $\lambda_{\max}$  (in Å).

$\lambda_{\min}$ – $\lambda_{\max}$	Elements
3450–4070	Ca II H&K
3850–4150	Si II $\lambda 4131$ ; Co II
4000–4610	Mg II triplet
4350–5350	Fe II blend
5060–5700	S II W
5500–6050	Si II $\lambda 5972$
5800–6400	Si II $\lambda 6355$
5500–6400	Si II $\lambda 5972$ ; 6355
6500–8800	O I triplet; Ca II IR

**Table A.2.** Definition of the wavelength regions (in Å) where the extrema are expected to be found.

Indicators	$\lambda_{\min}^b$ – $\lambda_{\max}^b$	$\lambda_{\min}^r$ – $\lambda_{\max}^r$
Ca II H&K	3504–3687	3830–3990
Si II $\lambda 4131$	3830–3990	4030–4150
Mg II	4030–4150	4450–4650
Fe $\lambda 4800$	4450–4650	5050–5285
S II W	5050–5285	5500–5681
Si II $\lambda 5972$	5550–5681	5850–6015
Si II $\lambda 6355$	5850–6015	6250–6365
O I $\lambda 7773$	7100–7270	7720–8000
Ca II IR	7720–8000	8300–8800

**Notes.** The  $b$  and  $r$  exponents respectively represent the left (blue) and right (red) peak.

**Table A.3.** Wavelength regions (in Å) used to compute the feature velocities, together with their rest-frame wavelengths,  $\lambda_0$ .

Velocity	$\lambda_{\min}$ – $\lambda_{\max}$	$\lambda_0$
$v(\text{Si II } \lambda 4131)$	3963–4034	4131
$v(\text{S II W } \lambda 5454)$	5200–5350	5454
$v(\text{S II W } \lambda 5640)$	5351–5550	5640
$v(\text{Si II } \lambda 6355)$	6000–6210	6355

### A.2.1. Formalism

Smoothing a noisy spectrum  $Y$  consists of the determination of the smoothed spectrum  $Y'$ , which is a function of the initial flux  $Y' = f(Y)$ .  $Y'$  is an estimator of the noise-free and unknown original flux, convolved by the instrumental resolution:  $\hat{Y}$ . It is related to the observed flux by:

$$Y = \hat{Y} + N, \quad (\text{A.1})$$

where  $N$  is a realization of the spectral noise vector. The goodness of the smoothing being represented by the error function,

$$\|Y' - \hat{Y}\| \triangleq (Y' - \hat{Y})^T W (Y' - \hat{Y}), \quad (\text{A.2})$$

where  $W$  is the inverse of the noise covariance matrix, this requires minimization of a quantity which depends both on the noise properties and the signal shape. As the true spectrum,  $\hat{Y}$ , is unknown, we need to build an estimator of  $\|Y' - \hat{Y}\|$  which is independent of  $\hat{Y}$ . For a linear regularization, one can write:

$$Y' = f(Y) = B_p Y, \quad (\text{A.3})$$

where  $B_p$  is a smoothing matrix which only depends on the chosen smoothing technique and a smoothing parameter  $p$ . Introducing Eqs. (A.1) and (A.3) into (A.2), one can show that:

$$\|Y' - \hat{Y}\| = (Y' - Y + N)^T W(Y' - Y + N) \quad (\text{A.4})$$

$$= \|Y' - Y\| + 2N^T W(B_p - I)(\hat{Y} + N) + N^T W N. \quad (\text{A.5})$$

An estimator of the error function,  $\epsilon$ , that has to be minimized can be constructed by noticing that

$$E[\|Y' - \hat{Y}\|] = E[\|Y' - Y\|] + 2 \text{Tr}(B_p) - n, \quad (\text{A.6})$$

where  $E$  is the mathematical expectation value,  $\text{Tr}$  is the Trace operator, and  $n$  is the rank of the vector  $Y$ . As we have only one realization of  $Y$ , this translates to:

$$\epsilon = \|Y' - Y\| + 2 \text{Tr}(B_p) - n, \quad (\text{A.7})$$

which is the quantity that is minimized with respect to  $p$ .

Given a class of linear smoothing methods  $B_p$ , we have thus defined a procedure to find the value of  $p$  for which the smoothed spectrum best reproduces the original unknown one given an observed spectrum. This method is not exempt from possible overtraining, however this is mitigated by restricting the optimization to a single parameter. This was tested with Monte Carlo simulations.

## A.2.2. B matrix estimation

Among all possibilities, we chose a Savitzky–Golay regularization (Savitzky & Golay 1964, SG) as it allows the reduction of the high frequency noise, while keeping the original shape of the feature. This method relies on fitting a  $k$ -order polynomial function for each point  $i$  in a fixed window size  $p$  (with  $p > k + 1$ ) centered on the current point and is designed to preserve the locus of maxima for even values of  $k$ . Each observed data point is replaced by its fit value, and the window then moves to the next data point until the spectrum is completely smoothed. As we have fixed the degree of the polynomial function to  $k = 2$ , the only parameter that has to be optimized is the window size,  $p$ . This transformation is linear, and the optimal window size  $p$  and corresponding smoothing matrix  $B_p$  are found by minimizing Eq. (A.7). The value of  $p$  is estimated individually for each spectral region of a given spectrum, and the spectral indicator measurements are performed on the corresponding smoothed spectrum, as indicated above.

## A.3. Uncertainties

The uncertainty on a given spectral-indicator measurement arises from the statistical noise of the data and the induced uncertainties on the measurement method parameters. In our approach, these two uncertainties are independently measured using the smoothed spectrum as a reference: the SG window size is typically large enough so that the residual noise can be neglected in the simulations.

A random noise matching the statistical properties of the observed data is then added in order to generate a mock spectrum, and the spectral indicators are measured on this new spectrum as if it were the observed one. After one thousand generations, we are able to derive the statistical fluctuations of the obtained values, which we quote as the statistical uncertainty. This is computed as the standard deviation from the value measured on the real spectrum, thus taking into account a potential

bias. This bias typically corresponds to 10–20% of the total error budget.

In order to save computation time, the initial SG window size  $p$  is determined once and for all using the original spectrum. It is then kept constant when measuring spectral indicators on all the simulated spectra. However, the noise affecting the initial spectrum induces an uncertainty in the smoothing parameter. This has been studied on a reduced set of simulations for which the optimal  $p$  was derived for each realization of the noise, and the corresponding uncertainty was found to range from 15% to 20% depending on the spectral zone. We then propagate this uncertainty by computing the spectral-indicator values on the original spectrum for several values of  $p$  in this 15–20% range. The standard deviation of the resulting spectral-indicator value distribution gives an estimate of the systematic error introduced by the arbitrariness of the smoothing method, which is found to be  $\sim 20\%$  of the total error on average. These two uncertainties are quadratically added together to derive the final uncertainty for a given spectral indicator. This estimate takes into account the statistical noise of the spectrum, as well as the induced scatter in wavelength and flux of the extrema. Our measurement errors thus include all nonlinear effects due to limited signal to noise and can be trusted for subsequent statistical analysis. More details could be found in Chotard (2011) and Nordin et al. (2011).

## A.4. Performances

### A.4.1. Failure rate

When one of the extrema defining a spectral indicator lies on a flat and/or noisy section of the spectrum, its measurement has a chance of failure. In that case, the measurement is automatically rejected and a visual scan using control plots is performed to confirm the actual lack of an extremum. The rejection of the “bad” measurements is performed using a  $3\sigma$  clipping in the measured-uncertainty space: if the uncertainty made on a given measurement is larger than  $m + 3 \times \text{std}$ , where  $m$  and  $\text{std}$  are the average and standard deviation of this spectral-indicator uncertainty distribution (for the whole sample), the corresponding measurement is rejected. Considering all the spectral indicators measured on the 113 input spectra in a range of phase of  $\pm 2.5$  days around maximum light and presented in this paper, the global failure rate of the measurement procedure is less than  $10^{-3}$  for the automatic selections mentioned above. If it does happen, the spectral indicators are set to the average value and assigned an infinite error.

### A.4.2. Quoted uncertainties

A simple test has been performed to confirm the robustness of the method. In our selected sample, 18 SNe Ia have two or more spectra taken in the same night in a phase interval of  $\pm 5$  days around maximum light. We then computed the distribution of the pull,

$$\frac{\delta I}{\sigma_{\delta I}} = \frac{I_1 - I_2}{\sqrt{\sigma_{I_1}^2 + \sigma_{I_2}^2}}, \quad (\text{A.8})$$

for all the spectral indicators  $I$  (feature velocity and absorption ratio) measured on each spectrum of a same night (for a same supernova). This distribution is centered around  $-0.01$  with a dispersion of 1.06 which indicates that our estimation of the uncertainty is valid up to a possible underestimation of the error

by 6%. This number is small enough so that we can trust our uncertainty estimation for the main analysis.

## Appendix B: Expectation-maximization factor analysis

Dimensionality reduction in the presence of noisy data is often an overlooked problem. However, standard methods like PCA tend to fail at capturing the intrinsic variability of the data and the principal components align with the direction of the noise when the latter becomes important. Factor analysis on the other hand is a statistical method designed to model the covariance structure of high dimensional data using a small number of latent variables. It estimates both the natural variability of the sample and the noise arising from the measurements, under the assumption that the statistics are the same for all data records. Our case is slightly different: on one hand, the noise statistics are different for each measurement, but on the other hand, their variance is already known. We thus adapted the expectation-minimization algorithm presented in Ghahramani & Hinton (1997) to accommodate for the specifics of our problem. The resulting method is also known as probabilistic PCA. The formalism is the following:  $\mathbf{x}_i$  is a vector of rank  $l$  representing the measurement  $i$ . It is linked to the factor  $\mathbf{q}_i$ , vector of rank  $k \leq l$  and the noise  $\eta_i$  by:

$$\mathbf{x}_i = \mathbf{x}^0 + \Lambda \mathbf{q}_i + \eta_i, \quad (\text{B.1})$$

where  $\mathbf{x}^0$  is a central value which can be further neglected without loss of generality ( $\mathbf{x}^0 = 0$ ),  $\mathbf{q}_i$  is assumed to follow a normal distribution of unit variance, and  $\eta_i$  follows a multivariate normal distribution of variance  $\Psi_i$ . In our case,  $\Psi_i$  is diagonal, a property that can be used to speed-up computations.  $\Lambda$  is the matrix containing the  $k$  explicative vectors that we need to determine. While  $\Lambda$  is not uniquely defined,  $\Lambda \Lambda^T$  is and represents the intrinsic covariance of the data, that is, the one we would observe in the absence of noise. The eigenvectors of  $\Lambda \Lambda^T$  thus correspond to the  $k$  first eigenvectors that PCA would have found in the absence of noise.

To find  $\Lambda$ , instead of directly maximizing the likelihood of observing  $\mathbf{x}_i$ , the expectation-maximization algorithm introduces the latent variable  $\mathbf{q}_i$  and then maximizes the expected likelihood over  $\mathbf{q}_i$ . The joint probability of  $\mathbf{x}_i$  and  $\mathbf{q}_i$  is the following multivariate normal distribution:

$$P \left( \begin{bmatrix} \mathbf{x}_i \\ \mathbf{q}_i \end{bmatrix} \right) = \mathcal{N} \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \Lambda \Lambda^T + \Psi_i & \Lambda \\ \Lambda^T & \mathbf{I} \end{bmatrix} \right). \quad (\text{B.2})$$

The block-diagonal elements of the covariance matrix represent, respectively, the covariances of  $\mathbf{x}_i$  and  $\mathbf{q}_i$ , and the non block-diagonal elements represent the covariance arising from the relation (B.1). Expectation-maximization is an iterative procedure which ensures that the likelihood increases at each iteration and it has been shown that the convergence is faster than using a gradient method (Dempster et al. 1977). Each iteration proceeds in two steps. The first step, called the E-step, consists of calculating the expectation of the conditional first and second moments of  $\mathbf{q}_i$  for a given  $\Lambda$ :

*E-step:*

$$\mathbf{q}_i \hat{=} E[\mathbf{q}_i | \mathbf{x}_i] = \Lambda^T (\Psi_i + \Lambda \Lambda^T)^{-1} \mathbf{x}_i \quad (\text{B.3})$$

$$E[\mathbf{q} \mathbf{q}^T | \mathbf{x}_i] = \mathbf{I} - \Lambda^T (\Psi_i + \Lambda \Lambda^T)^{-1} \Lambda + \mathbf{q}_i \mathbf{q}_i^T. \quad (\text{B.4})$$

Once these two quantities are computed, the second step, or M-step, consists of estimating the  $\Lambda$  matrix by maximizing the like-

lihood expectation, which provides the condition:

$$\sum_{i=0}^N \Psi_i^{-1} \Lambda E[\mathbf{q} \mathbf{q}^T | \mathbf{x}_i] = \sum_{i=0}^N \Psi_i^{-1} \mathbf{x}_i \mathbf{q}_i^T. \quad (\text{B.5})$$

In the case where  $\Psi_i$  is diagonal, one can independently calculate each row of  $\Lambda$ , noted  $\Lambda^j$ , according to the relation

*M-step:*

$$\Lambda^j = \left[ \sum_{i=0}^N \frac{x_i^j}{\psi_i^{jj}} \mathbf{q}_i^T \right] \left[ \sum_{i=0}^N \frac{1}{\psi_i^{jj}} E[\mathbf{q} \mathbf{q}^T | \mathbf{x}_i] \right]^{-1}, \quad (\text{B.6})$$

After the last iteration, the internal degeneracies of the description are lifted with the transformation of  $\Lambda$  into an orthogonal  $\Lambda'$  matrix which satisfies the condition:

$$\Lambda' \Lambda'^T = \Lambda \Lambda^T. \quad (\text{B.7})$$

The columns of  $\Lambda'$  are aligned with the eigenvectors of  $\Lambda \Lambda^T$  and the square of their norm are the respective eigenvalues.

## Appendix C: Orthogonal distance regression

### C.1. Expectation and maximization steps

Similarly to the factor analysis,  $\mathbf{h}_i$  and  $\mathbf{A}$  are estimated iteratively. The first step consists of minimizing the  $\chi^2$  with respect to  $\mathbf{h}_i$ , which amounts to solving the equation:

$$\frac{\partial \chi^2}{\partial \mathbf{h}_i} = 0 \quad (\text{C.1})$$

and consequently gives:

*E-step:*

$$\mathbf{h}_i = (\mathbf{A}^T \mathbf{W}_{M_i} \mathbf{A} + \mathbf{W}_{x_i})^{-1} (\mathbf{A}^T \mathbf{W}_{M_i} \mathbf{M}_i + \mathbf{W}_{x_i} \mathbf{x}_i). \quad (\text{C.2})$$

Once the E-step is performed we can estimate the matrix  $\mathbf{A}$  by

$$\frac{\partial \chi^2}{\partial \mathbf{A}} = 0, \quad (\text{C.3})$$

which gives:

*M-step:*

$$\tilde{\mathbf{A}} = \left( \left[ \sum_{i=0}^N \mathbf{h}_i \mathbf{h}_i^T \otimes \mathbf{W}_{M_i} \right] \right)^{-1} \left[ \sum_{i=0}^N \mathbf{W}_{M_i} \mathbf{M}_i \mathbf{h}_i^T \right], \quad (\text{C.4})$$

where  $\tilde{\mathbf{A}}$  is the vector that contains all columns of  $\mathbf{A}$  put end-to-end and the  $\otimes$  operator is the Kronecker product. The estimation of the  $\mathbf{A}$  matrix and the orthogonal projections  $\mathbf{h}_i$  is done by reiterating the E-step and M-step until  $\chi^2$  convergences.

### C.2. Fixing the degeneracies

The free parameters of Eqs. (5) and (22) that we find through orthogonal distance regression contain degenerate degrees of freedom. These must be fixed in order to present an unique an interpretable solution.

### C.2.1. Degeneracies of the extinction fit

The term describing the extinction,  $A_{\lambda_0,i}\gamma(\lambda)$ , is not directly constrained by external observations. As a consequence, Eq. (5) is invariant under several transformations. The first one is

$$A_{\lambda_0,i}^{\text{new}} = A_{\lambda_0,i} + \sum_j h_i^j c^j, \quad (\text{C.5})$$

$$\alpha_\lambda^{j,\text{new}} = \alpha_\lambda^j - \gamma_\lambda c^j. \quad (\text{C.6})$$

We must therefore fix the  $c^j$ . A natural choice is to impose that the  $A_{\lambda_0,i}$  are decorrelated with the  $h_i^j$ . Indeed, any correlation would imply that the extinction term contains information linked to the intrinsic properties described by  $h_i^j$ . Thus after each E-step, we imposed a correlation of zero between the  $A_{\lambda_0,i}$  and the  $h_i^j$ , which amounts to applying the transformations (C.5) and (C.6) taking for the  $c^j$

$$\mathbf{c} = [\text{cov}(\tilde{\mathbf{h}})]^{-1} \text{cov}(\tilde{\mathbf{h}}, \mathbf{A}_{\lambda_0}), \quad (\text{C.7})$$

where  $\tilde{\mathbf{h}}$  is the vector that contains the  $h_i$ ,  $\text{cov}(\tilde{\mathbf{h}})$  is the covariance matrix of the  $\tilde{\mathbf{h}}$ ,  $\text{cov}(\tilde{\mathbf{h}}, \mathbf{A}_{\lambda_0})$  is the vector that contains the covariances between the  $h_i^j$  and the  $A_{\lambda_0,i}$ . The second degeneracy to be fixed is the scale of  $\gamma_\lambda$ : only the product  $A_{\lambda_0,i}\gamma_\lambda$  plays a role, and we then impose after each iteration

$$\gamma_{\lambda_0} = 1 \quad (\text{C.8})$$

by rescaling accordingly  $A_{\lambda_0,i}$  and  $\gamma_\lambda$ .

Finally, we need a prescription for the mean value of  $A_{\lambda_0,i}$ : we choose it to be centered at zero. This corresponds to the following transformation

$$A_{\lambda_0,i}^{\text{new}} = A_{\lambda_0,i} - \frac{1}{N} \sum_i A_{\lambda_0,i}, \quad (\text{C.9})$$

$$M_{\lambda,0}^{\text{new}} = M_{\lambda,0} + \frac{\gamma_\lambda}{N} \sum_i A_{\lambda_0,i}. \quad (\text{C.10})$$

which leaves the  $\chi^2$  invariant and amounts to placing the average spectrum at the average extinction.

### C.2.2. Degeneracies of the global fit

The estimation of the grey parameter,  $\Delta M_{\text{grey } i}$ , is affected by different degeneracies which can be fixed after the E or M iterations. We choose to fix them after each E-step. The first one arises from the invariance of Eq. (22) under the transformation:

$$\Delta M_{\text{grey } i}^{\text{new}} = \Delta M_{\text{grey } i} + \sum_j h_i^j c^j, \quad (\text{C.11})$$

$$\alpha_{t,\lambda}^{j,\text{new}} = \alpha_{t,\lambda}^j - c^j. \quad (\text{C.12})$$

We must therefore fix the  $c^j$ . Similarly to the prescription made for the extinction fit, a natural choice is to impose that the  $\Delta M_{\text{grey } i}$  and the  $h_i^j$  are uncorrelated. Indeed, the interpretation is made easier with  $\alpha^j$  containing all the information driven by  $h_i^j$ , including its global impact on magnitudes. We thus impose after

each E-step that the correlation between the  $\Delta M_{\text{grey } i}$  and the  $h_i^j$  is zero. This is obtained by applying the transformations (C.11) and (C.12) taking for the  $c^j$ :

$$\mathbf{c} = [\text{cov}(\tilde{\mathbf{h}})]^{-1} \text{cov}(\tilde{\mathbf{h}}, \Delta \mathbf{M}_{\text{grey}}), \quad (\text{C.13})$$

where  $\tilde{\mathbf{h}}_i$  is the vector that contains the  $h_i^j$ ,  $\text{cov}(\tilde{\mathbf{h}})$  is the observed covariance matrix computed on the set of  $\tilde{\mathbf{h}}_i$  vectors and  $\text{cov}(\tilde{\mathbf{h}}, \Delta \mathbf{M}_{\text{grey}})$  is the vector that contains the covariances between  $h_i^j$  and  $\Delta M_{\text{grey } i}$ . Finally, by convention, the  $\Delta M_{\text{grey } i}$  are centered on zeros at each step. This is obtained by the following transformation, which leaves the  $\chi^2$  invariant:

$$\Delta M_{\text{grey } i}^{\text{new}} = \Delta M_{\text{grey } i} - \frac{1}{N} \sum_i \Delta M_{\text{grey } i}, \quad (\text{C.14})$$

$$M_{t,\lambda,0}^{\text{new}} = M_{t,\lambda,0} + \frac{1}{N} \sum_i \Delta M_{\text{grey } i}. \quad (\text{C.15})$$

## Appendix D: Matrix dispersion estimation

Once the minimum of the  $\chi^2$  defined in Eq. (9) is reached, the dispersion matrix is estimated. For this purpose, we use the same method as in Chotard et al. (2011) and which is described in detail in Chotard (2011). The approach is to calculate  $\mathbf{D}$  from the observed dispersion of residuals and to subtract the average dispersion due to uncertainties:

$$\mathbf{D} = \frac{1}{N} \sum_i (\delta \mathbf{M}_i \delta \mathbf{M}_i^T - \mathbf{C}_i) \quad (\text{D.1})$$

where  $\delta \mathbf{M}_i$  are the residuals of the model once the minimum of the  $\chi^2$  is reached and is defined for a given wavelength as:

$$\delta M_{\lambda,i} = M_{\lambda,i} - M_{\lambda,0} - \sum_j q_i^j \alpha_\lambda^j - A_{\lambda_0,i} \gamma_\lambda \quad (\text{D.2})$$

and  $\mathbf{C}_i$  is the covariance matrix that accounts for the total propagation of residuals error and is defined as:

$$\mathbf{C}_i = \begin{pmatrix} \ddots & & 0 \\ & \sigma_{\lambda i}^2 & \\ 0 & & \ddots \end{pmatrix} + (\sigma_{\text{cal}}^2 + \sigma_z^2) \begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{pmatrix} + \alpha \text{cov}(\mathbf{q}_i) \alpha^T \quad (\text{D.3})$$

where  $\alpha$  is the matrix that contains the intrinsic vectors and is defined as:

$$\alpha = (\alpha^1, \alpha^2, \alpha^3, \dots) \quad (\text{D.4})$$

In order to ensure that the matrix  $\mathbf{D}$  is positive definite, the negative eigenvalues of the matrix are set to zero. Once the matrix  $\mathbf{D}$  has been calculated, we add it in the expression of the Eq. (9) in order to recalculate the spectral distribution in energy and we iterate the calculations of  $\mathbf{D}$  and the computation of the SED, until reaching the maximum of the restricted maximum likelihood (Guy et al. 2010).