FUNCTIONAL LIGHT CURVE MODELS FOR TYPE IA SUPERNOVAE AND MIRA

VARIABLES, WITH THEIR APPLICATION OF DISTANCE DETERMINATION

A Dissertation

by

SHIYUAN HE

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Jianhua Huang
Committee Members,	James Long
	Lifan Wang
	Raymond J. Carroll
Head of Department,	Valen Johnson

May 2017

Major Subject: Statistics

ABSTRACT

Both type Ia supernovae and variable stars are important distance indicators in astronomy. The peak luminosity of type Ia supernovae and the period-luminosity relation of Miras can be employed for relative distance determination. For both SNIa and Mira, we develop light curve models with noisy, sparse and irregularly-sampled data.

We develop a functional principal component method for SNIa light curves. Each SNIa light curve is expressed as a linear combination of a mean function and several principal component functions. The coefficients of the principal component functions are called scores. The proposed method takes into account peak registration, shape constraints and is equipped with a fast training algorithm. The resulting model provides high quality fit to each light curve. In addition, the scores present powerful characterization of SNIa. They demonstrate connection with interstellar dusting, spectral classes and other physical properties. Moreover, the method provides a functional linear form in place of the commonly used ΔM_{15} parameter for distance predictions.

We also develop a semi-parametric model for Mira period estimation. The proposed method has a close relation with a Gaussian process model, and is solved in an empirical Bayesian framework. The empirical Bayesian is solved by a fast quasi-Newton algorithm with warm start, and combined with a grid search in the frequency parameter due to the related high multimodality. The proposed method is compared with the traditional Lomb-Scargle method in a large-scale simulation and shows considerable improvement.

ACKNOWLEDGMENTS

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supported by a dissertation committee consisting of Professor Jianhua Z. Huang (advisor), Professor James Long and Professor Raymond J. Carroll of the Department of Statistics and Professor Lifan Wang of the Department of Physics and Astronomy.

The work of Chapter 2 and Chapter 3 was conducted with Professor Jianhua Z. Huang and Professor Lifan Wang. The work of Chapter 4 is a collaborative product with Professor Jianhua Z. Huang, Professor James Long, as well as Professor Lucas Z. Macri and graduate student Wenlong Yuan of the Department of Physics and Astronomy. All the research work of the dissertation was completed by the student as the first author.

Funding Sources

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

Both Type Ia supernovae and Mira variables are important tools to measure the distance between their hosting galaxies and us. SNIa have consistent peak magnitude (brightness) when they explode, and our observed brightness is inverse proportional to the squared distance. These two facts make SNIa "standardizable candles" for distance measurement. On the other hand, Miras and other variable stars have their brightness changes periodically. They can also be employed for distance measurement because of the period luminosity relation, which states that their average magnitudes and the logarithm of their periods follow a linear relation. When we observe SNIa or Mira variables, we are also looking into the past due to the light has a constant and finite speed. As a result, we are measuring the scale of the universe at its different ages. Astronomers are interested in these measurements because they want to test cosmology models at a large scale. It is these measurements that give the evidence that our universe is expanding at an accelerating rate. The accuracy of these measurement and data processing reduces the distance prediction uncertainty, and as a result, to provide tight constraints for cosmology parameters.

1.1 Type Ia Supernova and its Light Curve Models

Type Ia supernovae (SNIa) have relatively constant intrinsic luminosity when they explode. They provide us a a very important tool for distance measurement based on the inverse square law, which states that the observed luminosity is inverse proportional to the squared distance. When observing SNIa, we are looking into the past due to constant speed of light. This helps us to measure the scale of the universe at its different ages. The observation of SNIa provided the direct evidence of acceleratingly expanding universe (Riess et al., 1998; Perlmutter et al., 1999).

The most common observation of SNIa is its light curves. The light curves are sparse

and irregularly sampled time series data. They record how the object brightness changes over time, and the brightness is measured in terms of "magnitude". Moreover, the light curves are usually observed in several filters (bands), covering different wavelength ranges. In the left panel of Figure 1.1 are four light curves of supernova SN 1998dh, observed by the Lick Observatory Supernova Search (LOSS) (Ganeshalingam et al., 2010). These four light curves are recorded in four optical bands, denoted by B, V, R, I. Their observation points are plotted as red, blue, green and black, respectively. Notice the vertical axis is plotted in reverse order. This is because brighter object has smaller value of magnitude. In the right panel of Figure 1.1 are the transmission curves of the four filters. The transmission curve depicts the proportion of light passes through the filter at a given wavelength.

The key interest of cosmology studies lies around the peak epoch, the time when the supernova has brightest luminosity. However, SNIa have small inhomogeneity in their peak magnitudes. Generally, the brighter SNIa have wider light curve and bluer color, and the dimmer SNIa have narrower light curves and redder color. One classical measure of light curve width is the ΔM_{15} parameter (Phillips, 1993). It gauges the change in magnitude 15 days after the peak. Adjusting these light curve width and color effects, the dispersion of distance prediction can be reduced to about 15% (Fakhouri et al., 2015).

When the SNIa samples are employed for distance prediction, the general workflow consists of the following steps. Firstly, their light curve is corrected for dust extinction and K-corrected to rest-frame magnitude (Nugent et al., 2002; Hsiao et al., 2007). After that, the peak magnitude is adjusted for the observed color and light curve width. All these processes rely on an accurately estimated peak magnitude, color evolution and light curve shape. Either directly or indirectly, the uncertainty in these steps depends on an employed light curve model. These uncertainties will accumulate to the final uncertainty of distance prediction and cosmology parameter constraints. The performance of the employed light curve model is therefore critical for the accuracy of cosmology study.
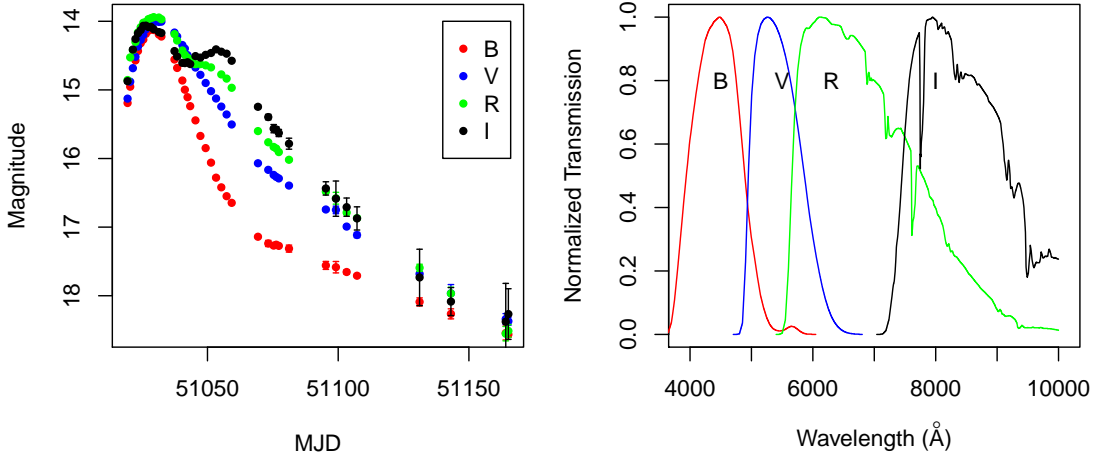
Figure 1.1: The left are the light curves of SN 1998dh observed by LOSS. The light curves are observed in four optical filters, B, V, R, I. The right plots the wavelength coverage (transmission) of the four filters.

The light curve models also underly many supernova classification algorithms (Kessler et al., 2010). Several classification schemes simply rely on naive spline fitting to individual light curve (Ishida and de Souza, 2013; Richards et al., 2012). After independent spline fitting, the fitted curve is vectorized on a grid for their next step analysis. This approach is problematic especially when the light curves are too sparse or with a low signal-to-noise ratio. A pre-specified fixed templates can be used to match actual observations. Newling et al. (2011) match the observation to a template with the gamma function shape. The templates can also be drived from observation empirically (Hamuy et al., 1996; Prieto et al., 2006). These empirical templates are parametrized by a single ΔM_{15}, and they are among the classical light curve models. Although these pre-specified templates are easy to use, they also neglect the diversity of light curve shapes and may suffer a low fitting quality for some light curves.

Several sophisticated light curve models exists in the astronomy literature. Among them are the popular SALT2 model (Guy et al., 2007), SiFTO model (Conley et al., 2008),

3

and MLCS2k2 model (Riess et al., 1996; Jha et al., 2007). The commonly used parameters of these light curve models include B band maximum epoch, B band maximum magnitude, a shape parameter and a color parameter. The SALT2 and SiFTO model build on spectrum template. On the other hand, the MLCS2k2 are empirically derived light curve model with second order correction terms. These methods also have a relatively rigid underlying structure, and still can not account for various light curve shapes. The MLCS2k2 also has vectorized template structure, as the naive spline fitting approach. Sampling curves to a grid lose accuracy, and it is cumbersome for modeling with the resulting vectors. This vectorized template approach also constitutes the hierarchical Bayesian model of Mandel et al. (2009) and Mandel et al. (2011).

Fortunately, the SNIa light curve observation fits into the well-established functional data analysis framework in statistics. For sparse and irregularly sampled functional data, there exists a collection of tools for dimensional reduction and the whole curve reconstruction. Among them is the functional principal component analysis (FPCA, see for example Ramsay and Silverman, 2010). With FPCA, each sample curve can be reconstructed by a linear combination of a mean function and several principal component functions. The principal component functions are orthonormal to each other and capture most of the variability in the dataset. Besides, the mean function and principal component functions can be constructed from spline, and therefore avoid the cumbersome effort to vectorize observations.

However, the FPCA can not be directly applied to SNIa light curves. Light curves of distinct supernovae are collected over years of survey. A curve registration step to align the peak is required before the analysis. The accuracy to estimate the peak is also the cornerstone for cosmology study. Moreover, when reconstructing a sparsely sampled light curve, the FCPA sometimes results in a weird and impossible SNIa light curve shape. Additional shape constraints should be built into the reconstruction process.

We will modify the functional principal component analysis for SNIa light curves. The resulting method incorporates the necessary curve registration and the shape constraints method proposed by Meyer (2012). We also develop a fast model training algorithm for FPCA. The model is trained through penalized least squares, with the nuclear norm penalty and the roughness penalty. These penalties encourage the low rank structure and smoothness of the solution.

The employed nuclear penalty is a popular tool for low rank structure learning (Cai et al., 2010; Chen et al., 2014; Rohde et al., 2011; Wright et al., 2009). Suppose the singular values of a matrix $\mathbf{A} \in \mathcal{R}^{p \times q}$ are $\sigma_1(\mathbf{A}) \geq \sigma_2(\mathbf{A}) \geq \cdots \geq \sigma_{p \wedge q}(\mathbf{A}) \geq 0$, then the nuclear penalty is defined as $\|\mathbf{A}\|_{S_1} = \sum_{i=1}^{p \wedge q} \sigma_j(\mathbf{A})$. This is a convex approximation to $\mathrm{rank}(\mathbf{A})$. An efficient algorithm in application of matrix completion is via soft thresholding of its singular values (Cai et al., 2010).

Our functional data modeling approach is purely data driven, without any specific structure for color and stretching. The resulting model from training dataset still provides abundant information about the intrinsic color and light curve shapes for SNIa. Through a numeric analysis on the real data set and simulated data set, we demonstrate an improvement in estimating peak magnitude and light curve shape.

1.2 Mira Variables and the M33 Survey*

The determination of reliable periods for variable stars has been an area of interest in astronomy for at least four centuries, since the discovery of the variability of Mira (*o* Ceti) by Fabricius in 1596 and the first attempts to determine its period by Holwarda & Bouillaud in the mid-1600s. The availability of electronic computers for astronomical research half a century ago enabled the development of many algorithms to estimate periods quickly

---

*Reprinted with permission from "Period estimation for sparsely-sampled quasi-periodic light curves applied to Mira" by Shiyuan He, et al., 2016. The Astronomical Journal, Volume 152, Number 6, 152–164. Copyright 2016 by the American Astronomical Society.

and reliably, such as Lafler and Kinman (1965); Lomb (1976); Scargle (1982).

The aforementioned algorithms work best in the case of periodic variations with constant amplitude and Mira variables present several challenges in this regard. While their periods of pulsation are stable except for a few intriguing cases (Templeton et al., 2005), Mira light curves can exhibit widely varying amplitudes from cycle to cycle (see, for example, the historical light curve of Mira compiled by Templeton and Karovska, 2009). In the case of C-rich Miras, the stochastic changes in mean magnitude across cycles (e.g., Marsakova, 1999) only complicate the problem further. The wide variety of light curves for long-period variables, already recognized by Campbell (1925) and Ludendorff (1928), may complicate the identification of Miras among other stars. Lastly, from a purely practical standpoint, it is simpler to obtain light curves spanning several cycles for RR Lyraes or Cepheids (with periods ranging from ~ 0.5 to ~ 100 d) than for Miras (with periods ranging from ~ 100 to ~ 1500 d).

Despite these challenges, the identification and determination of robust periods for Miras — especially in the regime of sparsely sampled, low signal-to-noise light curves — would be very beneficial for the determination of distances to galaxies of any type. Thanks to the unprecedented temporal coverage of the Large Magellanic Cloud (LMC) by microlensing surveys, the availability of large samples of extremely well-observed Miras has led to a thorough characterization of their period-luminosity relations at various wavelengths (Wood et al., 1999; Ita et al., 2004; Soszynski et al., 2007). The dispersion of the K-band period-luminosity relation (Glass and Evans, 2003, $\sigma = 0.13$ mag), is quite comparable to that of Cepheids at the same wavelength (Macri et al., 2015, $\sigma = 0.09$ mag) and makes them competitive distance indicators.

The third phase of the OGLE survey (Udalski et al., 2008) imaged most of the LMC with little interruption over 7.5 years and resulted in the discovery of 1663 Miras (Soszyński et al., 2009) with a median of 466 photometric measurements per object. The temporal

6
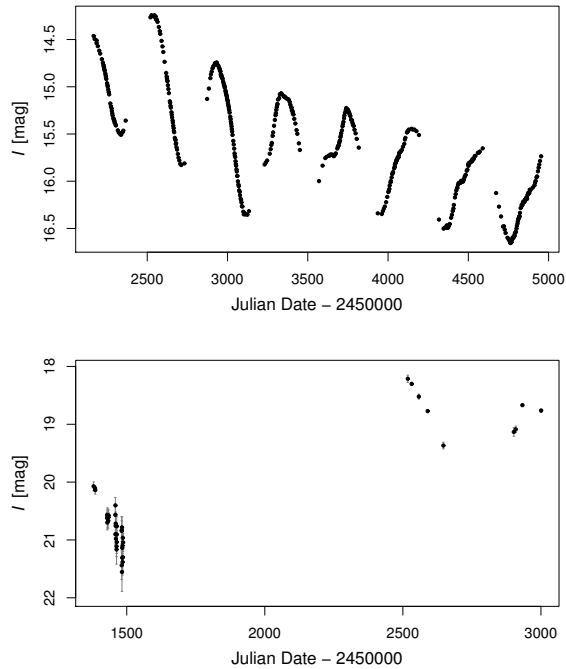
Figure 1.2: Representative Mira light curves observed by OGLE-III in the Large Magellanic Cloud (top) and DIRECT/M33SSS in M33 (bottom).

sampling of these light curves and their photometric precision are exceptional relative to typical astronomical surveys and make period estimation relatively easy. In comparison, a similar span of observations of M33 by the DIRECT (Macri et al., 2001) and M33SSS projects (Pellerin and Macri, 2011) in the I-band consists of a median number of 44 somewhat noisy measurements, heavily concentrated in a few observing seasons. Representative Mira light curves from the OGLE & DIRECT/M33SSS surveys are shown in Fig. 1.2. There are several reasons for the striking difference in quality between these two data sets. The LMC Miras are among the brightest objects in the OGLE fields, whereas their M33 counterparts are among the faintest in the aforementioned surveys of this galaxy. While the effective exposure times of all these surveys are quite comparable, after taking into account differences in collecting area of their respective telescopes, M33 lies approximately

7

6.2 mag farther in terms of its I-band apparent distance modulus. Furthermore, the main goal of the OGLE project (detection of microlensing events) requires a very dense temporal sampling of the survey fields; this is achieved by using a dedicated telescope and is helped by the fact that the LMC is observable nearly all year long from the site. In contrast, the observations of M33 were carried out using shared facilities (available only a few nights per month) with the primary purpose of studying Cepheids and eclipsing binaries (which do not require exceptionally dense temporal sampling), and the galaxy is only observable all night long for $\sim 1/3$ of the year. Standard period estimation algorithms, which work well for high signal-to-noise, well-sampled light curves such as those obtained by OGLE, will fail on more typical data sets represented by the M33 observations. The purpose of this work is to develop and test a methodology for estimating periods for sparsely sampled, noisy, quasi-periodic light curves such as those of Miras observed in M33 by the aforementioned projects.

2. FUNCTIONAL PRINCIPAL COMPONENT METHOD FOR TYPE IA SUPERNOVA LIGHT CURVES

In this chapter, we develop a light curve model based on functional principal component analysis. This model is incorporated with peak registration, shape constraints, as well as a fast training algorithm. We will demonstrate its power to capture the shape of light curves on real data and simulated data.

2.1 Review of FPCA

This section reviews the functional principal component analysis (FPCA). A thorough discussion can be found in Ramsay and Silverman (2010), Tran (2008) and Shang (2014)

Generally, the functional data analysis models random function X in $L^2(\mathcal{T})$, and the function X has domain $\mathcal{T}$. Its mean function and covariance function can be defined. The mean function is given by $\phi_0 = \mathrm{E}(X)$, and the covariance function $K(u,v) : \mathcal{T} \times \mathcal{T} \to \mathcal{R}$ by

$$C(u,v) = \mathrm{E}\left[(X(u) - \phi_0(u))(X(v) - \phi_0(v))\right] .$$

Suppose there exists an sequence of functions $(\phi_k)_{k=1}^{\infty}$ and a sequence of decreasing positive numbers $(\kappa_k)_{k=1}^{\infty}$ such that $C(u,v) = \sum_{k=1}^{\infty} \kappa_k \phi_k(u)\phi_k(v)$. The function ϕ_k's are orthonormal, which means $\langle \phi_k, \phi_j \rangle = \int \phi_k(u)\phi_j(u)\,\mathrm{d}u = \delta_{jk}$, and $\delta_{jk} = 1$ if $j = k$ and $\delta_{jk} = 0$ otherwise. The random function X has corresponding Karhunen-Loève expansion, $X(u) = \phi_0(u) + \sum_{k=1}^{K} \beta_k \phi_k(u)$. The β_k is a random variable with variance κ_k, and they are uncorrelated with each other $\mathrm{E}(\beta_k \beta_j) = \delta_{kj}\sqrt{\kappa_k \kappa_j}$.

For the sequence of (κ_k), suppose its first few elements have large values and then decrease very fast toward zero. To some extant, the variability related to $\phi_k(k > K)$ can be ignored for some value of $K > 0$. Functional PCA aims to estimate the leading compo-

nents of the covariance expansion and the Karhunen-Loève expansion. The truncated version corresponds to $C(u, v) = \sum_{k=1}^{K} \kappa_k \phi_k(u) \phi_k(v)$, and $X(u) = \phi_0(u) + \sum_{k=1}^{K} \beta_k \phi_k(u)$. The first K leading functions are called principal component functions, and the random variable β_k's are called principal component scores or scores.

Suppose there are n samples of the random function $X_1, \cdots, X_n$, and they are observed on a dense grid of $\mathcal{T}$. The mean and covariance function ca be directly estimated by the sample mean and sample covariance. The principal component functions can then be found by solving eigensystems, and finally the k-th score for the n-th sample are calculated via numerical integration $\beta_{nk} = \int X_n(u) \phi_k(u) \mathrm{d}u$. The smoothness of the estimated ϕ_k's can be encouraged by incorporating a roughness penalty $\int (\phi_k''(u))^2 \mathrm{d}u$ (Pezzulli and Silverman, 1993; Silverman et al., 1996).

When each sample is observed sparsely at irregular grid points, one estimation approach is to employ a reduced rank model that solved by the EM algorithm (James et al., 2000; Zhou et al., 2008; Kayano and Konishi, 2010). Peng and Paul (2009) proposed the restricted maximum likelihood by solving over Stiefel manifold. They noticed in their paper their algorithm is not guaranteed to converge and may be trapped in local optimal solution. The work of Yao et al. (2005) relied on local linear and quadratic regression to estimate the covariance function $C(\cdot, \cdot)$, and the scores are predicted by conditional expectation instead of numerical integration.

2.2 The FPCA for SNIa

This section develops the functional principal component analysis for SNIa. The FPCA method in statistics assumes the function is defined on a common domain $\mathcal{T}$. This section extends it with a curve registration mechanism. In particular, the model will convert a function of observation date to a function of phase relative to the peak. A fast algorithm for FPCA estimation with rank penalty and roughness penalty will be developed in

Section 2.3. The score prediction with shape constraints will be addressed in Section 2.4.

Suppose we need to model the light curves of S SNIa. For $s = 1, 2, \cdots, S$, let $l_{s\lambda}(t)$ be the light curve of the s-th supernova with band $\lambda \in \{B, V, R, I\}$. It is a function of the Julian date t. Let $b_{s\lambda}$ be the epoch of its peak magnitude $m_{s\lambda}$, i.e.,

$$m_{s\lambda} = \min_t l_{s\lambda}(t) = l_{s\lambda}(b_{s\lambda}). \tag{2.1}$$

The observed light curve is usually recorded at sparsely sampled time points and corrupted by noise. Suppose for this light curve there are totally $n_{s\lambda}$ observation points. They are observed at time $t_{s\lambda j}$ with magnitude $y_{s\lambda j}$ and magnitude uncertainty $\sigma_{s\lambda j}$ for $j = 1, 2, \cdots, n_{s\lambda}$. This time series of magnitude is an observation of the underlying light curve $l_{s\lambda}(t)$ corrupted by gaussian noise, i.e.,

$$y_{s\lambda j} = l_{sf}(t_{s\lambda j}) + \sigma_{sfj}\epsilon_{s\lambda j}.$$

The noise $\epsilon_{s\lambda j}$'s are assumed to follow the independent standard normal distribution.

The light curve around the peak epoch is of central interest for cosmology. We can transform the light curve $l_{s\lambda}(t)$, as a function of time t, to a function of phase q. Let z_s be the redshift of the corresponding supernova, and the transformation of time into phase is defined by $q = (t - b_{s\lambda})/(1 + z_s)$. The denominator $1 + z_s$ accounts for time dilation for SNIa with various redshifts. Our model concentrates on the phase range from $q_{min}(< 0)$ to $q_{max}(> 0)$ around the peak epoch ($q = 0$). Let us denote the light curve in phase q as $g_{s\lambda}(q) = l_{s\lambda}(t) = l_{sf}(q(1 + z_s) + b_{s\lambda})$ for $q \in [q_{min}, q_{max}]$.

From the truncated version of Karhunen-Loève expansion, each light curve in phase $g_{s\lambda}$ can be represented as a linear combination of the mean function and several principal

component functions,

$$g_{s\lambda}(q) = m_{s\lambda} + \phi_0(q) + \sum_{k=1}^{K} \beta_{s\lambda}^{(k)} \phi_k(q) \, . \qquad (2.2)$$

In the above, the $\beta_{s\lambda}^{(k)}$'s are the principal component scores. For each light curve, all its scores are denoted by a vector $\boldsymbol{\beta}_{s\lambda} = (\beta_{s\lambda}^{(1)}, \beta_{s\lambda}^{(2)}, \cdots, \beta_{s\lambda}^{(K)})^T$. We assume this score vector $\boldsymbol{\beta}_{s\lambda}$ is a random variable following $\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$ with unknown diagonal covariance matrix $\boldsymbol{\Sigma}$. As in classical FPCA, the principal component function $\phi_k(q)$'s are set to be orthonormal to each other, and they are ordered with decreasing ability to explain the observation. Furthermore, recall from (2.1) $m_{s\lambda}$ is the peak magnitude at phase $q = 0$. For the identifiability issue, we also require that $\phi_k(0) = 0$ for $k = 0, \cdots, K$.

Now we have finished the discussion of the proposed model structure (2.2), and we conclude this section by summarizing its components. In particular, its the mean function $\phi_0(q)$, the principal component functions $\phi_k(q)$'s, and the covariance matrix $\boldsymbol{\Sigma}$ of $\boldsymbol{\beta}$ will be trained from a dataset. When fitting a new light curve observation, the following parameters should be determined: (1) the peak magnitude m_{sf}; (2) the peak Julian date b_{sf}; (3) the scores vector $\boldsymbol{\beta}_{sf}$.

2.3 Model Training

This section develops the procedure of model training on a collection of SNIa light curves. The first step is to represent the model (2.2) by spline. Then after initial peak alignment, the mean function and principal component functions will be learned in a two-step procedure.

2.3.1 Spline Representation

In our work, all the template functions are represented by spline. This avoids the effort to vectorize the observation or template as in the work of Ishida and de Souza (2013),

Richards et al. (2012) and Jha et al. (2007).

The the spline basis is constructed from natural quadratic spline. We put $P - 2$ equally spaced knots in the phase interval $[q_{\min}, q_{\max}]$. These knots are denoted as $d_1, d_2, \cdots, d_{P-2}$. The natural quadratic spline is orthornormalized by a QR decomposition of its dense evaluation matrix. The resulting orthonormal spline basis is denoted as

$$\mathbf{b}(q) = \big(b_1(q), \ b_2(q), \ \cdots, \ b_Q(q)\big)^T$$

for $q \in [q_{\min}, q_{\max}]$. These spline bases satisfy $\int b_j(q)b_k(q)\mathrm{d}q = \delta_{jk}$, with $\delta_{jk} = 1$ if $j = k$ and $\delta_{jk} = 0$ if $j \neq k$. Let $\boldsymbol{\theta}_0$ be a P dimensional vector such that the mean function can be expressed as $\phi_0(q) = \mathbf{b}(q)^T \boldsymbol{\theta}_0$. Let $\boldsymbol{\Theta}_\phi$ be a $Q \times K$ matrix of spline coefficients, such that the K principal component functions can be represented by $\boldsymbol{\phi}(q)^T = (\phi_1(q), \cdots, \phi_K(q)) = \mathbf{b}(q)^T \boldsymbol{\Theta}_\phi$. The matrix $\boldsymbol{\Theta}_\phi$ is required to have orthonormal columns, i.e. $\boldsymbol{\Theta}_\phi^T \boldsymbol{\Theta}_\phi = \mathbf{I}_R$. This guarantees the K principal component functions are also orthonormal to each other. Now, the model (2.2) can be expressed by the spline basis as

$$
\begin{aligned}
g_{s\lambda}(q) &= m_{s\lambda} + \phi_0(q) + \sum_{k=1}^{K} \beta_{s\lambda}^{(k)} \phi_k(q) \\
&= m_{s\lambda} + \mathbf{b}(q)^T \boldsymbol{\theta}_0 + \mathbf{b}(q)^T \boldsymbol{\Theta}_\phi \boldsymbol{\beta}_{s\lambda} \,.
\end{aligned}
\tag{2.3}
$$

In this representation, the fixed and unknown parameters $\boldsymbol{\theta}_0$ and $\boldsymbol{\Theta}_\phi$ will be estimated from a training dataset. The parameters $m_{s\lambda}$, $b_{s\lambda}$ and $\boldsymbol{\beta}_{s\lambda}$'s are unique to each light curve. Notice the parameter $b_{s\lambda}$ is implicitly coded in phase q.

2.3.2 Initial Peak Alignment

Recall we need to transform all light curves into phase by the transformation $q_{s\lambda j} = (t_{s\lambda j} - b_{s\lambda})/(1 + z_{s\lambda})$. This transformation aligns all the light curves such that their peaks

are at phase zero. However, the peak epoch $b_{s\lambda}$ is unknown. An initial estimate of the peak magnitude $m_{s\lambda}$ and peak epoch $b_{s\lambda}$ are obtained by local quadratic regression. With this initial estimates, a two-step procedure is carried out for model training, i.e., separately learning the mean function $\phi_0(q)$ and learning the principal component functions $\phi_k(q), k = 1, \cdots, K$.

2.3.3 Learning the Mean Function

For the light curve of s-th supernova with band λ, If we stack related spline basis vectors into a matrix $\mathbf{B}_{s\lambda} = \left(\mathbf{b}(q_{i\lambda 1}), \cdots, \mathbf{b}(q_{i\lambda n_{s\lambda}}) \right)^T$, the model (2.3) in matrix form is

$$\mathbf{y}_{s\lambda} = m_{s\lambda} \mathbf{1}_{n_{s\lambda}} + \mathbf{B}_{s\lambda} \boldsymbol{\theta}_0 + \mathbf{B}_{s\lambda} \boldsymbol{\Theta}_\phi \boldsymbol{\beta}_{s\lambda} + \mathbf{W}_{s\lambda} \boldsymbol{\epsilon}_{s\lambda}, \tag{2.4}$$

where $\mathbf{1}_{n_{s\lambda}}$ is a vector of length $n_{s\lambda}$ with all elements equal to one, $\boldsymbol{\epsilon}_i$ is a random vector of length $n_{s\lambda}$ following a standard normal distribution, and $\mathbf{W}_{s\lambda} = \text{diag}\{\sigma_{s\lambda 1}, \cdots, \sigma_{s\lambda n_{s\lambda}}\}$ is a diagonal matrix of magnitude uncertainty. This representation summarizes all the light curve observations of the s-th supernova with band λ.

The last two terms in (2.4) has an expectation of zero. Although their covariance matrix is unknown, the general least square estimate is still consistent with least square estimation. With the estimated $\widehat{m}_{sf}$ and $\widehat{b}_{sf}$ in the previous subsection, the mean function $\phi_0(q)$ is learned by minimizing

$$\min_{\boldsymbol{\theta}_0} \sum_{s,\lambda} \frac{1}{n_{s\lambda}} \|(\mathbf{W}_{s\lambda})^{-1} (\mathbf{y}_{s\lambda} - \widehat{m}_{s\lambda} \mathbf{1}_{n_{s\lambda}} - \mathbf{B}_{s\lambda} \boldsymbol{\theta}_0)\|_2^2 + \mu \, \text{tr}\big(\boldsymbol{\theta}_0^T \boldsymbol{\Omega} \boldsymbol{\theta}_0\big), \tag{2.5}$$

where $\boldsymbol{\Omega} = \int \mathbf{b}''(t)^T \mathbf{b}''(t) dt$. The last term is the roughness penalty to encourage a smooth solution, and μ is the tuning parameter. Essentially, the smoothness is achieved by controlling the integral of the squared second order derivative of the solution, i.e.,

$\int (\phi_0''(q))^2 \mathrm{d}q = \mathrm{tr}\big(\boldsymbol{\theta}_0^T \boldsymbol{\Omega} \boldsymbol{\theta}_0\big)$. The explicit solution of $\boldsymbol{\theta}_0$ in (2.5) is

$$\widehat{\boldsymbol{\theta}}_0 = \left(\sum_{s,\lambda} (\mathbf{B}_{s\lambda})^T (\mathbf{W}_{sf})^{-2} \mathbf{B}_{s\lambda} + \lambda \boldsymbol{\Omega} \right)^{-1} \times$$

$$\left(\sum_{s,\lambda} (\mathbf{B}_{s\lambda})^T (\mathbf{W}_{s\lambda})^{-2} \big(\mathbf{y}_{s\lambda} - \widehat{m}_{s\lambda} \mathbf{1}_{n_{s\lambda}}\big) \right).$$

The resulting mean function is $\hat{\mu}(q) = \mathbf{b}(q)^T \widehat{\boldsymbol{\theta}}_\mu$.

2.3.4 Learning the Principal Component Functions

Now we discuss the algorithm to learn the principal component functions $\phi_k(q)$ for $k = 1, \cdots, K$. In the following, the rank K is encouraged to be small by a low rank penalty, and principal component functions are also encouraged to be smooth by the roughness penalty.

Now, with the estimated information from previous steps, align the peak and subtract the mean function from the observed light curves by $\widetilde{\mathbf{y}}_{s\lambda} = \mathbf{y}_{s\lambda} - \widehat{m}_{s\lambda} \mathbf{1}_{n_{s\lambda}} - \mathbf{B}_{s\lambda} \widehat{\boldsymbol{\theta}}_\mu$. This is the remaining difference to be fitted by the principal component functions, $\phi_k(q)$ $(k \geq 1)$.

Let $\mathbf{s}_{s\lambda} = \boldsymbol{\Theta}_\phi \boldsymbol{\beta}_{s\lambda}$ and put them a in $P \times 4$ matrix $\mathbf{S}_s = (\mathbf{s}_{sB}, \mathbf{s}_{sV}, \mathbf{s}_{sR}, \mathbf{s}_{sI})$ for the s-th supernova. Combine all of them in a $P \times (4S)$ matrix $\mathbf{S} = \big(\mathbf{S}_1, \mathbf{S}_2, \cdots, \mathbf{S}_S\big)$ for all supernovae. Notice the column $\mathbf{s}_{s\lambda}$'s of $\mathbf{S}$ are indexed by (s, λ), and the rank of the matrix $\mathbf{S}$ is $K \leq \min\{P, 4S\}$. This matrix is estimated by the following penalized least square,

$$\min_{\mathbf{S}} \sum_{s,f} \frac{1}{n_{sf}} \|(\mathbf{W}_{sf})^{-1} (\widetilde{\mathbf{y}}_{sf} - \mathbf{B}_{sf} \mathbf{s}_{sf})\|_2^2 + \mu_1 \|\mathbf{S}\|_{S_1} + \mu_2 \mathrm{tr}\big(\mathbf{S}^T \boldsymbol{\Omega} \mathbf{S}\big). \qquad (2.6)$$

In the above, the nuclear norm $\|\cdot\|_{S_1}$ serves to encourage a small value of K, and the last term encourages smoothly fitted light curves. The μ_1 and μ_2 are tuning parameters chosen by cross-validation.

The optimization problem (2.6) is solved by the ADMM algorithm (Boyd et al., 2011).

This is done by breaking the objective function into two parts: one part involving the quadratic forms and the other part involving the nuclear norm. Notice (2.6) is equivalent to

$$\min_{\mathbf{S}=\mathbf{P}} \sum_{s,\lambda} \frac{1}{n_{s\lambda}} \|\mathbf{W}_{s\lambda}^{-1}(\mathbf{y}_{s\lambda} - \mathbf{B}_{s\lambda}\mathbf{s}_{s\lambda})\|_2^2 + \mu_1 \|\mathbf{P}\|_{S_1} + \mu_2 \operatorname{tr}(\mathbf{S}^T \boldsymbol{\Omega} \mathbf{S}),$$

with an additional matrix $\mathbf{P}$, being forced to equal to $\mathbf{S}$. The augmented Lagrangian form of this problem is

$$\min_{\mathbf{S},\mathbf{P}} \sum_{s,\lambda} \frac{1}{n_{s\lambda}} \|\mathbf{W}_{s\lambda}^{-1}(\mathbf{y}_{s\lambda} - \mathbf{B}_{s\lambda}\mathbf{s}_{s\lambda})\|_2^2 + \mu_1 \|\mathbf{P}\|_{S_1} + \mu_2 \operatorname{tr}(\mathbf{S}^T \boldsymbol{\Omega} \mathbf{S})$$
$$+ \rho\|\mathbf{P} - \mathbf{S}\|_F^2 + \operatorname{tr}\left(\mathbf{U}^T(\mathbf{P} - \mathbf{S})\right),$$

where $\mathbf{U}$ is the Lagrangian parameter. The matrices $\mathbf{P}$ and $\mathbf{U}$ have the same size as $\mathbf{S}$, and their columns are also indexed by (s, λ) accordingly.

The ADMM algorithm then iterates through steps updating $\mathbf{S}$, $\mathbf{P}$ and $\mathbf{U}$ sequentially. This algorithm is summarized in Algorithm 1. In the algorithm, Line 2 updates the matrix $\mathbf{S}$ while $\mathbf{P}$ and $\mathbf{U}$ are fixed. The updating is through directly finding the minimizer of a quadratic form for each column of $\mathbf{S}$. Line 4 updates the matrix $\mathbf{P}$ while $\mathbf{S}$ and $\mathbf{U}$ are fixed. This is done with the singular value soft-thresholding operator (Cai et al., 2010). Finally, Line 5 updates the Lagrangian parameter $\mathbf{U}$.

Now, let $\widehat{\mathbf{S}}$ be the solution of the optimization problem (2.6), and $\widehat{\mathbf{S}} = \widehat{\mathbf{U}}\widehat{\mathbf{D}}\widehat{\mathbf{V}}^T$ be its SVD decomposition. Suppose $\widehat{\mathbf{u}}_j$ is the j-th column of $\widehat{\mathbf{U}}$ for $j = 1, 2, \cdots, \min\{P, 4S\}$. Define $\widehat{\boldsymbol{\Theta}}_\phi = (\widehat{\mathbf{u}}_1, \widehat{\mathbf{u}}_2, \cdots, \widehat{\mathbf{u}}_K)$ with the first K columns of $\widehat{\mathbf{U}}$, then the estimated principal component functions are $\widehat{\boldsymbol{\phi}}(s)^T = \mathbf{b}(s)^T \widehat{\boldsymbol{\Theta}}_\phi$. Also recall the columns of $\mathbf{S}$ are indexed by (s, λ), the rows of $\widehat{\mathbf{V}}$ are indexed in the same way corresponding to the SVD decomposition. Let $\widehat{\mathbf{v}}_{s\lambda}$ be the row vector of $\widehat{\mathbf{V}}$ indexed by (s, λ). For the light curve of s-th supernova with band λ, its score vector is computed as $\widehat{\boldsymbol{\beta}}_{s\lambda} = \widehat{\mathbf{D}}\widehat{\mathbf{v}}_{s\lambda}^T$, and the covariance

---
Algorithm 1 The ADMM algorithm for learning FPCA.

1: **repeat**
2:     Update $\mathbf{s}_{s\lambda}$ for $s = 1, 2, \cdots, S$ and $\lambda \in \{B, V, R, I\}$, by solving

$$\left(\mu_2\boldsymbol{\Omega} + \frac{1}{n_{s\lambda}}\mathbf{B}_{s\lambda}^T\mathbf{W}_{s\lambda}^{-2}\mathbf{B}_{s\lambda} + \rho\mathbf{I}\right)\mathbf{s}_{s\lambda} = \frac{1}{n_{s\lambda}}\mathbf{B}_{s\lambda}^T\mathbf{W}_{sf}^{-2}\mathbf{y}_{s\lambda} + \rho\mathbf{p}_{s\lambda} + \frac{1}{2}\mathbf{u}_{s\lambda} \, .$$

3:     Update $\mathbf{R} := \mathbf{S} - \frac{1}{2\rho}\mathbf{U}$.
4:     Update $\mathbf{P}$ by solving $\|\mathbf{P} - \mathbf{R}\|_F^2 + \frac{\mu_1}{\rho}\|\mathbf{P}\|_*$. With the SVD of $\mathbf{R} = \mathbf{U}\mathbf{D}\mathbf{V}^T$,

$$\mathbf{P} := \sum_{s\lambda}\left(d_{s\lambda} - \frac{\mu_1}{\rho}\right)_+ \mathbf{u}_{sf}\mathbf{v}_{sf}^T \, .$$

5:     $\mathbf{U} := \mathbf{U} + 2\rho \times \left(\mathbf{P} - \mathbf{S}\right)$.
6: **until** Convergence
---

matrix of $\boldsymbol{\beta}_{s\lambda}$ estimated by $\widehat{\boldsymbol{\Sigma}} = \sum_{s,\lambda}\widehat{\boldsymbol{\beta}}_{s\lambda}\widehat{\boldsymbol{\beta}}_{s\lambda}^T/(4S)$.

2.4   Fitting a New Light Curve

When our model fits a new light curve, the same procedure is applied to individual light curve across different filters and different supernovae. In the following, the subscript sf is dropped from the notation in Section 2.3 for simplicity.

Suppose we have a new light curve observation. Its magnitude $\mathbf{y} = (y_1, y_2, \cdots, y_n)^T$ is observed at time $t_1, \cdots, t_n$ with the standard deviation of measurement error $\sigma_1, \sigma_2, \cdots, \sigma_n$. The supernova with this light curve has redshift z.

Again, with the transformation $q_j = (t_j - b)/(1 + z)$ for $j = 1, \cdots, n$, the new light curve in phase is expanded by the template functions as

$$g(q) = m + \phi_0(q) + \sum_{k=1}^{K}\beta^{(k)}\phi_k(q) \, . \tag{2.7}$$

Define $\mathbf{B} = \big(\mathbf{b}(q_1), \cdots, \mathbf{b}(q_n)\big)^T$. The matrix form of the model (2.3) is

$$\mathbf{y} = m\mathbf{1}_n + \mathbf{B}\boldsymbol{\theta}_\mu + \mathbf{B}\boldsymbol{\Theta}_\phi\boldsymbol{\beta} + \mathbf{W}\boldsymbol{\epsilon}$$

where the noise $\boldsymbol{\epsilon}$ follows a standard normal distribution and $\mathbf{W} = \mathrm{diag}\{\sigma_1, \cdots, \sigma_n\}$ is a diagonal matrix representing measurement error of the magnitude. Recall the score vector $\boldsymbol{\beta} = (\beta^{(1)}, \cdots, \beta^{(K)})^T$ is regarded as random coefficient and it follows $\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$.

In Section 2.3, we have obtained estimates of $\boldsymbol{\theta}_0$, $\boldsymbol{\Theta}_\phi$ and $\boldsymbol{\Sigma}$ from a group of training samples. In this section, we treat them as fixed and known quantities. Our primary interests include: (1) to estimate the peak magnitude m and peak Julian date b, (2) to predict the random coefficient $\boldsymbol{\beta}$ for the new light curve, and (3) to assess all the associated uncertainty.

As regards to these tasks, the starting point is the joint distribution of the observed magnitude $\mathbf{y}$ and the score vector $\boldsymbol{\beta}$. They jointly follow a normal distribution, $\mathbf{z} = (\mathbf{y}^T, \boldsymbol{\beta}^T)^T \sim \mathcal{N}(\boldsymbol{\mu}_{\mathbf{z}}, \boldsymbol{\Sigma}_{\mathbf{z}})$, with mean and covariance,

$$\boldsymbol{\mu}_{\mathbf{z}} = \begin{pmatrix} \boldsymbol{\mu}_{\mathbf{y}} \\ \mathbf{0} \end{pmatrix}, \ \boldsymbol{\Sigma}_{\mathbf{z}} = \begin{pmatrix} \boldsymbol{\Sigma}_{\mathbf{y}} & \mathbf{B}\boldsymbol{\Theta}_\phi\boldsymbol{\Sigma} \\ \boldsymbol{\Sigma}\boldsymbol{\Theta}_\phi^T\mathbf{B}^T & \boldsymbol{\Sigma} \end{pmatrix}, \tag{2.8}$$

where $\boldsymbol{\mu}_{\mathbf{y}} = m\mathbf{1}_n + \mathbf{B}\boldsymbol{\theta}_\mu$ and $\boldsymbol{\Sigma}_{\mathbf{y}} = \mathbf{B}\boldsymbol{\Theta}_\phi\boldsymbol{\Sigma}\boldsymbol{\Theta}_\phi^T\mathbf{B}^T + \mathbf{W}^2$.

2.4.1 Peak Search

At first, the estimates of the peak date b and peak magnitude m are obtained by maximizing the marginal likelihood of $\mathbf{y}$. From (2.8), we know the marginal distribution of $\mathbf{y}$

is $\mathcal{N}(\boldsymbol{\mu_y}, \boldsymbol{\Sigma_y})$. The maximum likelihood estimates of m and b are computed by

$$
\hat{b}, \hat{m} = \arg\max_{m,b} \log p(\mathbf{y}|b, m), \tag{2.9}
$$

$$
\log p(\mathbf{y}|b, m) \propto -(\mathbf{y} - m\mathbf{1}_n - \mathbf{B}\boldsymbol{\theta}_\mu)^T \boldsymbol{\Sigma_y}^{-1}(\mathbf{y} - m\mathbf{1}_n - \mathbf{B}\boldsymbol{\theta}_\mu).
$$

Notice the matrix $\mathbf{B}$ implicitly depends on b through the phase transformation $q_j = (t_j - b)/(1 + z)$. The optimization (2.9) is evaluated by a profile likehood approach. At each value of b, the m has explicit solution by generalized least squares method, i.e.,

$$
m(b) = (\mathbf{1}_n^T \boldsymbol{\Sigma}_y^{-1} \mathbf{1}_n)^{-1} \mathbf{1}_n^T \boldsymbol{\Sigma}_y^{-1}(\mathbf{y} - \mathbf{B}\boldsymbol{\theta}_\mu).
$$

Plug in this $m(b)$ into (2.9) and compute $p(\mathbf{y}|b, m(b))$ at a dense grid of b. Suppose $p(\mathbf{y}|b, m(b))$ is maximized at $\hat{b}$. In this way, we get the optimizer of (2.9) as $(\hat{b}, \hat{m}) = (\hat{b}, m(\hat{b}))$.

2.4.2 Fitting Template

The next step involves the prediction of the random coefficients $\boldsymbol{\beta}$ to fit the shape of the light curve. With the joint normal distribution (2.8), it is well known that the best unbiased linear prediction of $\boldsymbol{\beta}$ given $\mathbf{y}$ is computed by the conditional expectation, $\mathrm{E}(\boldsymbol{\beta}|\mathbf{y}, \widehat{m}, \widehat{b})$. The uncertainty is quantified by the conditional variance $\mathrm{Var}(\boldsymbol{\beta}|\mathbf{y}, \widehat{m}, \widehat{b})$. Direct computation formulas for both of them are available as

$$
\mathrm{E}(\boldsymbol{\beta}|\mathbf{y}, \widehat{m}, \widehat{b}) = \boldsymbol{\Sigma}\boldsymbol{\Theta}_\phi^T \mathbf{B}^T \boldsymbol{\Sigma_y}^{-1}(\mathbf{y} - \widehat{m}\mathbf{1}_l - \mathbf{B}\boldsymbol{\theta}_\mu), \tag{2.10}
$$

$$
\mathrm{Var}(\boldsymbol{\beta}|\mathbf{y}, \widehat{m}, \widehat{b}) = (\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Theta}_\phi^T \mathbf{B}^T \mathbf{W}^{-2} \mathbf{B}\boldsymbol{\Theta}_\phi)^{-1}. \tag{2.11}
$$

Although $\mathrm{E}(\boldsymbol{\beta}|\mathbf{y}, \widehat{m}, \widehat{b})$ provides a direct way to predict the random coefficients $\boldsymbol{\beta}$, the final fitted light curve $g(q)$ in (2.7) could have weird light curve shape when there are not

enough data points. This sometimes happens at the left and right boundary of the phase interval $[q_{\min}, q_{\max}]$. Ideally, $g(s)$ should be monotone decreasing and concave in the phase interval $[q_{\min}, 0]$, where the supernova luminosity gradually increases. The monotoncity and concavity stated here are in terms of the numerical values of the magnitude. The light curve function in phase $g(q)$ should also be monotone increasing in the phase interval $[35, q_{\max}]$. In terms of derivatives, we have $g'(q) \leq 0$ and $g''(q) \leq 0$ for $s \in [q_{\min}, 0]$; meanwhile $g'(q) \geq 0$ for $s \in [35, q_{\max}]$.

To impose the above shape constraint, we adopt the method of Meyer (2012). Because our basis functions are constructed from natural quadratic spline (Section 2.3.1), we only need to impose constraint at the spline knots $d_1, d_2, \cdots, d_{P-2}$. The constraints $g'(q) \leq 0$ and $g''(q) \leq 0$ at $d_p \in [q_{\min}, 0]$ are equivalent to

$$g'(d_p) = \phi_0'(d_p) + \sum_{k=1}^{K} \beta^{(k)} \phi_k'(d_p) \leq 0, \tag{2.12}$$

$$g''(d_p) = \phi_0''(d_p) + \sum_{k=1}^{K} \beta^{(k)} \phi_k''(d_p) \leq 0, \tag{2.13}$$

for all the knots in the related phase interval $[s_{\min}, 0]$. At the same time, we have

$$g'(d_p) = \phi_0'(d_p) + \sum_{k=1}^{K} \beta^{(k)} \phi_k'(d_p) \geq 0 \tag{2.14}$$

for all the knots $d_p \in [35, q_{\max}]$. Notice (2.12), (2.13) and (2.14) impose a group of linear constraints on random coefficient $\boldsymbol{\beta}$.

Our final prediction of $\boldsymbol{\beta}$ is the one maximizing the conditional probability $p(\boldsymbol{\beta}|\mathbf{y}, \widehat{b}, \widehat{m})$ under the group of constraints (2.12), (2.13) and (2.14). Denote the conditional expectation $\mathrm{E}(\boldsymbol{\beta}|\mathbf{y}, \widehat{m}, \widehat{b})$ in (2.10) by $\mathbf{e}$, and the conditional variance matrix $\mathrm{Var}(\boldsymbol{\beta}|\mathbf{y}, \widehat{m}, \widehat{b})$ in (2.11) by $\mathbf{V}$. The distribution of $\boldsymbol{\beta}$ given $\mathbf{y}, \widehat{m}, \widehat{b}$ is multivariate normal $\mathcal{N}(\mathbf{e}, \mathbf{V})$. The

most probable $\boldsymbol{\beta}$ satisfying the shape constraints is obtained by

$$\widehat{\boldsymbol{\beta}} = \arg\max_{\boldsymbol{\beta}} \ \log p(\boldsymbol{\beta}|\mathbf{y}, \widehat{b}, \widehat{m})$$

$$= \arg\min_{\boldsymbol{\beta}} \ (\boldsymbol{\beta} - \mathbf{e})^T \mathbf{V}^{-1} (\boldsymbol{\beta} - \mathbf{e}),$$

subject to the constraints

$$\begin{cases} \sum_{k=1}^{K} \beta^{(k)} \phi_r'(d_p) \leq -\phi_0'(d_p), & d_p \in [q_{\min}, 0] \\ \sum_{k=1}^{K} \beta^{(k)} \phi_r''(d_p) \leq -\phi_0''(d_p), & d_p \in [q_{\min}, 0] \\ \sum_{k=1}^{K} -\beta^{(k)} \phi_j'(d_p) \leq \phi_0'(d_p), & d_p \in [35, q_{\max}]. \end{cases}$$

However, the conditional variance matrix (2.11) is no longer a valid quantification of the uncertainty of this $\widehat{\boldsymbol{\beta}}$. This issue will be addressed in Section 2.4.3.

The shape constraints above are chosen for our data set in Section 2.5.1. The constraints can be adjusted to specific light curve type and observation cadence. For example, to fit a B band light curve without a re-brightening second peak, it can impose the fitted light curve to be concave over the entire phase range, monotone decreasing over $[q_{\min}, 0]$, and monotone increasing over $[0, q_{\max}]$.

2.4.3 Fitting Uncertainty

In this subsection, we discuss the method to quantify the uncertainty of the estimates $\widehat{m}, \widehat{b}$ and the predicted $\widehat{\boldsymbol{\beta}}$. This is computed via the bootstrap method.

During the bootstrap, we fix the $\widehat{m}, \widehat{b}$ and $\widehat{\boldsymbol{\beta}}$ from the actual observation. Then several light curve samples are generated from a parametric bootstrap procedure. Suppose the total number of generated light curves is G. For $g = 1, 2, \cdots, G$, we generate a sequence

of magnitude $y_1^{(g)}, \cdots, y_n^{(g)}$ at the original observation time. They are computed as

$$y_j^{(b)} = \widehat{m} + \phi_0(q_j) + \sum_{k=1}^{K} \widehat{\beta}^{(k)} \phi_k(q_j) + \epsilon_j^{(g)}$$

where $\epsilon_j^{(b)}$ is sampled from $\mathcal{N}(0, \sigma_j^2)$ for $j = 1, \cdots, n$. The observation uncertainty for the g-th generated light curve $\mathbf{y}^{(g)} = (y_1^{(g)}, \cdots, y_n^{(g)})^T$ is unchanged as $\sigma_1, \cdots, \sigma_n$.

The same estimation and prediction procedure is applied to the generated light curve magnitude $y_1^{(g)}, \cdots, y_n^{(g)}$ with observation time $t_1, \cdots, t_n$ and uncertainty $\sigma_1, \cdots, \sigma_n$. Suppose its parameters are computed as $\widehat{m}^{(g)}, \widehat{b}^{(g)}, \widehat{\boldsymbol{\beta}}^{(g)}$. The standard deviation of all $\widehat{m}^{(1)}, \cdots, \widehat{m}^{(G)}$ is an estimation of the uncertainty of our actual estimate $\widehat{m}$. The uncertainty of $\widehat{b}$ and $\widehat{\boldsymbol{\beta}}$ is evaluated in a similar way.

2.5 Numerical Analysis and Methods Comparison

2.5.1 The Real Data Analysis

We select 115 relatively well-observed SNIa from the release of the Lick Observatory Supernova Search (LOSS) (Ganeshalingam et al., 2010), the Carnegie Supernova Project (CSP) (Contreras et al., 2010; Stritzinger et al., 2011), and the Harvard-Smithsonian Center for Astrophysics (CfA) (Hicken et al., 2009, 2012). Each selected sample has at least one observation within 5 days before the peak, and at least one observation within 5 days after the maximum. These SNIa samples are nearby supernovae, with cosmic microwave background (CMB) redshift $z < 0.1$. The light curve bands we use are four optical bands. A cross-filter K-correction is applied so that all the light curve magnitude are transformed to the standard rest-frame Kron-Cousins filters. This K-correction is performed using the SNooPy package of Burns et al. (2010). The Galactic extinction is corrected with the dust map of Schlafly and Finkbeiner (2011).

The proposed FPCA model is trained on the selected light curves, and constrained to
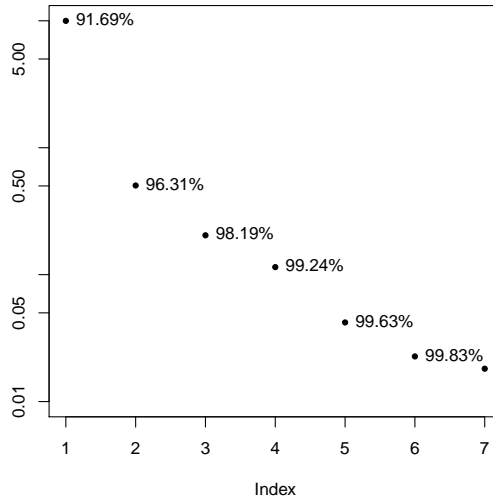
Figure 2.1: The first seven singular values and cumulatively explained proportion of variability. The first four bases together accounts for 99.24% variability in the dataset. The vertical axis is on the logarithmic scale.

the phase range $(-10, 40)$. The tuning parameters are selected by cross-validation. The number of basis functions is determined to be $R = 7$. Figure 2.1 plots the eigenvalues of the covariance matrix Σ of the scores β. The vertical axis is in the logarithmic scale. The percentage number is the cumulative proportion of variability explained in the dataset. For example, the first four bases together explained 99.24% of the total variability.

The estimated template functions are in Figure 2.2. Each panel shows the mean function and the effects of one basis function. The solid line is the the mean function $\phi_0(q)$. The "+" points represent $\phi_0(q) + 2\phi_j(q)$, and the "−" points represent $\phi_0(s) - 2\phi_j(s)$ for $j = 1, 2, 3, 4$. The first template basis curve $\phi_1(q)$ (in the upper left panel) reflects the the decline rate about 15 days after the peak. The second template basis curve $\phi_2(q)$ (in the upper right panel) is sensitive to the light curve width around the peak. Meanwhile, it also reflects a constrast of decline rate before and after 20 days in phase. The third template (in the lower left panel) basis curve adjusts the bump around 20 days after the peak.

The left panel of Figure 2.3 indicates the first two scores are nonlinearly correlated.

23
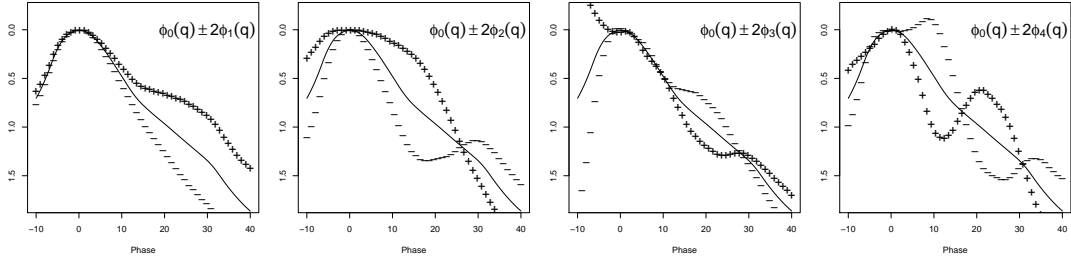
Figure 2.2: Plots of $\phi_0(q) \pm 2\phi_j(q)$ for the first four template basis functions. This shows how the template basis curves change the shape of the template mean curve. The solid line is the template mean curve. The "+" points represent $\phi_0(q) + 2\phi_j(q)$, and the "−" points represent $\phi_0(q) - 2\phi_j(q)$
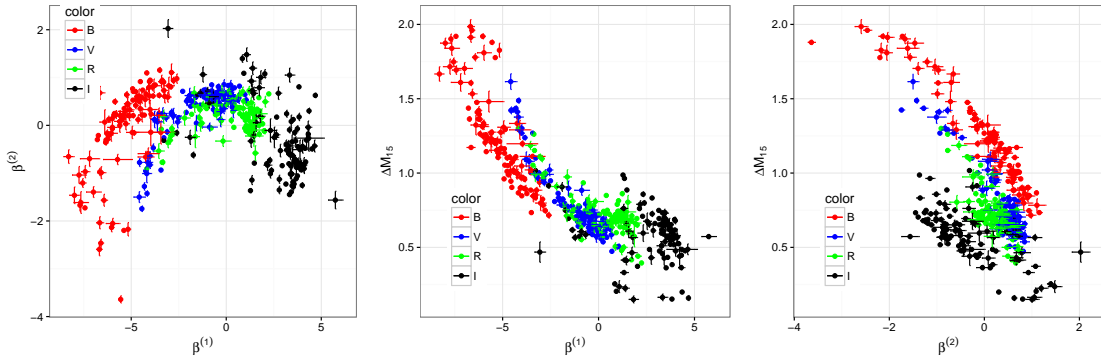
Figure 2.3: Relation of first the two scores the $\beta^{(1)}, \beta^{(2)}$ and ΔM_{15}. The left panel is the plot of $\beta^{(1)}$ and $\beta^{(2)}$; the middle panel is the plot of $\beta^{(1)}$ and ΔM_{15}; and the right panel is the plot of $\beta^{(2)}$ and ΔM_{15}. The B, V, R, I band are colored as red, blue, green and black respectively.

This suggests nonlinear dimensional reduction technique should be more effective, and this is left for future research. Notice in Figure 2.2, the first two basis functions adjust the width of different parts of the light curve. It is not a surprise that their scores $\beta^{(1)}$ and $\beta^{(2)}$ are correlated with the ΔM_{15} parameter (Phillips, 1993). It is shown as in the middle and right panels of Figure 2.3.

Figure 2.4 show the first and second scores are nonlinearly correlated the observed color $(B - V)_{\max}$ at the B band maximal. Phillips et al. (1999), Nugent et al. (2002) and
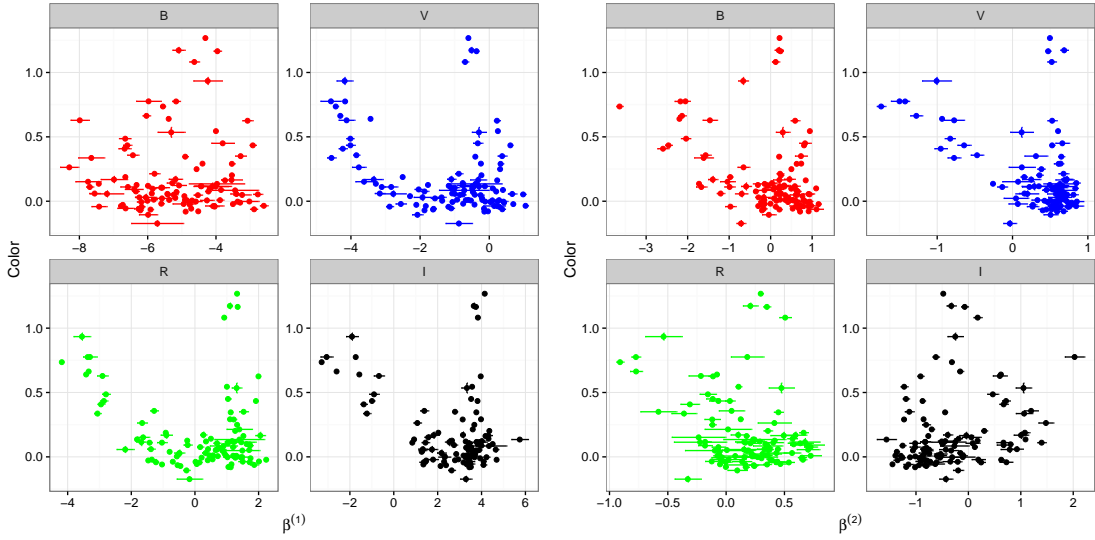
Figure 2.4: Plot of the first two scores $\beta^{(1)}, \beta^{(2)}$ against color.

etc have noticed the reddening affects the shape of the light curves, For V, R and I band light curves, *intrinsically* redder supernovae tend to have smaller value of $\beta^{(1)}$. At the same time, for B, V and R band light curves, *intrinsically* redder supernovae tend to have smaller value of $\beta^{(2)}$.

2.5.2 Method Comparison

Now, we compare the proposed FPCA model with several existing methods for SNIa in the literature. These methods include the popular SALT2 model (Guy et al., 2007), and the MLCS2k2 method (Jha et al., 2007). In addition, we will also compare with the naive independent spline fitting approach (Spline Method). In some works such as Ishida and de Souza (2013) and Richards et al. (2012), this approach provided the basis for their next step analysis. For this approach in our comparison, the natural cubic spline is used with six equally spaced knots inside the phase range $(-10, 40)$. Then spline is fitted to the observations with a roughness penalty, and the tuning parameter is chosen to minimize overall cross-validation error.

The quality of model fitting is assessed by cross-validation in the real dataset. In each round of the cross-validation procedure, one supernova is selected at random, and then for this supernova one observation point of each band light curve is randomly chosen and removed. For each round of the cross-validation, the following three steps are carried out

1. The model is trained on the all SNIa samples *except* the one selected for prediction.

2. The trained model is fitted to this selected supernova sample *without* the removed observation points.

3. The fitted light curves predict the magnitude at the removed observation points. This gives the absolute deviation error for the prediction.

This procedure is repeated 4,000 times, each time with one randomly selected supernova. The prediction errors are averaged for each band, and the standard errors are also computed. Lastly, note the model training step only applies to the FPCA method. The spline method is applied to the new data directly. The SALT2 and MLCS2k2 already have fixed trained models.

This cross-validation result is in Table 2.1. The table reports the mean absolute error (MAE), and its corresponding standard errors in parentheses. The results for the four bands are reported in four columns separately. Notice the proposed FPCA method consistently has smaller mean absolute CV error across all bands. The naive spline method has a close performance. For the selected SNIa samples, they have relatively low redshifts and high signal-to-noise ratio. The naive spline method takes the advantage of its flexibility and the data quality.

On the other hand, the proposed FPCA method captures the light curve shape much better than the well established SALT2 and MLCS2k2 methods. The latter two methods have a very rigid underlying struture. They can not guarantee the fitting quality for a

26

Table 2.1: Cross validation error (MAE) of light curve fitting on the real data. All numbers are of order 10^{-2} mag.

Method	Band			
	B	V	R	I
FPCA	4.002(0.086)	2.965(0.070)	3.072(0.056)	3.763(0.072)
Spline Method	4.615(0.119)	3.307(0.086)	3.443(0.087)	4.435(0.088)
SALT2	8.535(0.162)	6.877(0.110)	8.881(0.167)	19.818(0.250)
MLCS2k2	14.211(0.214)	13.264(0.173)	9.415(0.162)	15.891(0.211)

large number of SNIa light curves. For our dataset, some light curve fitting examples are presented in Figure 2.6. In the figure, the first, second and third columns correspond to the results of FPCA, SALT2, and MLCS2k2 methods, respectively. Each row corresponds to one selected SNIa sample. For some light curves, the fitted light curves of SALT2 and MLCS2k2 have high discrepancy with the observed light curve points.

Figure 2.5 shows the CV error against the phase. The lines are fitted locally for absolute CV errors against the phase of the prediction points. These plots show the model performance at various phase ranges across bands. It is easy to see the SALT2 and MLCS2k2 methods have relatively poor performance, especially for the I band light curves. The I band light curve may have a second bump around 15 days to 35 days in phase, as shown by the black points in Figure 2.6. The exact phase of this bump varies across different SNIa. This bump is difficult for light curve model to capture. The CV errors are highest during this phase for SALT2 and MLCS2k2 methods. Also notice, toward the later days in phase for all bands, especially when phase > 20, as the observation points get sparser. the CV errors grow larger for all methods. This is evident at the right end of all panels in Figure 2.5. Although the naive spline method takes the advantage of the high dataset quality, its performance deteriorates fast toward the right boundary.
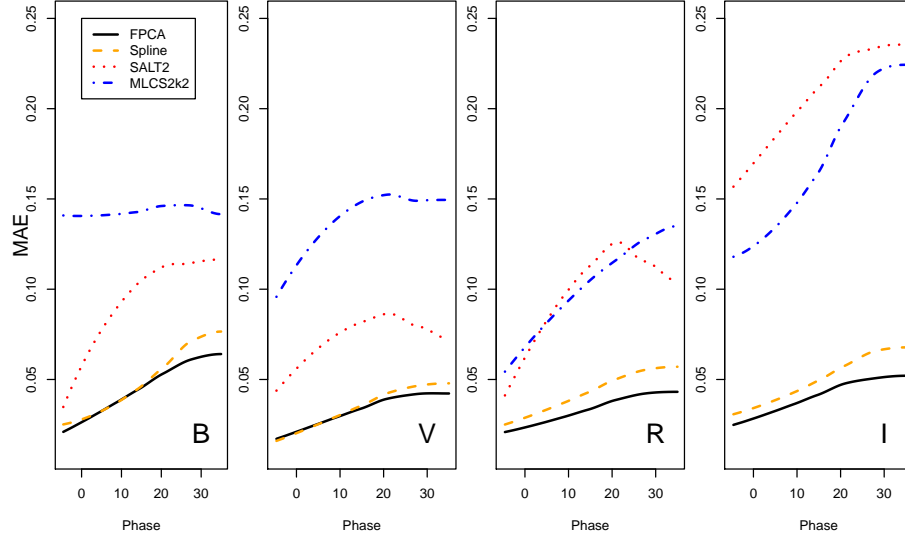
Figure 2.5: Compare methods on the real data set by cross-validation. The plots show the phase of the removed points against its prediction error. The lines are fitted locally for the prediction points selected in the cross-validation. The prediction error is measured by mean absolute error (MAE).

2.5.3   Simulation Comparison

We have assessed the four methods using the real dataset. In this subsection, we take a further analysis based on simulated dataset.

Estimating the peak magnitude m and the ΔM_{15} parameter is of core importance for the SNIa distance prediction models (Kattner et al., 2012; Wang et al., 2003). We wish to compare the light curve models' ability to estimate them. However, these quantities are unknown in the actual observations. In this subsection, a simulation study is carried out for comparison purpose. Note in practice, the peak magnitude and ΔM_{15} for B band light curve are the most commonly used. For completeness, these parameters of other bands are also considered in the simulation.

Generating SNIa light curve observation is not an easy task. For SNIa samples, the peak magnitudes, color, light curve shape, redshift and etc. are all correlated in a non-

(a) SN 1998de

(b) SN 1998dh
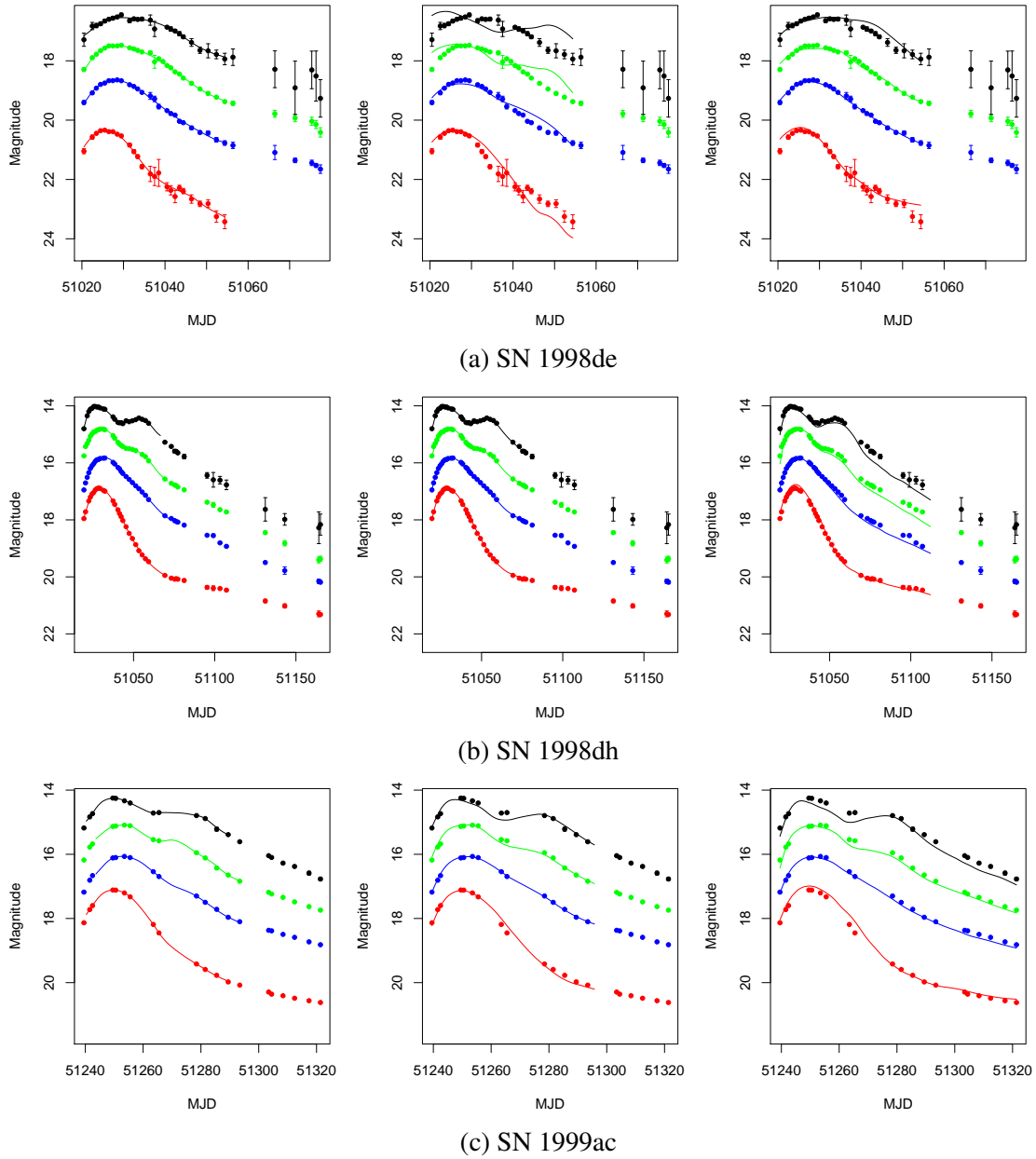
(c) SN 1999ac

Figure 2.6: Light curve fitting comparison for SNFT (first column), SALT (second column) and MLCS2k2 (third column). Four SNIa are illustrated here. From the first row to the last row are: SN 1998de, SN 1998dh, SN1999ac. The red, blue, green and black points are $B + 3$, $V + 2$, $R + 1$ and I band data.
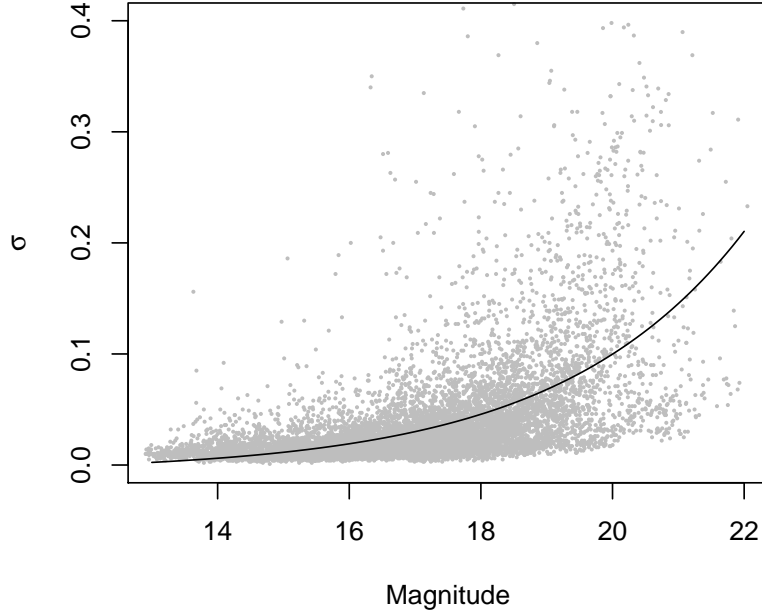
Figure 2.7: The relation between magnitude m and measurement uncertainty σ. An empirical relation $\sigma = a^{m-b} + c$ is fitted with $a = 1.428$, $b = 26.288$ and $c = 0.006$.

trivial way. Their observation cadence and noise are also affected by the weather and equipment condition. In this simulation, we take the approach to generate light curve of different bands separately. The generated light curves are made to match actual senario as much as possible.

To generate the B band light curve, the empirical model (2.7) trained in Section 2.5.1 is exploited. For the fairness to compare methods, after generating a collection of simulated light curves, the trained empirical model is blind to all methods. The FPCA method is re-trained on the simulated light curves.

To generate one fake light curve with trained model (2.7), we need to (1) specify its peak magnitude $\bar{m}$, its redshift $\bar{z}$ and its sampling cadence in phase $\bar{q}_1, \cdots, \bar{q}_n$; (2) to generate its score vector $\bar{\boldsymbol{\beta}}$, and then compute the magnitude values at the sample cadence; (3) to add noise to the computed magnitude, to get the observed magnitude $\bar{y}_1, \bar{y}_2, \cdots, \bar{y}_n$.

Table 2.2: Simulation result. Estimation accuracy of the peak magnitude m and ΔM_{15} of each band. The accuracy is measured in mean absolute error. The L_2 loss of the whole reconstructed light curve is also reported.

Method	B			V		
	m	ΔM_{15}	L_2	m	ΔM_{15}	L_2
FPCA	0.013	0.032	0.297	0.012	0.041	0.348
	(0.001)	(0.003)	(0.021)	(0.001)	(0.005)	(0.030)
Spline Method	0.028	0.078	0.872	0.021	0.070	0.919
	(0.003)	(0.004)	(0.042)	(0.002)	(0.005)	(0.042)
SALT	0.015	0.093	0.622	0.015	0.089	0.587
	(0.000)	(0.003)	(0.013)	(0.000)	(0.002)	(0.012)
MLCS2k2	0.019	0.109	0.631	0.020	0.103	0.622
	(0.001)	(0.003)	(0.011)	(0.001)	(0.003)	(0.011)
Method	R			I		
	m	ΔM_{15}	L_2	m	ΔM_{15}	L_2
FPCA	0.014	0.039	0.330	0.013	0.038	0.275
	(0.001)	(0.003)	(0.030)	(0.001)	(0.004)	(0.016)
Spline Method	0.026	0.076	0.984	0.027	0.078	0.866
	(0.003)	(0.005)	(0.046)	(0.004)	(0.005)	(0.039)
SALT	0.014	0.095	0.614	0.016	0.100	0.624
	(0.000)	(0.003)	(0.013)	(0.001)	(0.003)	(0.013)
MLCS2k2	0.020	0.111	0.667	0.020	0.110	0.634
	(0.001)	(0.003)	(0.011)	(0.001)	(0.003)	(0.012)

For the first part, one B band light curve is randomly selected from the real data set in Section 2.5.1. Its peak magnitude, redshift and sampling cadence in phase are reserved to generate the simulated light curve. This procedure also helps to match the real world relation between the redshift and peak magnitude. The actual sampling cadence is directly employed, because the result will have the closest match to actual survey conditions. Besides, for each generated light curves, it has peak epoch $\bar{b} = 0$, such that the observed time points are $\bar{t}_1 = \bar{z}\bar{q}_1, \cdots, \bar{t}_n = \bar{z}\bar{q}_n$. The peak epoch information $\bar{b} = 0$ is blind to all methods.

In the second step, we need to compute the magnitudes at selected sampling cadence

in phase. However, the light curve shape has not been specified yet. With the empirical model trained in Subsection 2.5.1, we only need to generate the score vector $\boldsymbol{\beta}$ to specify the light curve shape. In particular, the B band score vector is drawn from a multivariate normal distribution $N(\hat{\boldsymbol{\mu}}_B, \hat{\boldsymbol{\Sigma}}_B)$. The mean $\hat{\boldsymbol{\mu}}_B$ and the covariance matrix $\hat{\boldsymbol{\Sigma}}_B$ of B band scores are obtained from the training data in Subsection 2.5.1. They are computed by $\hat{\boldsymbol{\mu}}_B = \frac{1}{S}\sum_s \hat{\boldsymbol{\beta}}_{sB}$ and $\hat{\boldsymbol{\Sigma}}_B = \frac{1}{S-1}\sum_s (\hat{\boldsymbol{\beta}}_{sB} - \hat{\boldsymbol{\mu}}_B)(\hat{\boldsymbol{\beta}}_{sB} - \hat{\boldsymbol{\mu}}_B)^T$ for $\lambda = B$, where the B band score vector $\hat{\boldsymbol{\beta}}_{sB}$ belongs to the s-th supernova in the training data set. With a $\bar{\boldsymbol{\beta}}_B$ sampled from this distribution, the constructed whole light curve signal is

$$g(q) = \bar{m} + \hat{\phi}_0(q) + \sum_{k=1}^{K} \bar{\beta}_B^{(k)} \hat{\phi}_k(q)$$

where $\hat{\phi}_0(q), \cdots, \hat{\phi}_K(q)$ are from the empirical model Subsection 2.5.1. Now, we are able to compute the signal magnitude at the sampling cadence in phase $\bar{q}_1, \cdots, \bar{q}_n$. The signal corresponding magnitude is computed as $g(\bar{q}_1), \cdots, g(\bar{q}_n)$.

The last step is to add noise to the signal magnitude. Figure 2.7 plots the sigma-magnitude relation for all the observation points in the real data set. Generally, larger uncertainty sigma values are associated with larger magnitude values. The uncertainty is also affected by actual weather and equipment condition. An empirical relation $\sigma = a^{m-b} + c$ is fitted with $a = 1.428$, $b = 26.288$ and $c = 0.006$. This is shown as the black curve in Figure 2.7. From this empirical relation, the $\bar{\sigma}_j$ value can be computed for each signal magnitude value $g(\bar{q}_j)$. Then a gaussian noise to the signal magnitude value via $\bar{y}_j = g(\bar{q}_j) + \bar{\sigma}_j \bar{\epsilon}_j$, where $\bar{\epsilon}_j \sim \mathcal{N}(0, 1)$.

Now, from the above procedure, we generate a light curve with sampling cadance $\bar{t}_1, \cdots, \bar{t}_n$, observed magnitude $\bar{y}_1, \cdots, \bar{y}_n$ and magnitude uncertainty $\bar{\sigma}_1, \cdots, \bar{\sigma}_n$. This procedure is repeated to generate 2,000 B band light curves.

All the methods are applied these light curves, with only the information of the sam-

pling cadence, observed magnitudes and magnitude uncertainties. We assess methods'
ability to estimate the peak magnitude $\bar{m}$ and ΔM_{15}. The accuracy is measured by the
mean absolute error (MAE). In addition, we also assess methods' ability to reconstruct the
whole light curve, and this accuracy is measured by L_2 loss. The L_2 loss is the integral of
the squared difference between the true light curve and reconstructed light curve, and the
integral is taken over the phase range $(-10, 40)$.

Table 2.2 presents the result of the simulation. This simulation procedure is also re-
peated for V, R, I band separately. The message is the same as in the previous subsection.
The proposed FPCA method consistently produces more accurate estimates of the peak
magnitude m and the ΔM_{15} parameter, as well as better reconstruction of the whole light
curve. The empirical relation fitted in Figure 2.7 makes the generated light curves have
relatively lower signal to noise ratio than the real dataset. This makes the naive spline
method has the worst performance among all methods.

3.   SNIA CHARACTERIZATION AND DISTANCE MODELLING

In the previous chapter, we have dicussed a funcitonal principal component model for SNIa light curves. The obtained principal component scores parametrize the shape of the light curve and provide abundant information of SNIa. The primary goal of this chapter is to explore the potential of the extracted scores in explaining important physical quantities, such as intrinsic color, interstellar dust reddening, spectral line strength, and spectral classes. We show that our FPCA light curve model provides a flexible and effective light curve shape characterization. One direct application is a better constrained color excess estimation. Moreover, by exploring the relations between the scores and spectral features, we open the possibility of inferring SNIa spectral information from purely light curve data and therefore provide the opportunity of more precise K-correction and distance prediction within subgroups of SNIa. Finally, our light curve model is used to produce a new light curve shape parameterization as a better constrained alternative to the classical ΔM_{15} parameter in distance prediction.

3.1   Estimation of Color Excess

This section explores the potential of using the principal component scores to determine the intrinsic color and color excess of SNIa. The first panel in Figure 2.3 exhibits a nonlinear relation between the first two dominant scores $\beta^{(1)}, \beta^{(2)}$. Thus, a more effective dimension reduction is possible for SNIa light curves with nonlinear dimension reduction techniques. Both $\beta^{(1)}$ and $\beta^{(2)}$ are correlated with color. We would expect all the color information could be absorbed into one score after the nonlinear dimension reduction.
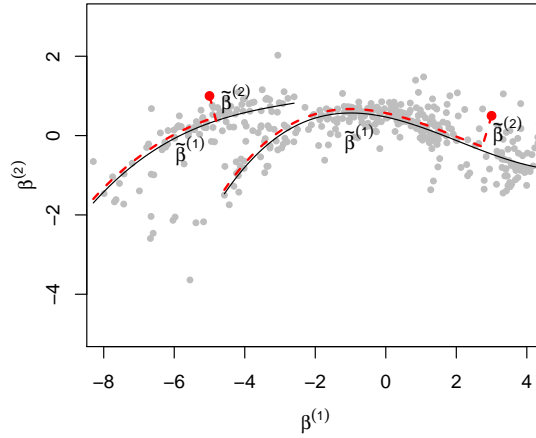
Figure 3.1: The nonlinear scores $\tilde{\beta}^{(1)}, \tilde{\beta}^{(2)}$ are calculated via fitting curves to the original scores $\beta^{(1)}, \beta^{(2)}$.
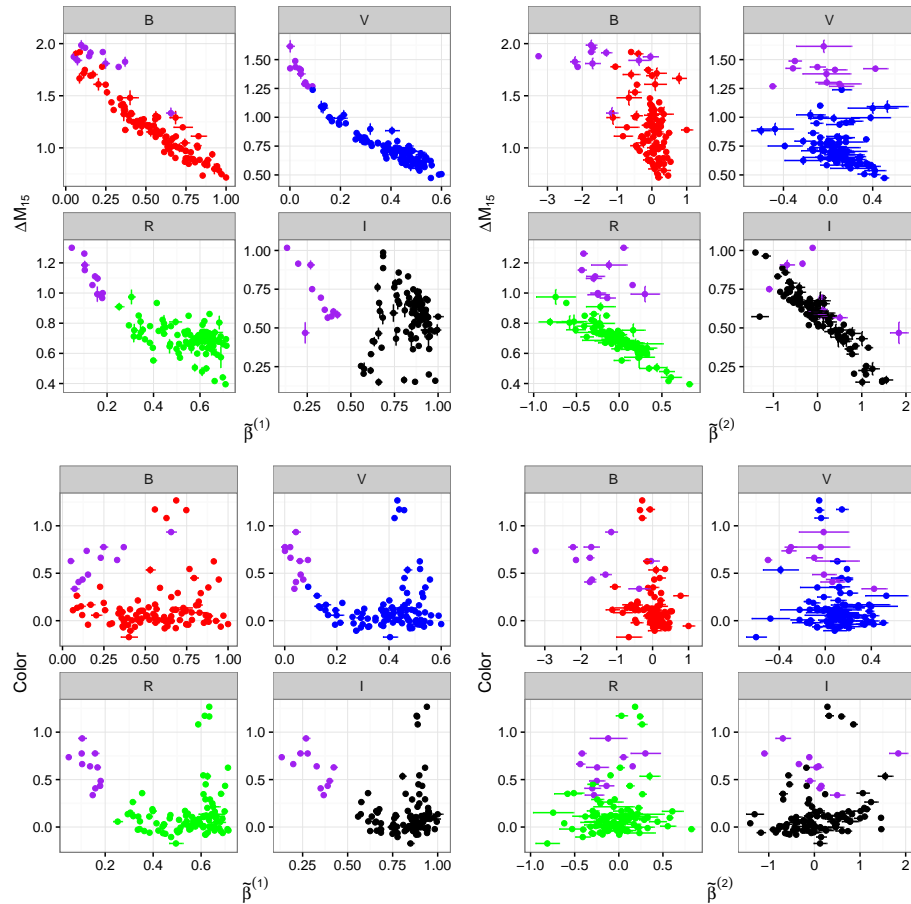
Figure 3.2: Relation between the nonlinear scores $\tilde{\beta}^{(1)}, \tilde{\beta}^{(2)}$ and ΔM_{15}, color. The SNIa with $\tilde{\beta}_R^{(1)} < 0.2$, shown as purple points, are SN 1991bg-like events.
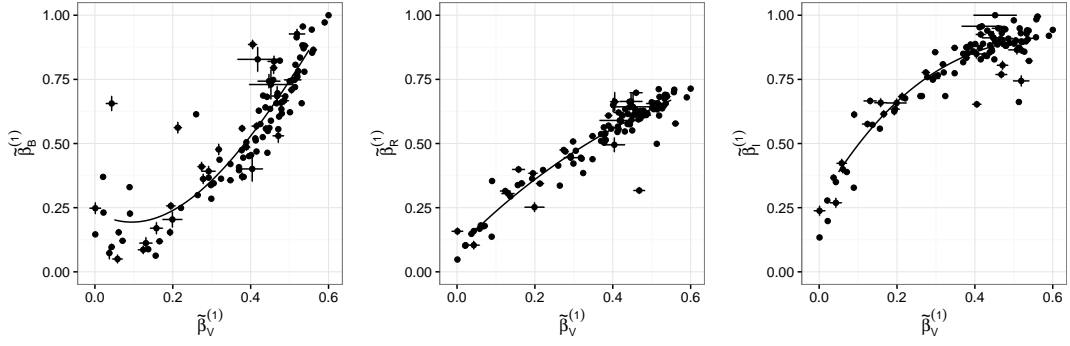
Figure 3.3: The relation between the first nonlinear V band score $\tilde{\beta}_V^{(1)}$ (on horizontal axes) with $\tilde{\beta}_B^{(1)}, \tilde{\beta}_R^{(1)}, \tilde{\beta}_I^{(1)}$ (on vertical axes).

3.1.1 Nonlinear Dimension Reduction and Nonlinear Scores

For simplicity, consider the two dimensional space of $\beta^{(1)}$ and $\beta^{(2)}$. These first two dimensions alone already account for 96.31% cumulative variance of the dataset. We adapt the concept of principal ridge (Ozertem and Erdogmus, 2011). A curve is fitted for $\beta_B^{(1)}$ and $\beta_B^{(2)}$ of B band, as the left black solid line in Figure 3.1. This nonlinear curve is treated as the first nonlinear dimension for B band scores. The second nonlinear dimension is the one locally perpendicular to the curve. Given the original scores $\beta^{(1)}, \beta^{(2)}$ from our model, the new *nonlinear scores* $\tilde{\beta}^{(1)}, \tilde{\beta}^{(2)}$ are calculated as follows. Consider the left red point in Figure 3.1, it is projected onto the B band curve. The projection is identified by the nearest point on the curve. The nonlinear score $\tilde{\beta}^{(1)}$ is the geodesic distance from the leftmost point of the curve to the projection point. The geodesic distance along the curve is indicated by the red dashed curve. The new score $\tilde{\beta}^{(2)}$ is the usual Euclidean distance of the original point to the projection point, as indicated by the red vertical dashed line. In a similar manner, a curve is fitted for V, R, I band together (the right solid curve in Figure 3.1), and the linear scores for V, R, I band are projected onto this curve to obtain the new nonlinear score.
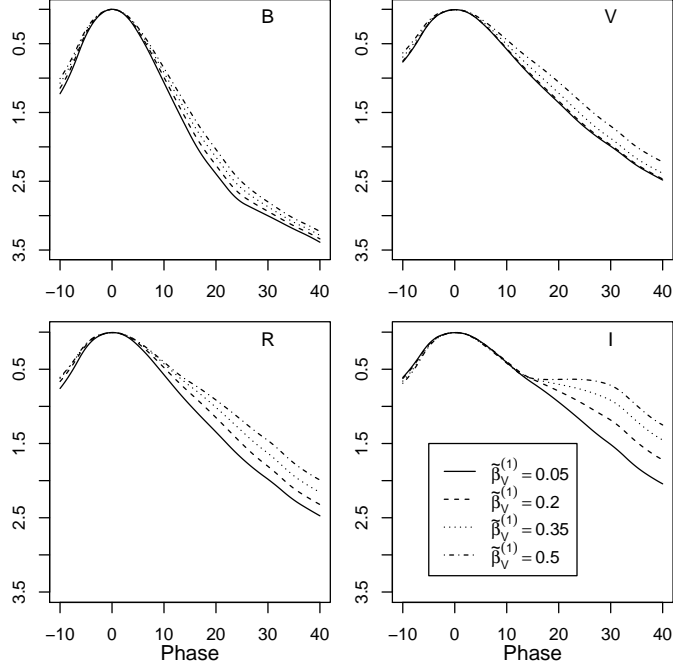
Figure 3.4: Average light curve shapes parametrized by the nonlinear score $\tilde{\beta}_V^{(1)}$.

In the original linear system, the first dimension explains 91.69% of the total variability. The first two dimension together explains 96.31% of the total variability. Now in the nonlinear system, the explained proportion of the first nonlinear score should be larger than 91.69%, but smaller than 96.31%. More light curve information is absorbed into the first dimension. The first score alone can provide adequate fit to SNIa light curves. Figure 3.2 plots the relation of the nonlinear scores $\tilde{\beta}^{(1)}, \tilde{\beta}^{(2)}$ with ΔM_{15} and the color $(B - V)_{\max}$. Compared with the orignal score $\beta^{(2)}$, the second nonlinear score $\tilde{\beta}^{(2)}$ shows less correlation with the color and ΔM_{15} parameters.

The more thorough dimension reduction can be made by describing the correlation among $\tilde{\beta}_B^{(1)}, \tilde{\beta}_V^{(1)}, \tilde{\beta}_R^{(1)}, \tilde{\beta}_I^{(1)}$. The shape of light curves across these optical bands should be well correlated due to their common spectral evolution. The relation of $\tilde{\beta}_V^{(1)}$ with $\tilde{\beta}_B^{(1)}, \tilde{\beta}_R^{(1)}, \tilde{\beta}_I^{(1)}$ is plotted in Figure 3.3. For most of the cases, the V band nonlinear score

$\tilde{\beta}_V^{(1)}$ is a reliable predictor of $\tilde{\beta}_B^{(1)}$, $\tilde{\beta}_R^{(1)}$ and $\tilde{\beta}_I^{(1)}$. Therefore we can use a single parameter, $\tilde{\beta}_V^{(1)}$, to describe all the SNIa light curve shapes. For each $\tilde{\beta}_V^{(1)}$, the predicted values of $\tilde{\beta}_B^{(1)}$, $\tilde{\beta}_R^{(1)}$ and $\tilde{\beta}_I^{(1)}$ are obtained by the fitted curves in Figure 3.3. This single parameterization of light curve shape is especially useful for fitting high redshift supernovae with sparse and noisy observations, because only one parameter is required to be constrained by the data. Using this scheme with a single parameter $\tilde{\beta}_V^{(1)}$, Figure 3.4 depicts all four band light curve shapes for $\tilde{\beta}_V^{(1)} = 0.05, 0.2, 0.35, 0.5$. These light curves are the expected shapes accounting for first order correction of the mean curve with $\tilde{\beta}_V^{(1)}$.

3.1.2 Nonlinear Scores and Color Excess

Figure 3.2 reveals a relation between the observed color at B maximum, $(B - V)_{\mathrm{max}}$ and the first nonlinear score $\tilde{\beta}^{(1)}$, especially for V, R, I band light curves. This relation can be exploited to obtain an estimate of the color excess of the supernova. For example, here we use the relationship between the R band nonlinear score $\tilde{\beta}_R^{(1)}$ and the observed color at B maximum. As in Figure 3.5, we try to estimate a lower envelope (the black solid line), and treat it as an extinction free curve for SNIa. This lower envelope is estimated by lower 10% quantile regression with B-spline basis. The quantile regression is iterated by removing points with large positive residuals. This lower envelop $e_R(\tilde{\beta}_R^{(1)})$ as a function of R band $\tilde{\beta}_R^{(1)}$ serves to estimate the intrinsic color $(B - V)_0$ of the supernova. The color excess is obtained by

$$E(B - V) = (B - V)_{\mathrm{max}} - e_R(\tilde{\beta}_R^{(1)}) \,. \tag{3.1}$$

In other words, the color excess is the vertical distance from the observation points to the lower envelope, as illustrated in Figure 3.5.

The same method can be used to estimate extinction using similar relations appeared in other filter bands. The values and precisions from these different measurements however,

can be quite different. For example, the lower bound to the B band $\tilde{\beta}_B^{(1)}$ shows very little correlation with color, and one would get an estimate of $E(B - V)$ by approximately assuming the intrinsic $B - V$ is nearly zero. In the V, R, I bands, $\tilde{\beta}^{(1)}$ appears to produce very good intrinsic color estimators.

The popular method to estimate the color excess is from the work of Phillips et al. (1999). They used the empirical linear relation of the intrinsic color $(B - V)_0$,

$$(B - V)_0 = 0.725 - 0.0118 \, (q_V - 60) \,,$$

for the phase q_V with respect to the V band maximum. This linear relation holds for $30 \leq q_V \leq 90$. The color excess can be estimated via the observed color minus the $(B - V)_0$ above. The observed color should be corrected by K-correction and Galactic reddening. Figure 3.6 compares the color excess computed via this classical method and the color excess estimated from the R band $\tilde{\beta}_R^{(1)}$ at a reference phase $q_V = 35$. The figure shows that the color excess given by the method of Phillips et al. (1999) has negative values for a considerable portion of the supernovae. Riess et al. (1998) eliminated this problem by imposing a one-sided (positive-sided) Gaussian prior with zero mean and sigma 0.3. The figure also shows that the uncertainty deduced from $\tilde{\beta}_R^{(1)}$ is much smaller than that based on Phillips et al. (1999). This is not surprising, because the method of Phillips et al. (1999) uses only light curve observations beyond 30 days after V band maximum, which are usually very sparse, while the estimation of $\tilde{\beta}_R^{(1)}$ uses observations from the whole light curve.

3.2   Spectral Information

This section examines the relationship between the scores in model (3) and spectral features and discuss the possibility of using the scores for identifying spectral classes. There exists some analysis regard light curve width and spectral features such as Si II $\lambda 4000$
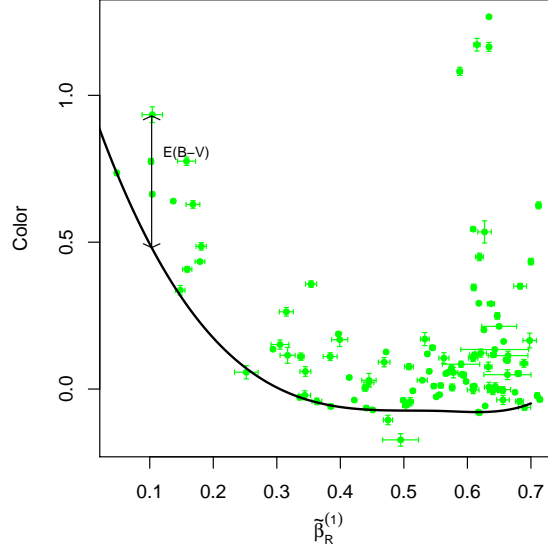
Figure 3.5: The lower envelope (black solid line) for the observed color and R band non-linear score $\tilde{\beta}_R^{(1)}$. The color excess is estimated by the vertical distance of the observation points to the envelope line.

in the literature. With the aid of model (2.2), we are able to present more details on how light curve shape (not just its width) changes with spectral features. We will also demonstrate the light curve scores can be linked to spectroscopically different SNIa, although with limited precision. This linkage is important. With refined spectral sub-classification, the K-correction can be applied with higher precision. Identifying sub-classes of SNIa based is of ultimate importance in assessing systematic evolutionary effect in applying them as standard candles.

3.2.1 The Scores and Spectral Features

Now we compare the first four scores across all bands with the spectral features. The dataset of SNIa with comprehensive spectral data is even more sparse, and the measurement of the strength of spectral features usually suffers from severe systematic errors due to difficulties in defining the level of continuum and observational noise, the latter is usu-
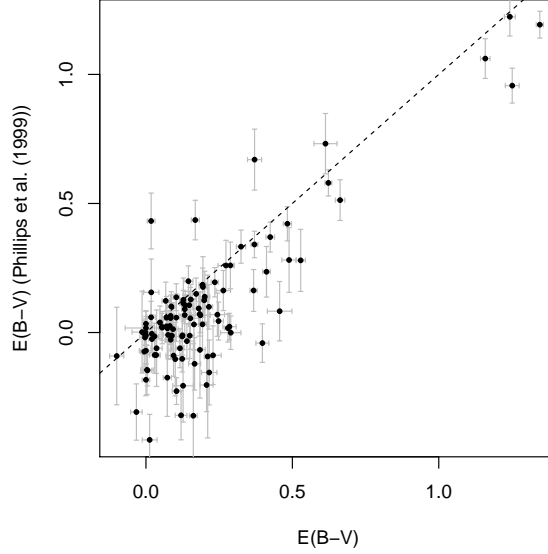
Figure 3.6: The y-axis is the color excess computed from Phillips et al. (1999), and the horizontal axis is the color excess computed from the R band score $\tilde{\beta}_R^{(1)}$.

ally not even available for most published SNIa spectra. Wagers et al. (2010) developed a mathematical framework based on wavelet decomposition of the spectra to reconstruct the noise and signal from published data. It was shown in Wagers et al. (2010) that large noise can easily bias estimate of spectral line strength, and Monte-Carlo simulation can be used to simulate the effect and correct the bias. However, there is no overlap of the SNIa sample in Wagers et al. (2010) and the current sample. A more recent derivation of spectral line strength is given in Zhao et al. (2016), but its data sample size is small.

High quality spectral feature measurements are scarce. They provide information on the details of spectral features which can significantly affect the measured colors and lead to abnormal extinction behavior. We show in this paper the correlations of the most significant spectral features with the light curve features produced from our method. We emphasize that spectral features provide further constraints on the properties of the intrinsic properties of supernovae and associated extinction. Measurement of spectral features

41
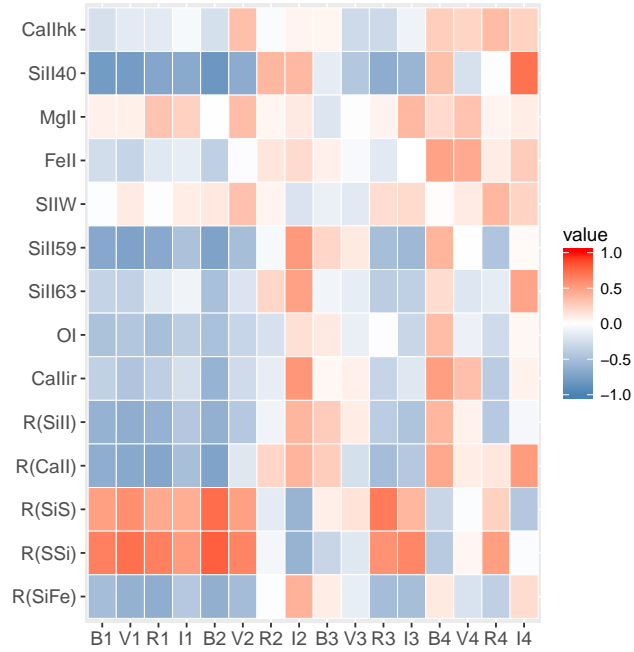
Figure 3.7: The heatmap of correlation between the scores of our model and the spectral features. The horizontal axis corresponds to the first four scores across all of the $B, V, R,$ I bands. The vertical axis corresponds to various spectral features.

may prove to be critical for the WFIRST program (Spergel et al., 2013) which aims at unprecedented precision.

Here we will only use the spectral features extracted by Silverman et al. (2012b). The work of Silverman et al. (2012b) provides pseudo-equivalent width (pEW), spectral feature depths, and fluxes at the center and end points of nine spectral feature complexes. The nine spectral feature complexes are Ca II H&K, Si II $\lambda 4000$, Mg II, Fe II, Si II 'W', Si II $\lambda 5972$, Si II $\lambda 6355$, O I triplet, Ca II near-IR triplet. We will compare the first four *original* scores $\beta^{(1)}, \beta^{(2)}, \beta^{(3)}, \beta^{(4)}$ with the pseudo-equivalent width (pEW) of these spectral feature.

The pEW of the spectral features within 5 days of the B maximum are obtained, and their Spearman correlation coefficients with the scores are computed. The correlation is visualized as heatmap in Figure 3.7. These nine spectral features corresponds to
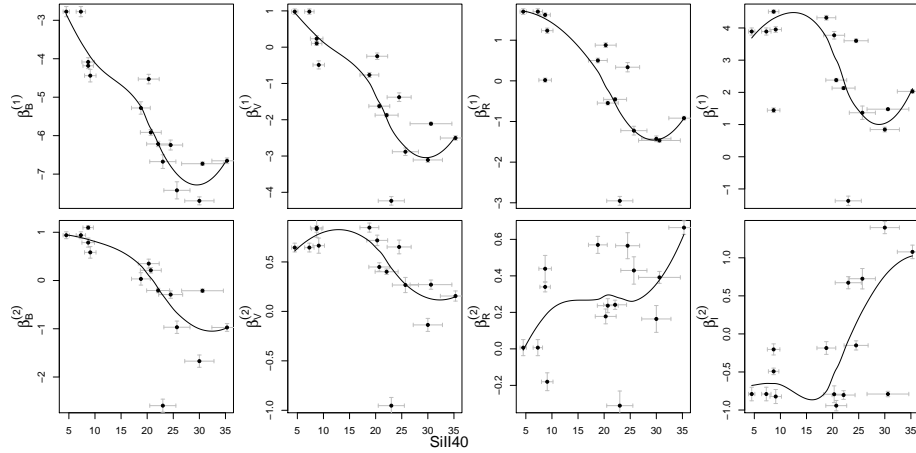
Figure 3.8: The first two scores of four bands against the pseudo-equivalent width (pEW) of Si II $\lambda4000$.

the first nine rows in the figure. Its columns from left to right correspond to the scores $\beta_B^{(1)}, \beta_V^{(1)}, \beta_R^{(1)}, \beta_I^{(1)}, \beta_B^{(2)}, \cdots, \beta_I^{(4)}$ respectively. Saturated red (blue resp.) implies a strong positive (negative resp.) correlation; and white color implies a very weak correlation.

The spectral feature Si II $\lambda4000$ and Si II $\lambda5972$ are important spectral luminosity indicator (Nugent et al., 1995). With the light curve width parameter from the SALT2 model, Silverman et al. (2012a) notice these two features are correlated with light curve width. This is also confirmed in our dataset. Both of them have strong (negative) anti-correlation with the first two scores across four optical bands. The exception is that $\beta_R^{(2)}$ has weak correlation with these two spectral features. We have also shown in Figure 2.3 the correlation of ΔM_{15} with the first two scores. The correlations among the spectral feature, the first two scores, and the ΔM_{15} imply that with sufficient amount of well calibrated data it would be possible to construct light curve templates for different spectral sub-classes of SNIa.

Our work provides more details on how these two spectral features correlate with light curve shape. Figure 3.8 is the scatter plot of the first two scores with the pEW of
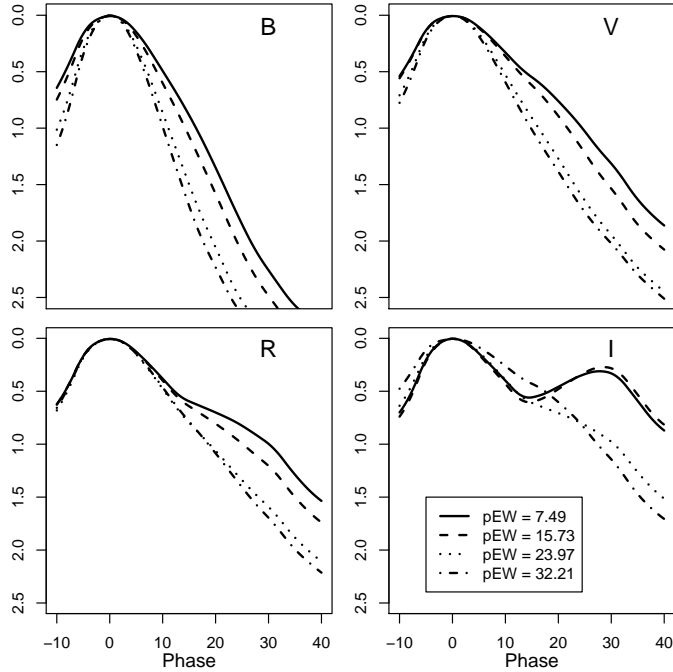
Figure 3.9: The "average" light curve shape at four different levels of the pseudo-equivalent width (pEW) of Si II λ4000.

Si II λ4000. In the first row of Figure 3.8, the first scores of all bands are negatively correlated with the pEW of Si II λ4000. However, as the central wavelength of the band increases, the first score $\beta^{(1)}$ becomes less sensitive to this spectral feacture. The I band score $\beta_I^{(1)}$ starts to drop when pEW(Si II λ4000) increase to 17; it almost remains at a constant level for pEW(Si II λ4000) in the interval $[5, 15]$.

Meanwhile, for the second score $\beta^{(2)}$ and pEW(Si II λ4000), their correlation changes from negative correlation to positive correlation as the central wavelength of the filter increases. This is shown in the second row of Figure 3.8. The B band score $\beta_B^{(2)}$ is negatively correlated with pEW(Si II λ4000); and the I band score $\beta_B^{(2)}$ is positively correlated with pEW(Si II λ4000)

Recall from Figure 2.2, the first score mainly affects the decline rate after 15 days from the peak. The second score affects the light curve width around the peak, and the decline

rate contrast before and after +15 days in phase. As both $\beta_B^{(1)}$ and $\beta_B^{(2)}$ are negatively correlated with pEW(Si II λ4000), stronger Si II λ4000 will shrink the B band light curve width across the entire phase range $[-10, 40]$. On the other hand, in the last column of Figure 3.8, the correlation pattern for $\beta_I^{(1)}$ and $\beta_I^{(2)}$ implies that strong Si II λ4000 tend to make the I band light curve wider around the peak and decrease faster after +15 days in phase.

This effect on light curve shape is illustrated in Figure 3.9. They are the "average" light curve shapes for each band at different levels of pEW(Si II λ4000). The first two scores as a function of pEW(Si II λ4000) is fitted by a LOESS curve (which is a robust local regression), as the solid line in Figure 3.8. Then we compute the value of $\beta_f^{(1)}, \beta_f^{(2)}$ at pEW$\approx$ 7.49, 15.73, 23.97, 32.21 for each band f. After that, the average light curve shapes are computed as $\phi_0(q) + \beta_f^{(1)}\phi_1(q) + \beta_f^{(2)}\phi_2(q)$. Of special interest is the lower right panel of Figure 3.9. As expected from previous analysis, when the strength of Si II λ4000 increases, the I band light curve becomes wider around the peak, but narrower after +15 days in phase. The B band light curve become uniformly narrower across the entire phase range.

Next, we also consider five spectral ratios as defined by Silverman et al. (2012a). The first is the Si II ratio as the pEW of Si II λ5972 divided by the pEW of Si II λ6355.

$$\mathcal{R}(\text{Si II}) = \frac{\text{pEW(Si II } \lambda 5972)}{\text{pEW(Si II } \lambda 6355)} \, .$$

The second is the ratio $\mathcal{R}(\text{Ca II})$ of the flux at the red and blue end of Ca II H&K.

$$\mathcal{R}(\text{Ca II}) = \frac{\text{F}_\text{r}(\text{Ca II H\&K})}{\text{F}_\text{b}(\text{Ca II H\&K})} \, .$$

These two spectral feature ratios are among the first spectral luminosity indicators (Nugent

et al., 1995). Three additional spectral ratios are defined as

$$\mathcal{R}(\text{SiS}) = \frac{\text{F}_\text{r}(\text{Ca II W})}{\text{F}_\text{r}(\text{Ca II } \lambda 6355)} \,,$$

$$\mathcal{R}(\text{SSi}) = \frac{\text{pEW}(\text{Ca II W})}{\text{pEW}(\text{Ca II } \lambda 5972)} \,,$$

$$\mathcal{R}(\text{SiFe}) = \frac{\text{pEW}(\text{Ca II } \lambda 5972)}{\text{pEW}(\text{Fe II})} \,.$$

These five ratios correspond to the last five rows in Figure 3.7. They also have strong correlation (or anti-correlation) with the first two scores across all four bands. Notice the exception is the second score of the R band $\beta_R^{(2)}$. It has very weak correlation with all the spectral features. Furthermore, the correlation of $\beta_I^{(2)}$ tend to have opposite sign with the correlations involving $\beta_B^{(1)}, \beta_V^{(1)}, \beta_R^{(1)}, \beta_I^{(1)}, \beta_B^{(2)}, \beta_V^{(2)}$.

3.2.2 The Scores and Spectral Classes
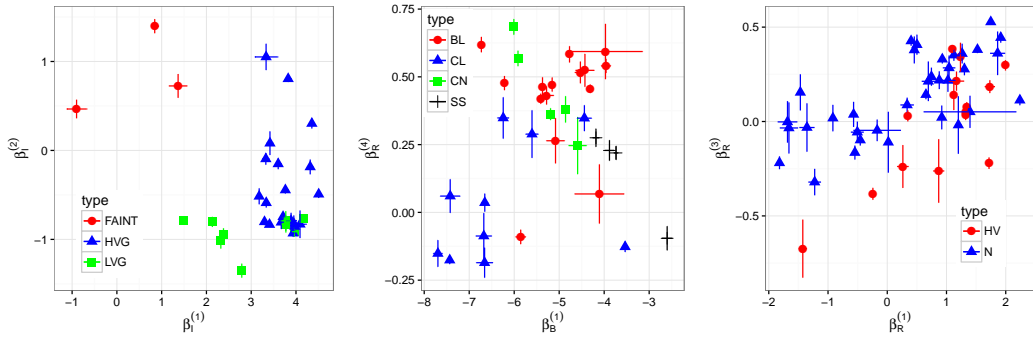
Figure 3.10: Spectral classes separation using the scores. The spectral classes from left to right corresponds to the results of Benetti et al. (2005), Branch et al. (2009), and Wang et al. (2009), respectively.

Section 3.2.1 explains that the scores from our model provide abundant information about spectral features. The next step is to determine supernova spectral class based on

the scores. This task could be modeled as a classification problem. We can possibly treat spectral classes as the response, and our scores as predictors. To keep the analysis simple, we try to separate spectral classes with only two selected scores. The two scores are selected with the aid of sparse linear discriminant analysis (LDA). A reference of LDA can be found in Murphy (2012). The sparse LDA tries to linearly combine predictors to separate classes. The linear combining coefficients are called loadings. The number of predictors used is encouraged to be small. We control the sparsity of the loading so that only one variable is allowed in each of the first two loading vectors (introducing more variables only gives marginal improvement). Then we examine the ability of the selected two scores to separate spectral classes. In the following, we take the spectral classes from Benetti et al. (2005), Branch et al. (2009), and Wang et al. (2009). We will use all the SNIa in our sample with spectral classes assigned by these papers.

Firstly, we consider the three spectral classes in Benetti et al. (2005). The three classes are FAINT, high temporal velocity gradient (HVG) and low temporal velocity gradient group (LVG). The average velocity gradients in the three groups are 87, 97, and 37, respectively. The three classes are plotted as red circle, blue triangle, and green square in the left panel of Figure 3.10. The classification is carried out with the first and second score of the I band light curve, $\beta_I^{(1)}, \beta_I^{(2)}$. The three classes are quite well separated. The HVG and LVG classes have a small overlap regions.

Branch et al. (2009) provided spectral classification on the basis of the absorption features near 5750Å and 6100Å. The absorption features are measured by pseudo equivalent width. Their four groups are core normal (CN), broad line (BL), cool (CL), and shallow silicon (SS). The CN is a homogeneous class, and its absolute magnitude has small correlation with light curve width ΔM_{15}. The BL class tends to have strong absorption near 6100Å. For the CL class, the absorption features near 5750Å and 6100Å are both strong; and the SS class is another extreme with both features weak. On average, the CL class

tends to have higher ΔM_{15} values and fainter absolute magnitude; on the contrary, the SS class tend to have lower ΔM_{15} values and brighter absolute magnitude.

The middle panel of Figure 3.10 presents the four class separation based on the B band first score and R band fourth scores $\beta_R^{(4)}$. The CL (blue triangle) and SS (black cross) classes are separated in the lower right and lower left corner of the plot. The CN (green square) and BL (red circle) classes are mixed in the top.

In Wang et al. (2009), the supernova samples are classified into two groups Normal (N) and high velocity (HV) according to the observed velocity of Si II λ6355. They found that the HV group has narrower distribution in peak luminosity and decline rate, and this group also prefers a lower extinction ratio. The distance prediction model is applied to these two groups separately to reduce the dispersion. The right panel in Figure 3.10 shows the spectral class separation by the first and third score of the R band light curve. These two classes have a considerable overlap in the scatterplot.

When only light curve observation is available, Figure 3.10 shows we may still be able to classify the observation into their corresponding spectral class. However, even if a quantitative scheme proves to be difficult, based on the class separation relation in Figure 3.10, we could still try to extract a subgroup of SNIa which could be more homogeneous than the entire sample. This can be useful in controlling systematic errors in distance determination.

3.3   Distance Prediction

Our FPCA light curve model can produce the entire shape of a light curve, which in turn can be used for intrinsic luminosity adjustment in distance prediction. In this section, we consider an adjustment using a functional linear form and compare it with the standard adjustment using ΔM_{15}. We consider both the peak magnitude and the CMAGIC magnitude as distance indicators in separate subsections. Note that in this comparative

study, the purpose is to show the potential advantage of using the entire light curve shape, and we still use our FPCA model to determine the value of ΔM_{15}.

3.3.1 Distance Models

Distance models fit a linear model for the distance modulus. The distance modulus μ is a function of redshift. In particular,

$$\mu(z) = 25 + 5\log_{10}(D_L(z)\mathrm{Mpc}^{-1}),$$

where $D_L(z)$ is the luminosity distance under a fixed cosmology with $\Omega_m = 0.3$, $\Omega_\Lambda = 0.7$ and Hubble constant $h = 0.7$. With type Ia supernova, the standard model for predicting distance is

$$(M1) \quad \mu = m_B - M - \alpha(C - \langle C \rangle) - \delta(\Delta M_{15} - \langle \Delta M_{15} \rangle). \tag{3.2}$$

In the above, m_B is the apparent B band peak magnitude, M is the absolute B band peak magnitude. All these are magnitudes in the rest-frame filter. C is the observed color at B maximum, $(B - V)_{\mathrm{max}}$. ΔM_{15} is the magnitude change 15 days after B maximum for the B band light curve.

Although simple by concept, ΔM_{15} is not a quantity that can be directly measured accurately. Its calculated value is highly influenced by light curve fitting errors. In practice, because of the rapid luminosity decay at 15 days past optical maximum, a small error in determining the peak epoch can lead to large inaccuracy of ΔM_{15}.

With our proposed FPCA light curve model, we consider an alternative adjustment of intrinsic luminosity using the entire shape of the light curve. The shape parameter ΔM_{15}

is replaced by a functional linear term

$$Q = \int \delta(q)(g_B(q) - m_B - \mu(q))\mathrm{d}q, \qquad (3.3)$$

where $\delta(s)$ is a fixed function to be determined by the data. Note that in Equation (3.3), the peak magnitude m_B and the mean function $\mu(q)$ are subtracted, so that only the light curve shape may influence the value of Q. With this functional linear form, the distance model becomes

$$(M2) \quad \mu = m_B - M - \alpha(C - \langle C \rangle) - (Q - \langle Q \rangle). \qquad (3.4)$$

Since the light curve shape enters the model as a functional linear term, we refer to (9) as a functional linear distance model. This model has some similarity to the functional linear regression model studied in statistics (Müller and Stadtmüller, 2005).

For each supernova, the associated distance prediction uncertainty includes the parts due to peculiar velocity, measurement error in the apparent magnitude, and intrinsic Type Ia supernova property variation. In the following we assume a peculiar velocity of $v_{\mathrm{pec}} = 300 \ \mathrm{km \ s^{-1}}$ and it introduces a magnitude uncertainty of $\sigma_{\mathrm{pec}} = (5/\ln 10)(v_{\mathrm{pec}}/cz) = 0.002173/z$. The uncertainty of the apparent peak magnitude σ_m is computed from the bootstrap method. The associated distance prediction uncertainty is computed as $\sigma_s^2 = \sigma_{\mathrm{pec}}^2 + \sigma_m^2$.

The standard distance model (3.2) is trained by minimizing the χ^2,

$$\sum_{s=1}^{S} \frac{1}{\sigma_s^2} [\mu - m_B + M + \alpha(C - \langle C \rangle) + \delta(\Delta M_{15} - \langle \Delta M_{15} \rangle)]^2,$$

where ΔM_{15} is determined using our FPCA model. The minimization is taken with respect to M, α, δ. A similar χ^2 minimization applies to model (3.4). The functional linear

distance model (3.4) is trained by minimizing

$$\sum_{s=1}^{S} \frac{1}{\sigma_s^2} [\mu - m_B + M + \alpha(C - \langle C \rangle) + (Q - \langle Q \rangle) + \lambda \int [\delta''(q)]^2 \mathrm{d}q.$$

The minimization is taken with respect to M, α and $\delta(q)$ in Q. The last term is the integration of the squared second order derivative of $\delta(q)$. This is a roughness penalty to encourage the smoothness of the solution of $\delta(q)$. The λ parameter controls the amount of penalty imposed and is chosen by the cross-validation. In the χ^2 minimization, the solution of $\delta(q)$ is searched among the span of the principal component functions $\phi_k(q)$, i.e., $\delta(q) = \sum_{k=1}^{K} \delta^{(k)} \phi_k(q)$. Using the principal component functions, we only need to solve for the scalar $\delta^{(k)}$'s to get an estimate of the function $\delta(q)$.

3.3.2 Comparing the Distance Prediction Models

The leave-one-out cross-validation is used to compare the distance prediction models: M1 (Equation 3.2), and M2 (Equation 3.4). Cross-validation is a commonly used method in statistics to evaluate the out-of-sample performance of a prediction model. It works as follows. Each time one sample (i.e. one SNIa) is removed from the dataset, and the remaining dataset is used to to train the distance prediction model. The resulted model is then applied to the removed sample to get a predicted distance modulus $\widehat{\mu}$. The cross-validated error of this prediction is denoted as $\Delta\mu = \mu - \widehat{\mu}$. We repeat this procedure for all SNIa samples in the dataset and summarize the cross-validated errors by the weighted mean squared errors (WMS),

$$\mathrm{WMS} = \left(\sum_s \Delta\mu_s^2 / \sigma_s^2 \right) / \left(\sum_s 1 / \sigma_s^2 \right),$$

whose square-root is $\mathrm{WRMS} = \mathrm{WMS}^{1/2}$.

Before applying the model to our dataset, we make a further selection of the dataset.

Table 3.1: Comparison of the distance models. Comparison of the WRMS and χ^2 for two models: M1 (Equation 3.2) and M2 (Equation 3.4). The comparison is based on different observed color $(B - V)_{\mathrm{max}}$ cutoffs and a common redshift cut $z > 0.01$. The column N is the sample size after the cutoff.

		M1		M2	
CC	N	WRMS	χ^2	WRMS	χ^2
0.05	35	0.109	136.30	0.110	128.85
0.1	48	0.123	245.48	0.123	224.78
0.2	65	0.127	347.48	0.119	278.85
0.3	69	0.130	381.49	0.126	333.51
0.4	71	0.131	391.44	0.127	345.79

We select the SNIa observations with CMB redshift $z > 0.01$ and the observed color $(B - V)_{\mathrm{max}} < CC$ for several values of CC. The cut 0.01 on redshift restricts the uncertainty due to peculiar velocity. In addiction, the cut CC on the observed color helps us to select a homogenous group of supernovae.

The result is shown in Table 4.1. In the table, the WRMS is computed using cross-validated errors. The χ^2 is computed using in-sample errors. For example, the result for $CC = 0.05$ is given on the first row. With the cut of $z > 0.01$ and $(B - V)_{\mathrm{max}} < (0.05)$, there are 35 SNIa in the remaining sample. The ΔM_{15} model (3.2) results in a WRMS of 0.109, and the functional linear distance model (3.4) has a close WRMS of 0.110.

In general, the difference of the two models is not significant. With only shape and color information, both models appear to have approached the statistical limits of the data in constructing a Hubble diagram with minimal dispersion. The dispersion is dominated by the peculiar velocity. However, the functional linear distance model (3.4) still consistently produces smaller WRMS across different sample groups.

At last, we present more detailed results for model (3.4) at the color cut $CC = 0.4$, where the sample size is the largest in our consideration. The upper panel of Figure 3.11 shows the predicted distance modulus versus redshift velocity. The lower panel of Fig-

ure 3.11 plots the cross-validated residuals and associated error bars. The dashed curves represent the uncertainty due to the assumed peculiar velocity. Note the scatter of the residuals is dominated by peculiar velocity at redshifts around $300 \mathrm{~km} \cdot \mathrm{s}^{-1}$. In addition, the estimated functional coefficient $\delta(q)$ is presented in Figure 3.12. This functional co-efficient $\delta(q)$ is positive over the phase range $(-10, 30)$. This suggests that the functional linear form still tries to measure the width of the light curve in its own way, and the measurement is adjusted by the phase range $(30, 40)$. Figure 3.13 compares the quantity ΔM_{15} with the calculated value of the the functional linear form $\int \delta(q)(g_B(q) - m_B - \mu(q))\mathrm{d}q$.
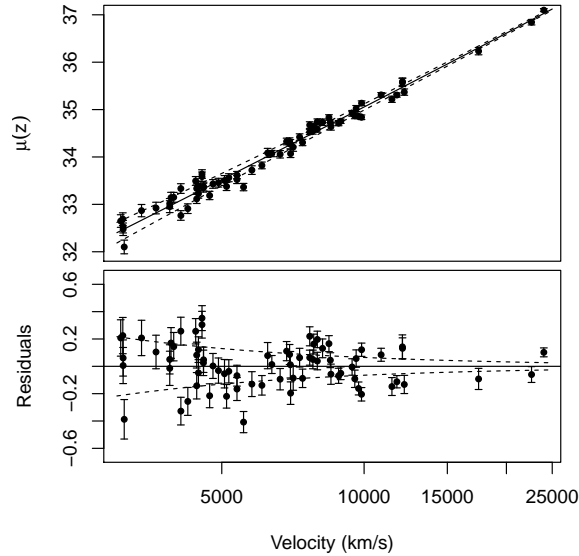
Figure 3.11: Predicted distance and residuals of the functional linear distance model. The dashed lines represent the uncertainty due to the peculiar velocity.

3.3.3 The CMAGIC for Distance Prediction

We now evaluate the effectiveness of using the entire light curve shape to adjust intrinsic luminosity for distance predication when the CMAGIC magnitude is used as the distance indicator. The CMAGIC magnitude is proposed by Wang et al. (2003). They
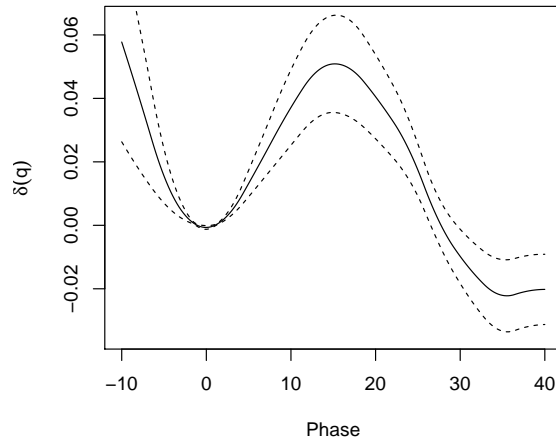
Figure 3.12: The estimated $\delta(q)$ of the functional linear form (3.3). The solid line is the estimated $\delta(q)$, the dashed lines represent one standard deviation uncertainty.
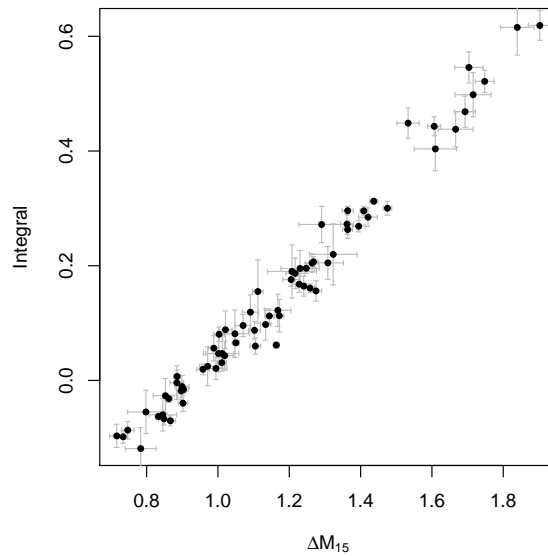
Figure 3.13: Comparing the ΔM_{15} on the horizontal axis with the value of the functional linear form $\int \delta(q)(g_B(q) - m_B - \mu(q))\mathrm{d}q$ on the vertical axis.
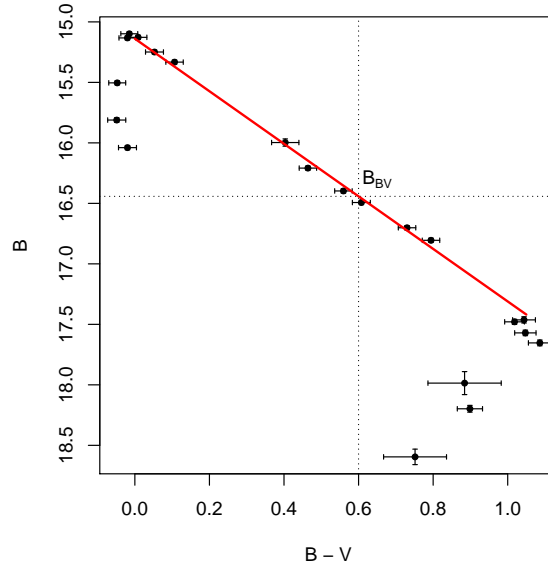
Figure 3.14: An example of the linear color magnitude evolution after B maximum. The red line is the fitted linear relation. The B band magnitude at $B - V = 0.6$ is extracted to get the CMAGIC magnitude B_{BV}.

exploited the linear relation of the color evolution for about 30 days after the B maximum. During this phase the B and $B - V$ magnitude follow a linear trend, as illustrated by the red line in Figure 3.14.

$$B = B_{BV} + \gamma(B - V - 0.6) \tag{3.5}$$

where γ is the slope of the linear relation. The exact starting and ending epochs of this linear evolution vary among supernovae with their intrinsic brightness. Some supernovae with a small ΔM_{15} show a "bump" feature in the color magnitude evolution immediately after B maximum. For our supernova samples, the observation points in the B band phase range $[+3, +25]$ are used to fit the linear relation. Wang et al. (2003) noticed that the slope γ has a small scattering 0.18 around the mean of 2.07.

The CMAGIC magnitude, denoted by B_{BV}, is defined as the B band magnitude when $B - V = 0.6$. Wang et al. (2003) replaced the peak magnitude m_B in (3.2) by B_{BV} and

Table 3.2: Comparison of the distance models. Comparison for the CMAGIC among two models: M3 (Equation 3.6) and M4 (Equation 3.7). The comparison is on different cutoff of the color at B maximum, $(B-V)_{\max} < CC$; and a redshift cut of $z > 0.01$. The column N indicates the number of selected samples.

		M3		M4	
CC	N	WRMS	χ^2	WRMS	χ^2
0.05	31	0.135	173.40	0.102	83.06
0.1	43	0.148	288.67	0.125	184.68
0.2	57	0.138	345.88	0.125	258.09
0.3	61	0.144	403.25	0.138	337.25

considered the following model

$$(M3) \quad \mu = B_{BV} - M - \delta(\Delta M_{15} - \langle\Delta M_{15}\rangle) - (b_2 - \gamma)\times$$
$$[\frac{m_B - B_{BV}}{\gamma} + 0.6 + 1.2(\frac{1}{\gamma} - \langle\frac{1}{\gamma}\rangle)]\,, \tag{3.6}$$

where ΔM_{15} is determined using our FPCA model. Now, as an alternative model, the ΔM_{15} is replaced by the functional linear form Q defined in Equation (3.3),

$$(M4) \quad \mu = B_{BV} - M - (Q - \langle Q\rangle) - (b_2 - \gamma)\times$$
$$[\frac{m_B - B_{BV}}{\gamma} + 0.6 + 1.2(\frac{1}{\gamma} - \langle\frac{1}{\gamma}\rangle)]\,. \tag{3.7}$$

In order to fit the linear color evolution, at least five observation points is required in the B band phase range $[+3, +25]$. We select those samples with $z > 0.01$ and make various levels of cut on the observed color $(B-V)_{\max}$ at B maximum. The color cut is necessary, due to the fact that the linear color evolution and CMAGIC can be best constrained among low $(B-V)_{\max}$ samples. For the 61 samples with color cut of 0.3, their slopes γ have a mean of 2.14 and standard deviation 0.17.

Over the selected SNIa samples, the model M3 (Equation 3.6), and model M4 in

(Equation 3.7) are tested using the leave-one-out cross-validation procedure as described in Section 3.3.2. Table 3.2 presents their WRMS and χ^2 at different levels of cutoff at the observed color at B maximum. The model M4 clearly outperforms the model M3 at all scenarios. At the cut 0.05 of the color at B maximum, the model M4 has a remarkable WRMS of 0.102. Notice the sample size N in Table 3.2 is smaller than in Table 4.1, because we require at least 5 observational points in the B band phase range $[+3, +25]$ to constraint the color evolution.

4. PERIOD ESTIMATION FOR SPARSELY SAMPLED QUASI-PERIODIC LIGHT CURVES APPLIED TO MIRAS*

In this chapter, we review several existing period estimation methods. Then we introduce a new semi-parametric (SP) model for Mira variables which uses a Gaussian process to account for deviations from strict periodicity. The maximum likelihood is used to estimate the period and the parameters of the Gaussian process, while other nuisance parameters in the model are integrated out with respect to some prior distributions using earlier studies. Since the likelihood is highly multimodal for the period/frequency parameter, we implement a hybrid method that applies the quasi-Newton algorithm for Gaussian process parameters and a grid search for the period/frequency parameter. In order to assess the effectiveness of the SP model, we carefully construct a simulated data set by fitting smooth functions to the light curves of well-observed OGLE LMC Miras and resampling them at the cadence, noise level, and completeness limits of the aforementioned M33 observations. Using the simulated data, we are able to compare the performance of existing period estimation methods to our SP model. We find that our proposed model shows an improvement over the generalized Lomb-Scargle (GLS) model under various metrics.

4.1 Period Estimation Techniques

Let y_i be the magnitude of a variable star observed at time t_i (in units of days) with uncertainty σ_i. The data set for this object, obtained as part of a time-series survey with n epochs is $\{(t_i, y_i, \sigma_i)\}_{i=1}^n$. One common approach to estimate the primary frequency of such an object is to assume some parametric model for brightness variation and then use maximum likelihood to estimate parameters. Zechmeister and Kürster (2009) define the

GLS model as

$$y_i = m + a\sin(2\pi f t_i + \phi) + \sigma_i \epsilon_i, \tag{4.1}$$

where $\epsilon_i \sim \mathcal{N}(0,1)$, m is the mean magnitude, a is the amplitude, $\phi \in [-\pi, \pi]$ is the phase, and f is the frequency (see Reimann, 1994, for early work in this model). Using the sine angle addition formula and letting $\beta_1 = a\cos(\phi)$ and $\beta_2 = a\sin(\phi)$ one obtains

$$y_i = m + \beta_1 \sin(2\pi f t_i) + \beta_2 \cos(2\pi f t_i) + \sigma_i \epsilon_i. \tag{4.2}$$

The likelihood function of this model is highly multimodal in f. However at a fixed f the model is linear in the parameters (m, β_1, β_2). These two facts motivate the computation strategy of performing a grid search across frequency and minimizing a weighted least squares

$$
\begin{aligned}
(\widehat{m}(f), \widehat{\beta}_1(f), \widehat{\beta}_2(f)) \\
= \underset{m,\beta_1,\beta_2}{\arg\min} \sum_{i=1}^{n} \frac{1}{\sigma_i^2} \left\{ y_i - m - \beta_1 \sin(2\pi f t_i) - \beta_2 \cos(2\pi f t_i) \right\}^2,
\end{aligned}
\tag{4.3}
$$

at every frequency f on the grid. Under the normality assumption, the weighted least squares minimization is equivalent to maximizing the likelihood. Since the model is linear, computation of $\widehat{m}(f), \widehat{\beta}_1(f), \widehat{\beta}_2(f)$ is straightforward. The residual sums of squares at f is

$$\mathrm{RSS}(f) = \sum_{i=1}^{n} \frac{1}{\sigma_i^2} \{ y_i - \widehat{m}(f) - \widehat{\beta}_1(f)\sin(2\pi f t_i) - \widehat{\beta}_2(f)\cos(2\pi f t_i) \}^2, \tag{4.4}$$

and the maximum likelihood estimator for f is

$$\widehat{f} = \underset{f}{\arg\min}\, \mathrm{RSS}(f). \tag{4.5}$$

Define RSS_0 as the (weighted) sum of squared residuals when fitting a model with only an intercept term m. The periodogram is defined as

$$S_{\mathrm{LS}}(f) = \frac{(n-3)(\mathrm{RSS}_0 - \mathrm{RSS}(f))}{2\mathrm{RSS}(f)}. \tag{4.6}$$

The periodogram has the property that if the light curve of the star is white noise (i.e., $y_i = m + \epsilon_i$), $S_{\mathrm{LS}}(f)$ has an $F_{2,n-3}$ distribution. Thus the periodogram may be used for controlling the "false alarm probability," the potential that a peak in the periodogram is due to noise (Schwarzenberg-Czerny, 1996).

A large number of period estimation algorithms in astronomy are closely related to GLS. The LS method is identical to GLS but first normalizes magnitudes to mean 0 and does not fit the m term (Lomb, 1976; Scargle, 1982). The "harmonic analysis of variance" includes an arbitrary number of harmonics in Equation (4.1) (Quinn and Thomson, 1991; Schwarzenberg-Czerny, 1996). Bretthorst (1988) incorporates Bayesian priors on the parameters β_1 and β_2. The method is similar to performing a discrete Fourier transform and selecting the frequency which maximizes the Deeming (1975) periodogram. However, Reimann (1994) showed that GLS has better consistency properties than the Deeming periodogram.

It is also possible to use non-sinusoidal models but compute and minimize the residual sum of squares as above. For example, Hall et al. (2000) consider the Nadaraya-Watson estimator and Reimann (1994) uses the Supersmoother algorithm. Wang et al. (2012) used Gaussian processes with a periodic kernel and found the period with maximum likelihood or minimum leave-one-out cross-validation error.

None of the above methods account for the non-periodic variation present in Miras. While these methods are adequate for densely sampled Mira light curves (where the quantity of data overwhelms model inadequacy), their performance deteriorates in the sparsely
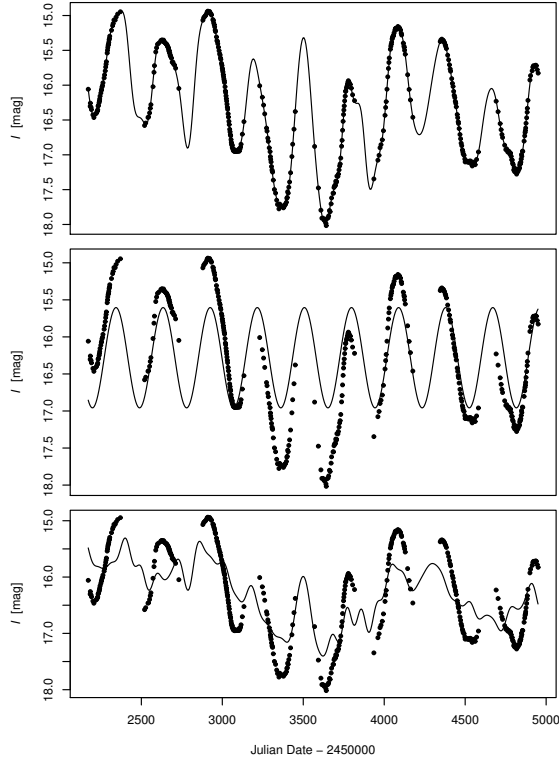
Figure 4.1: Light curve of a Mira in the LMC observed by OGLE (black points), decomposed following Eqn. 4.7. Top panel: fitted light curve; middle panel: periodic signal, $m + q(t)$; bottom panel: stochastic variations, $m + h(t)$.

sampled regime. In Section 4.4, we compare our proposed model with the LS method.

4.2 The SP Model

Suppose the data $\{(t_i, y_i, \sigma_i)\}_{i=1}^{n}$ are modeled by $y_i = g(t_i) + \sigma_i \epsilon_i$, where $g(t_i)$ is the light curve signal and the $\epsilon_i \sim N(0, 1)$ is independent of other ϵ_js. The signal of the light curve is further decomposed into three parts,

$$
\begin{aligned}
g(t) &= m + q(t) + h(t) \\
&= m + \beta_1 \cos(2\pi f t) + \beta_2 \sin(2\pi f t) + h(t) \,,
\end{aligned}
\tag{4.7}
$$

61

where m is the long-run average magnitude, $q(t) = \beta_1 \cos(2\pi f t) + \beta_2 \sin(2\pi f t)$ with frequency f is the exactly periodic signal, and $h(t)$ is the stochastic deviation from a constant mean magnitude, caused by the formation and destruction of dust in the cool atmospheres of Miras. Fig. 4.1 provides an example of the decomposition for a Mira light curve. The first two terms $m + q(t)$ in Eqn. 4.7 are exactly the same as the GLS model of Eqn. 4.2. To simplify notation, we define $\mathbf{b}_f(t) = (\cos(2\pi f t), \sin(2\pi f t))^T$, so that $q(t) = \mathbf{b}_f(t)^T \boldsymbol{\beta}$. The subscript in $\mathbf{b}_f(t)$ emphasizes that the basis is parameterized by the frequency f.

An SP statistical model is constructed in Eqn. 4.7 if we assume $h(t)$ is a smooth function that belongs to a reproducing kernel Hilbert space $\mathcal{H}$ with norm $\|\cdot\|_{\mathcal{H}}$ and a reproducing kernel $K(\cdot, \cdot)$. For this model, if the frequency f is known, we obtain a least squares kernel machine considered in Liu et al. (2007). Because the frequency is unknown, the response function is nonlinear in f. This nonlinearity and the multimodality in f of the residual sum of squares provide additional challenges that require a novel solution.

Besides the additive formulation in Eqn. 4.7, another possible solution to account for the quasi-periodicity is a multiplicative model such as $g(t) = m + h(t)q(t)$, where the amplitude of the strictly periodic term $q(t)$ is modified by a smooth function $h(t)$. However, the multiplicative model is more computationally intensive in nature and requires imposing a positive constraint on $h(t)$. As we will show in the following subsections, the $h(t)$ term in the additive model can be easily absorbed into the likelihood function. Nevertheless, the multiplicative approach is an interesting alternative approach to model formulation and is open to future study.

4.2.1 Equivalent Formulations

Following §5.2 of Rasmussen and Williams (2005), for fixed f, the parameters m, β_1, β_2 and $h(t)$ in Eqn. 4.7 are jointly estimated by minimizing

$$\sum_{i=1}^{n} \frac{1}{\sigma_i^2} [y_i - m - \beta_1 \cos(2\pi f t_i) \\ - \beta_2 \sin(2\pi f t_i) - h(t_i)]^2 + \lambda \|h(\cdot)\|_{\mathcal{H}}^2, \tag{4.8}$$

where λ is a regularization parameter. A smoothing/penalized spline model for $h(t)$ is a special case of the general formulation of Eqn. 4.8 with a specifically defined kernel; see §6.3 of Rasmussen and Williams (2005). For fixed λ, the solution of $h(t)$ is a linear combination of n basis functions $K(t_i, t)$, $i = 1, 2, \cdots, n$, by the representer theorem (Kimeldorf and Wahba, 1971; O'sullivan et al., 1986). It is still left for us to choose the regularization parameter λ to balance data fitting and the smoothness of the function $h(t)$.

An equivalent point of view to the above regularization approach is to impose a Gaussian process prior on the function $h(t)$; see §5.2.3 of Rasmussen and Williams (2005). The benefit of this view is that it provides an automatic method for selecting the regularization parameter λ. In particular, we can absorb λ into the definition of the norm $\| \cdot \|_{\mathcal{H}}$ and assume the term $h(t)$ in Eqn. 4.7 follows a Gaussian process, $h(t) \sim \mathcal{GP}(0, k_{\boldsymbol{\theta}}(t, t'))$, with the squared exponential kernel $k_{\boldsymbol{\theta}}(t, t') = \theta_1^2 \exp\left(-\frac{(t-t')^2}{2\theta_2^2}\right)$, and parameters $\boldsymbol{\theta} = (\theta_1, \theta_2)$. The Gaussian process assumption implies that at any finite number of time points $t_1, t_2, \cdots, t_s$, the vector $(h(t_1), \cdots, h(t_s))$ is multivariate normally distributed, with zero mean and covariance matrix $\mathbf{K} = (k(t_i, t_j))$. This imposes a prior on the function space of $h(t)$. We also impose priors on m and $\boldsymbol{\beta}$ in Eqn. 4.7. In particular, we assume $m \sim \mathcal{N}(m_0, \sigma_m^2)$ and $\boldsymbol{\beta} \sim \mathcal{N}(\mathbf{0}, \sigma_b^2 \mathbf{I})$. The prior mean m_0 can be interpreted as the average magnitude of Miras in a certain galaxy, and σ_m^2 is the variance of Miras in that galaxy;

the prior variance σ_b^2 is the variance of the light curve amplitude. These prior parameters can be determined using previous studies. For example, in §4.4, we use well-sampled light curves of LMC Miras (Soszyński et al., 2009) to obtain values of these parameters. It is advisable to check the sensitivity of these prior specifications.

The benefit of using priors on m and $\boldsymbol{\beta}$ is three-fold: first, they introduce regularization by using information from early studies; second, they provide a natural device for separating the estimation of frequency and the light curve signal component using Bayesian integration when the parameter of interest is the frequency; lastly, the regularization parameter $\boldsymbol{\theta}$ of the non-parametric function is allowed to be chosen by the maximum likelihood, without resorting to the computationally expensive cross-validation method.

In summary, we have built the following hierarchical model for a Mira light curve:

$$
\begin{aligned}
y_i|m, \boldsymbol{\beta}, g(t_i) &\sim \mathcal{N}(g(t_i), \sigma_i^2), \\
g(t) &= m + \mathbf{b}_f(t)^T\boldsymbol{\beta} + h(t), \\
m &\sim \mathcal{N}(m_0, \sigma_m^2), \boldsymbol{\beta} \sim \mathcal{N}(\mathbf{0}, \sigma_b^2\mathbf{I}), \\
h(t)|\boldsymbol{\theta} &\sim \mathcal{GP}(0, k_{\boldsymbol{\theta}}(t, t')),
\end{aligned}
\tag{4.9}
$$

where $\boldsymbol{\theta}$ and f are fixed parameters. In this model, the frequency parameter f is of key interest to our study. We do not perform a fully Bayesian inference by imposing a prior distribution on f because the likelihood function of f is highly irregular, with numerous local maxima, and Monte Carlo computation of the posterior is expensive and intractable for large astronomical surveys.

Previously, Baluev (2013) applied a Gaussian process model to study the impact of red noise in radial velocity planet searches. While his maximum likelihood method is a classical frequentist approach in statistics, our approach can be considered as a hybrid of

Bayesian and frequentist approaches. We treat the parameter of interest f, and the parameters for the kernel $\boldsymbol{\theta}$ of the Gaussian process as fixed, and impose a prior distribution on other parameters. This is similar to the type-II maximum likelihood estimation of parameters of a Gaussian process or regularization parameters in function estimation; see §5.2 of Rasmussen and Williams (2005). From the Bayesian point of view, $\boldsymbol{\theta}$ and f are treated as hyper-parameters that in turn are estimated by the empirical Bayes method. Because the Gaussian process plays a critical role in modeling departure of light curves from periodicity, we may also refer to our model more precisely as the nonlinear SP Gaussian process model.

4.2.2 Estimation of the Frequency and the Periodogram

Let $\mathbf{y} = (y_1, y_2, \cdots, y_n)$ be the observation vector of the magnitudes of a light curve. By integrating out $m, \boldsymbol{\beta}$ and $\mathbf{h}$ from the joint distribution given by Eqn. 4.9, we get the marginal distribution of $\mathbf{y}$, $p(\mathbf{y}|\boldsymbol{\theta}, f)$, which is a multivariate normal with mean $\boldsymbol{\mu} = m_0 \mathbf{1}$ and covariance matrix

$$\mathbf{K}_y = \left(\sigma_m^2 + \sigma_b^2 \mathbf{b}_f(t_i)^T \mathbf{b}_f(t_j) + k_{\boldsymbol{\theta}}(t_i, t_j) + \sigma_i^2 \delta_{ij}\right)_{n \times n},$$

where $\delta_{ij} = 1$ if $i = j$ and $\delta_{ij} = 0$ if $i \neq j$. Therefore, the log likelihood of $\boldsymbol{\theta}$ and f is

$$\begin{aligned} Q(\boldsymbol{\theta}, f) &= \log(p(\mathbf{y}|\boldsymbol{\theta}, f)) \\ &= -\frac{1}{2}(\mathbf{y} - m_0\mathbf{1})^T \mathbf{K}_y^{-1}(\mathbf{y} - m_0\mathbf{1}) - \frac{1}{2}\log\det\mathbf{K}_y - \frac{n}{2}\log(2\pi). \end{aligned} \tag{4.10}$$

The maximum likelihood estimator of $\boldsymbol{\theta}$ and f is obtained by maximizing $Q(\boldsymbol{\theta}, f)$. Since the likelihood function is differentiable with respect to $\boldsymbol{\theta}$ but highly multimodal in the parameter f, standard optimization methods cannot be directly used to jointly maximize over $\boldsymbol{\theta}$ and f.

We adopt a profile likelihood method as follows. For each frequency f over a dense grid, we compute the maximum likelihood estimator $\widehat{\boldsymbol{\theta}}_f = \arg\max_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}, f)$. This can be done using the quasi-Newton method. Then we perform a grid search to find the maximum profile likelihood estimator of f, i.e.,

$$\hat{f} = \arg\max_f Q(\widehat{\boldsymbol{\theta}}_f, f), \tag{4.11}$$

the estimated period is $\hat{P} = 1/\hat{f}$. The details of the algorithm are given in §4.2.3. The profile log-likelihood as a function of the frequency f is adopted as the *periodogram* for our model,

$$S_{SP}(f) = Q(\widehat{\boldsymbol{\theta}}_f, f). \tag{4.12}$$

It contains the spectral information of the signal. The frequency of the dominant harmonic component is expected to be the location of the peak of this profile likelihood.

4.2.3 Computation of the Periodogram

Now we present the details of computing the profile likelihood. Because $Q(\boldsymbol{\theta}, f)$ is highly multimodal in the frequency parameter f, we follow the commonly used strategy of optimization through grid search. On the other hand, since $Q(\boldsymbol{\theta}, f)$ is differentiable in parameter $\boldsymbol{\theta}$, the quasi-Newton method can be employed to optimize over $\boldsymbol{\theta}$ for fixed f, and obtain the profile likelihood (Eqn. 4.12). The gradient of the log likelihood (Eqn. 4.10) with respect to $\theta_j (j = 1, 2)$ is

$$\frac{\partial}{\partial \theta_j} Q(\boldsymbol{\theta}, f) = \frac{1}{2} \text{tr}\left((\boldsymbol{\alpha}\boldsymbol{\alpha}^T - \mathbf{K}_y^{-1}) \frac{\partial \mathbf{K}_y}{\partial \theta_j} \right)$$

where $\boldsymbol{\alpha} = \mathbf{K}_y^{-1}(\mathbf{y} - m_0 \mathbf{1})$. In general, the objective function for the Gaussian process model is not convex in its kernel parameters $\boldsymbol{\theta}$ and global optimization cannot be guaran-
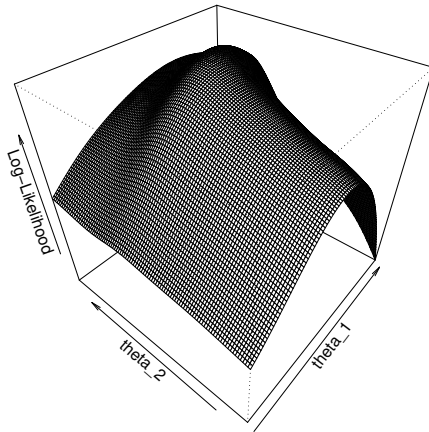
Figure 4.2: The three dimensional surface plot of $Q(\boldsymbol{\theta}, f)$ in Eqn. 4.10, for the simulated light curve in Fig. 4.6. Notice $Q(\boldsymbol{\theta}, f)$ is plotted as a function of $\boldsymbol{\theta}$ and f is fixed at its true frequency.

teed. Fig. 4.2 shows a surface plot of $Q(\boldsymbol{\theta}, f)$ as a function of $\boldsymbol{\theta}$ for one simulated light curve, with f fixed at the true frequency. The surface exhibits unimodality in this case, although it is not convex.

The computation involved in calculating the profile likelihood through the quasi-Newton method can be intensive. Since the objective function (Eqn. 4.10) is non-convex in $\boldsymbol{\theta}$, generally multiple starting points should be attempted to find the global optimizer when applying the quasi-Newton method. In addition, evaluating the objective function and the gradient function requires inversion of the covariance matrix whose computation cost is of the order $O(n^3)$. During each quasi-Newton iteration, these evaluations could be repeated several times because multiple step size might be attempted. To make the computation more challenging, all of the above needs to be repeated at hundreds or even thousands of densely gridded fs per light curve. Furthermore, the method may need to be applied to hundreds of thousands or millions of light curves from large astronomical surveys.

In order to speed up computation over the dense grid of frequency values, we use the

---
Algorithm 2 Quasi-Newton's Method with Grid Search.
---
procedure QUASINEWTON()

Input: Maximal and minimal trial frequencies $f_M > f_m > 0$; frequency step Δf; n observations $\{t_i, y_i, \sigma_i\}$.

Output: Periodogram $S(f)$ evaluated at the trial frequencies.

1: Initialize $\boldsymbol{\theta}^{(0)}$ and $\mathbf{H}^{(0)}$, and $f \leftarrow f_m$;

2: **for** $f \in \{f_m, f_m + \Delta f, f_m + 2\Delta f, \cdots, f_M\}$ **do**

3:    $p \leftarrow 0$;

4:    **repeat**

5:      $\mathbf{t}_p \leftarrow -\mathbf{H}^{(p)} \frac{\partial}{\partial \boldsymbol{\theta}} Q(\boldsymbol{\theta}^{(p)}, f)$;

6:      $\boldsymbol{\theta}^{(p+1)} \leftarrow \boldsymbol{\theta}^{(p)} + \alpha_p \mathbf{t}_p$ and the step size α_p satisfying the Wolfe condition;

7:      $\mathbf{d}_p \leftarrow \boldsymbol{\theta}^{(p+1)} - \boldsymbol{\theta}^{(p)}$, $\mathbf{e}_p \leftarrow \frac{\partial}{\partial \boldsymbol{\theta}} Q(\boldsymbol{\theta}^{(p+1)}, f) - \frac{\partial}{\partial \boldsymbol{\theta}} Q(\boldsymbol{\theta}^{(p)}, f)$;

8:      $\rho_p = 1/\mathbf{d}_p^T \mathbf{e}_p$;

9:      $\mathbf{H}^{(p+1)} \leftarrow (\mathbf{I} - \rho_p \mathbf{d}_p \mathbf{e}_p^T) \mathbf{H}^{(p)} (\mathbf{I} - \rho_p \mathbf{e}_p \mathbf{d}_p^T) + \rho_p \mathbf{d}_p \mathbf{d}_p^T$;   ▷ BFGS update

10:      $p \leftarrow p + 1$;

11:    **until** $\|\mathbf{d}_{p-1}\| < \epsilon$

12:    $\widehat{\boldsymbol{\theta}}_f \leftarrow \boldsymbol{\theta}^{(p)}$, and $S(f) \leftarrow Q(\widehat{\boldsymbol{\theta}}_f, f)$;

13:    $\mathbf{H}^{(0)} \leftarrow \mathbf{H}^{(p)}$ and $\boldsymbol{\theta}^{(0)} \leftarrow \boldsymbol{\theta}^{(p)}$;   ▷ Save warm start value for next trial frequency

14: **end for**
---

result of applying the quasi-Newton method at one frequency value as a warm start for the subsequent frequency value. Specifically, the optimizer $\widehat{\boldsymbol{\theta}}_f$ and its approximate inverse Hessian matrix are provided as quantities to start the quasi-Newton iterations for the next frequency value on the dense grid. When the initial point is near the local minimizer and the inverse Hessian matrix is a good approximation to the true Hessian matrix, the quasi-Newton algorithm will converge at superlinear rate; the step size of $\alpha = 1$ will be accepted by the Wolfe descent condition, avoiding evaluation of the objective function multiple times to determine the appropriate step size during each iteration (see Ch. 6 of Nocedal and Wright, 2006, for a more rigorous mathematical discussion). We find that a warm start can speed up the computation significantly but sometimes we need to restart with random initial values to ensure convergence to the global optimum. The pseudocode provided in the Algorithm X describes our algorithm.

4.2.4 Estimation of the Signal and its Components

After the parameters f and $\boldsymbol{\theta}$ are fixed at their maximum likelihood estimates $\widehat{f}$ and $\widehat{\boldsymbol{\theta}}_{\widehat{f}}$, we can perform the inference of the light curve signal $g(t)$ and its components in the standard Bayesian framework. Interested readers may consult Ch. 2 of Rasmussen and Williams (2005) for a detailed discussion of this topic.

Firstly, we could obtain the posterior distribution of $\boldsymbol{\gamma} = (m, \boldsymbol{\beta}^T)$, the parameters for the long run average magnitude and the exactly periodic term. The prior of $\boldsymbol{\gamma}$ is $\mathcal{N}(\boldsymbol{\gamma}_0, \boldsymbol{\Sigma}_\gamma)$ with $\boldsymbol{\gamma}_0 = (m_0, 0, 0)^T$ and $\boldsymbol{\Sigma}_\gamma = \mathrm{diag}(\sigma_m^2, \sigma_b^2, \sigma_b^2)$. Its posterior distribution is $\boldsymbol{\gamma}|\mathbf{y} \sim \mathcal{N}(\bar{\boldsymbol{\gamma}}, \bar{\boldsymbol{\Sigma}}_\gamma)$ with

$$
\begin{aligned}
\bar{\boldsymbol{\gamma}} &= \left(\mathbf{H}^T \mathbf{K}_c^{-1} \mathbf{H} + \boldsymbol{\Sigma}_\gamma^{-1}\right)^{-1} \\
&\quad \left(\boldsymbol{\Sigma}_\gamma^{-1} \boldsymbol{\gamma}_0 + \mathbf{H}^T \mathbf{K}_c^{-1} \mathbf{y}\right), \\
\bar{\boldsymbol{\Sigma}}_\gamma &= \left(\mathbf{H}^T \mathbf{K}_c^{-1} \mathbf{H} + \boldsymbol{\Sigma}_\gamma^{-1}\right)^{-1},
\end{aligned}
\tag{4.13}
$$

where

$$
\mathbf{h}(t) = (1, \mathbf{b}_{\widehat{f}}(t)^T)^T, \mathbf{H} = (\mathbf{h}(t_1), \mathbf{h}(t_2), \cdots, \mathbf{h}(t_n))^T,
$$

and $\mathbf{K}_c = \left(k_{\widehat{\boldsymbol{\theta}}_{\widehat{f}}}(t_i, t_j) + \sigma_i^2 \delta_{ij}\right)_{n \times n}$ with $\widehat{f}$ and $\widehat{\boldsymbol{\theta}}_{\widehat{f}}$ plugged in.

Consider the prediction of light curve magnitude at a specific time point t^*. Define the vector $\mathbf{k}^* = (k_{\widehat{\boldsymbol{\theta}}}(t^*, t_1), \cdots, k_{\widehat{\boldsymbol{\theta}}}(t^*, t_n))^T$. Conditional on $(\mathbf{y}, \boldsymbol{\gamma})$, the distribution of $g(t^*)|\mathbf{y}, \boldsymbol{\gamma}$ is a multivariate normal with mean $\mathbf{h}(t^*)^T \boldsymbol{\gamma} + \mathbf{k}_{\boldsymbol{\theta}}(t^*, \mathbf{t}) \mathbf{K}_c^{-1}(\mathbf{y} - \mathbf{H}\boldsymbol{\gamma})$ and variance $k_{\widehat{\boldsymbol{\theta}}}(t^*, t^*) - (\mathbf{k}^*)^T \mathbf{K}_c^{-1} \mathbf{k}^*$. With the posterior distribution of $\boldsymbol{\gamma}$ given in Eqn. 4.13, we are able the remove $\boldsymbol{\gamma}$ from the above conditional distribution of $g(t^*)$. Finally, we get

the posterior distribution of the signal at t^* as $g(t^*)|\mathbf{y} \sim \mathcal{N}(\bar{g}^*, \bar{\sigma}_{g^*}^2)$ with

$$
\begin{aligned}
\bar{g}^* &= \mathbf{h}(t^*)^T\bar{\gamma} + \mathbf{k}(t^*, \mathbf{t})\mathbf{K}_c^{-1}(\mathbf{y} - \mathbf{H}\bar{\gamma}), \\
\bar{\sigma}_{g^*}^2 &= k_{\widehat{\boldsymbol{\theta}}}(t^*, t^*) - (\mathbf{k}^*)^T\mathbf{K}_c^{-1}\mathbf{k}^* + \mathbf{r}^T\bar{\boldsymbol{\Sigma}}_\gamma\mathbf{r},
\end{aligned}
\tag{4.14}
$$

where $\mathbf{r} = \mathbf{h}(t^*) - \mathbf{H}^T\mathbf{K}_c^{-1}\mathbf{k}^*$.

4.3   Simulation of M33 Light Curves

It is not possible to evaluate the period estimation accuracy of our method directly on the M33 data because the "ground truth" is unknown. Instead, we construct a test data set by smoothing the well-sampled OGLE light curves to infer continuous functions, then re-sample these functions to match the observational patterns of the M33 data, and at last add noise to the light curves. This data set can serve as a testbed for future studies of comparing different period estimation methods. We will now describe the M33 observations and the construction of the test data set. As the whole simulation procedure is a complicated process, we will discuss its components in detail from §4.1 to §4.4. The whole simulation procedure will be summarized in §4.5.

4.3.1   Characteristics of the M33 Observations

Most of the disk of M33 was observed by the DIRECT (Macri et al., 2001) and M33SSS (Pellerin and Macri, 2011) projects in the BVI bands, with a combined baseline of $7 - 9$ years and a sampling pattern that depends on the exact location within the disk (see Fig. 4.3). The large area of coverage and long baseline of these observations make them suitable for Mira searches. We use the I-band observations to carry out the simulations, as this is the wavelength range where Miras are brightest (out of the three bands used by these projects). Detailed descriptions of the M33 observations can be found in the above referenced papers. We use the data products from a new reduction that will
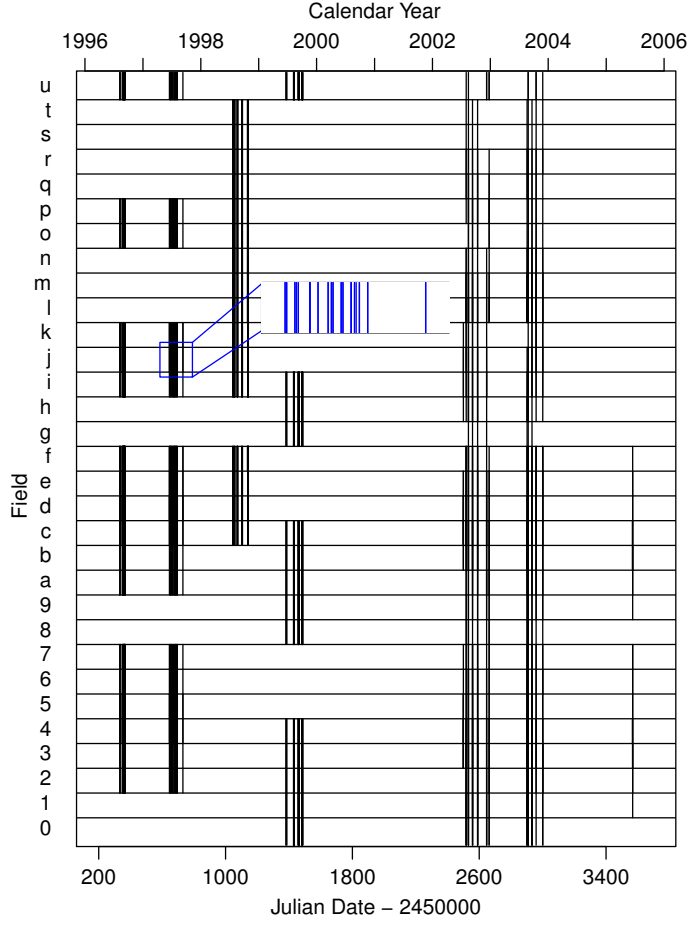
Figure 4.3: Observation patterns for 31 fields in M33, labeled as 0, 1, ..., 9, a, b, ..., u. The horizontal axis shows the Julian date (bottom) and the calendar year (top).

.

be presented in a companion paper (W. Yuan et al. 2016, in prep.). I-band light curves are available for $\sim 2.5 \times 10^5$ stars, with a median of 44 measurements and a maximum of 170. We model the relation between a magnitude measurement m and its uncertainty σ as

$$\sigma = a(t'_i, F)^{[m-b(t'_i,F)]} + c(t'_i, F) \,, \tag{4.15}$$

for each observation field F and each observation night t'_i, where $a(t'_i, F)$, $b(t'_i, F)$ and $c(t'_i, F)$ are field- and night-specific constants. There are 31 different fields in total, $F =$
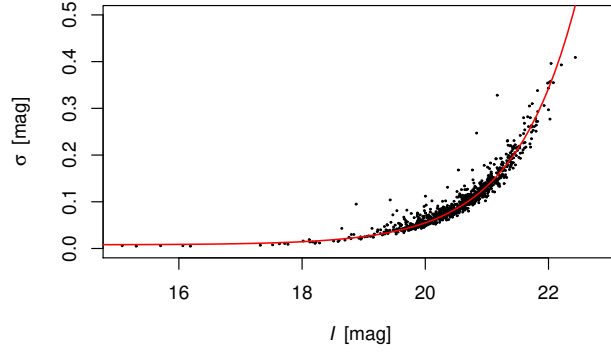
Figure 4.4: $m - \sigma$ relation for a given night and field within M33. The solid red line is the best-fit relation using the empirical function $\sigma = a^{(m-b)} + c$, with $a = 2.666$, $b = 23.117$, $c = 0.008$.

$0, 1, \cdots, 9, a, b, \cdots, u$. The parameters are determined via least-squares fitting using all the measurements for the specific field F and night t'_i. Fig. 4.4 shows the $m - \sigma$ relation for a typical field.

In order to test the SP periodogram we need sparsely sampled, moderately noisy Mira light curves with known periods. Thus, we characterize the sampling patterns and noise levels of the M33 observations and simulated Mira light curves of known periods using the OGLE observations of these objects in the LMC.

4.3.2 Matching the M33 Observation Pattern

The first step in simulating a Mira light curve is to randomly select a sampling pattern based on the light curve of an actual star in some field F, $\{t'_i\}_{i=1}^n$ with $n \in [10, 170]$. A random time shift s is added, $t_i = t'_i + s$ for $i = 1, 2, \cdots, n$. The random shift s follows a uniform distribution over the interval $[0, P_0]$, where P_0 is the true period of the LMC Mira selected during the artificial light curve generation process. This helps to simulate a large number of unique light curves sampled at random phases using the limited number of template light curves.

4.3.3 The Mira Template Light Curves

The template Mira light curves are obtained by using our SP model to fit the Mira light curves in the LMC, collected by the OGLE project (Soszyński et al., 2009). A total number of 1663 Miras have been observed in I with very high accuracy, excellent phase coverage, and a long baseline (the median and mean number of observations are 466 and 602, respectively, with a baseline of ~ 7.5 years for most fields). Because the LMC light curves are densely sampled with high quality, we can adopt a more complicated model to provide a higher fidelity fit. Following §5.4.3 of Rasmussen and Williams (2005), instead of Eqn. 4.7, the signal light curve $g(t)$ is decomposed into

$$g(t) = m + l(t) + q(t) + h(t), \tag{4.16}$$

where m is the long run average magnitude, $l(t)$ is the long-term (low-frequency) trend across different cycles, $q(t)$ is the periodic term, and $h(t)$ is small-scale (high-frequency) variability within each cycle. The latter three terms are modeled by the Gaussian process with different kernels. In particular, we use the squared exponential kernel $k_l(t_1, t_2) = \theta_1^2 \exp(-\frac{1}{2}\frac{(t_1-t_2)^2}{\theta_2^2})$ for $l(t)$, another squared exponential kernel $k_h(t_1, t_2) = \theta_6^2 \exp(-\frac{1}{2}\frac{(t_1-t_2)^2}{\theta_7^2})$ for $h(t)$, and lastly a periodic kernel

$$k_q(t_1, t_2) = \theta_3^2 \exp\left(-\frac{1}{2}\frac{(t_1-t_2)^2}{\theta_4^2} - \frac{2\sin^2(2\pi f(t_1-t_2))}{\theta_5^2} \right)$$

for $q(t)$. Note the periodic kernel allows the light curve amplitude to change across cycles. The maximum likelihood method is applied to fit each LMC light curve, fixing f to the OGLE value and solving for the unknown parameters $(\theta_1, \theta_2, \cdots, \theta_7)$. Fig. 4.5 is an illustration of the model fitting result using Eqn 4.16 based on the same light curve as in Fig. 4.1. Notice that the more complex model in Fig. 4.5 is only suitable for a densely
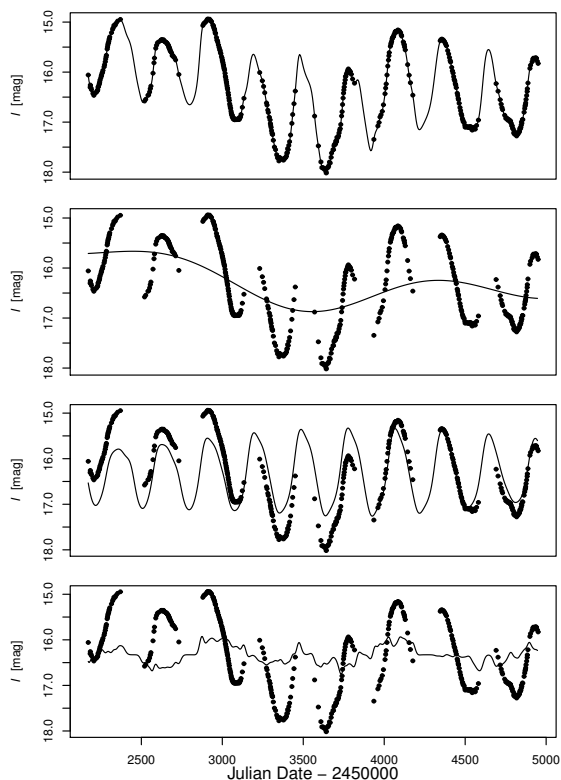
Figure 4.5: Light curve of a Mira in the LMC observed by OGLE (black points), decomposed following Eqn 4.16. Top panel: the fitted light curve; second panel: long-term signal, $m + l(t)$; third panel: periodic term, $m + q(t)$; bottom panel: stochastic variations, $m + h(t)$.

sampled light curve.

Once the sampling pattern is chosen, one of the template light curves will be selected according to the luminosity function described in the next subsection. With the selected template, the magnitude of the simulated light curve signal at t'_i with shift s is $g(t'_i + s)$, which is computed with Eqn. 4.16 in a similar way as Eqn. 4.14.

4.3.4 Matching the Luminosity Function to the M33 Observations

While the OGLE observations of LMC Miras are deep enough to detect these objects over their entire range of luminosities, the M33 observations become progressively more incomplete for fainter and redder objects. We derived an empirical completeness function

74

for the M33 observations as follows. We fitted the observed luminosity function $\mathcal{F}_0(I)$ using an exponential for $I \in [18.5, 20]$ mag and extrapolated to fainter magnitudes, obtaining $\mathcal{F}_1(I)$. The empirical completeness function is then $\mathcal{C}(I) = \mathcal{F}_1(I)/\mathcal{F}_0(I)$.

We randomly picked $\{t_i'\}_{i=1}^n$ from the M33 light curves. For each $\{t_i'\}_{i=1}^n$, we selected a (LMC-based) template using $\mathcal{C}(I + 6.2)$ as the probability distribution. The value of $+6.2$ mag accounts for the approximate difference in distance modulus between the LMC and M33. In this way the resulting luminosity function of the simulated light curves is statistically the same as that of the real M33 observations.

4.3.5 The Simulation Procedure

With all the components discussed above, we are able to present the whole simulation procedure here. In order to generate one simulated Mira light curve matching the sampling characteristics of the M33 observations, the first step is to randomly select a sampling pattern $\{t_i'\}_{i=1}^n$, and then add a random shift s, $t_i = t_i' + s$, $i = 1, 2, \cdots, n$. The second step is to randomly select a template light curve according to the luminosity function, then compute the light curve signal $g(t_i' + s)$ for the selected sampling pattern $\{t_i'\}_{i=1}^n$. The third step is to use the best-fit relations (Eqn. 4.15) to add photometric noise via

$$y_i = g(t_i' + s) + 6.2 + \sigma_i \epsilon_i \,,$$

where $+6.2$ mag is the approximate relative distance modulus, ϵ_i is drawn from $\mathcal{N}(0, 1)$, and σ_i is computed from

$$\sigma_i = a(t_i', F)^{[g(t_i)+6.2-b(t_i',F)]} + c(t_i', F).$$

for the selected observation pattern t_i' and field F. Following this procedure, we generate one simulated light curve $\{t_i', y_i, \sigma_i\}_{i=1}^n$. The procedure is repeated until 10^5 suitable light

curves are generated, excluding any with < 10 data points or sampling on < 7 nights.

4.4 Performance Evaluation

Having generated the test data set, we evaluate the performance of the SP model and compare it with the GLS model. We choose prior parameters for the SP model of $m_0 = 15.62 + 6.2$, $\sigma_m = 10$ and $\sigma_b = 1$. The adopted value of m_0 is the average I magnitude of Miras in the LMC and once again $+6.2$ is the approximate relative distance modulus between M33 and the LMC. The values of σ_m and σ_b are larger than those derived from the LMC samples in order to make those priors non-informative. Although fitting the SP model is computationally slower than the LS model, we find that our model gives an overall improvement in various metrics. For both methods, the periodograms are computed on a dense frequency grid from $1/2000$ to $1/100$ with a spacing of the order of 10^{-5}. For the GLS method, we chose a spacing of (0.05/time span) or $\sim 2.5 \times 10^{-5}$, which results in optimal performance for this simulation. For our SP method, we chose a slightly smaller value of 10^{-5} to facilitate the warm start mechanism in our algorithm (see Appendix) given that small changes in frequency result in tiny changes of the objective function.

4.4.1 The Aliasing Effect

We fit the entire simulated data set using the SP model. Fig. 4.6 gives an example of a simulated light curve and its SP periodogram (Eqn. 4.12). In this example, the true frequency (labeled by the blue dotted line) is successfully recovered.

Aliasing frequencies at $f \pm 1/365$ d affect most periodograms when dealing with sparsely observed astronomical data. The red dashed line in Fig. 4.6 indicates the aliasing frequency at $f + 1/365$ where a strong peak exists. This is not a rare case, and for some light curves the one-year beat aliasing frequencies have higher log likelihoods than the true frequencies. Fig. 4.7 compares the recovered and true frequencies for all simulated light curves. Two secondary strips parallel to the main one and offset by ± 0.00274 represent
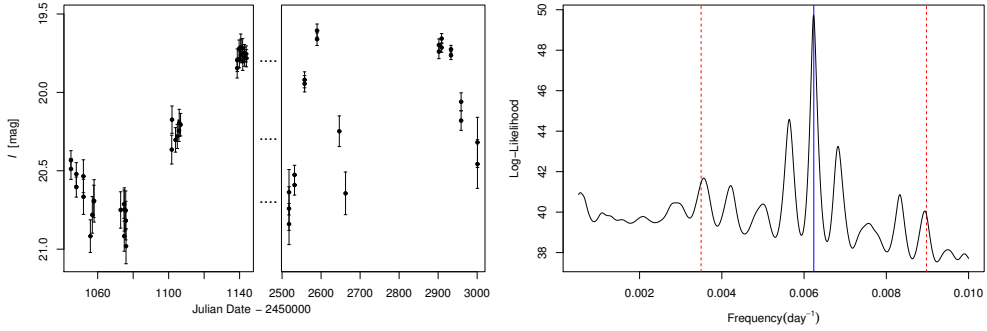
Figure 4.6: Simulated light curve (top) and periodogram $S_{SP}(f)$ based on our model (bottom). Error bars are derived from the M33 observations. In the top panel, a large gap in temporal coverage was removed to make the plot compact (also note the different time spans). The lines in the bottom panel correspond to the true (solid blue) and the one-year aliasing (dashed red) frequencies.

$\hat{f} = f \pm 1/365$, respectively. Other aliasing frequencies, such as $2f$, $3f$, $0.5f$, etc., are also noticeable. Lastly, due to the sampling pattern of some light curves, the side lobes of the main peak can be higher than the central value. These manifest as close parallel strips to the aforementioned features.

4.4.2 Accuracy Assessment

The estimated frequency is considered as correct if $\Delta f = |\hat{f} - f_0| < C_f$ for each light curve. The estimation accuracies for the two methods are summarized in Table 4.1 for several different values of C_f. We choose $C_f = 2.7 \times 10^{-4}$ to stringently bind the one-to-one strip in Fig. 4.7. Overall, the SP correctly estimates the period for 69.4% of the light curves, while the LS model has an accuracy of 63.6%. The improvement of SP over LS is more evident for C-rich Miras, with about 10% higher accuracy, while the improvement for O-rich Miras is smaller, with about 3% higher accuracy. The difference in performance arises because C-rich Miras often exhibit larger stochastic deviations that can be better captured by the SP model, while O-rich Miras have more stable light curves that can be modeled reasonably well with the LS method. We also compute the estimation accuracy
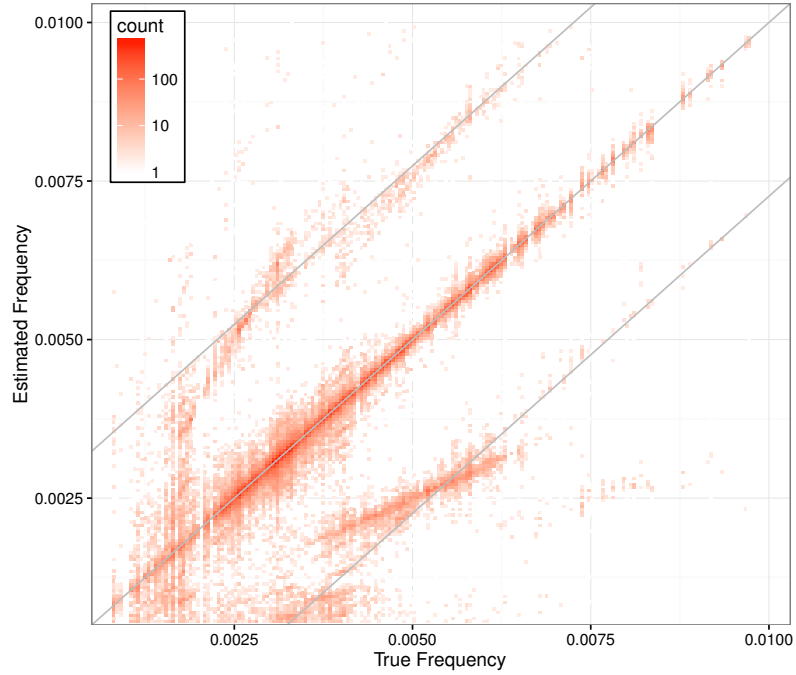
Figure 4.7: Model-recovered vs. true frequencies of the test data set. Features other than the one-to-one line are due to aliasing from one-year beat or harmonic frequencies. The three gray diagonal lines correspond to $\hat{f} = f + 1/365, f$ and $f - 1/365$ from top to bottom, respectively.

of each method by grouping the light curves according to the number of observations, as shown in the left panels of Fig. 4.8. The top and bottom rows show results for C- and O-rich Miras, respectively. The performance difference is once again more evident in the C-rich category.

Note that accuracy is not a monotonic function of the number of observations, implying this is not a good indicator *per se* of the information content of the light curves for frequency (period) estimation. Thus, we define another metric, called *phase coverage*. Recall that the times of observation for a given light curve are $t_1, t_2, \cdots, t_n$. Given a period of P, these are converted into corresponding phases by $s_i = (t_i \bmod P)/P$, $i = 1, 2, \cdots, n$ in
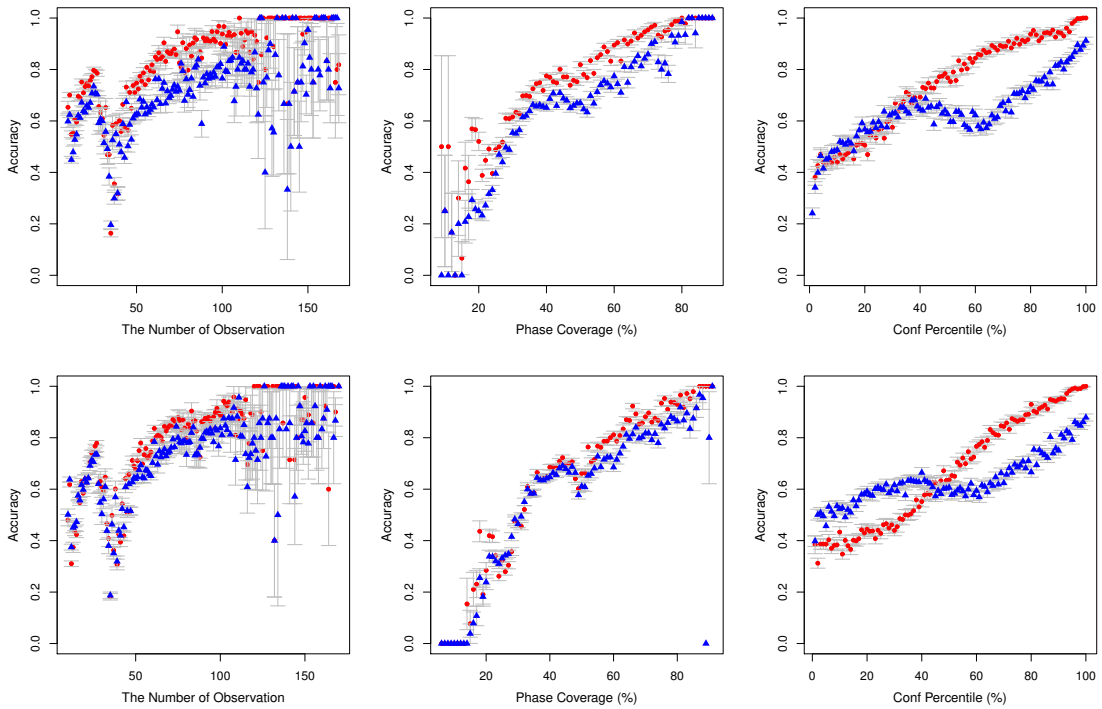
Figure 4.8: Accuracy comparison using different metrics. Light curves are grouped by number of observations (first column), phase coverage (second column), and `conf` values (third column). The estimation accuracies in each group is computed and plotted as above. The red circles represent the SP model, while blue triangles denote the GLS model. The upper panels are for C-rich Miras, and the lower panels are for O-rich Miras.
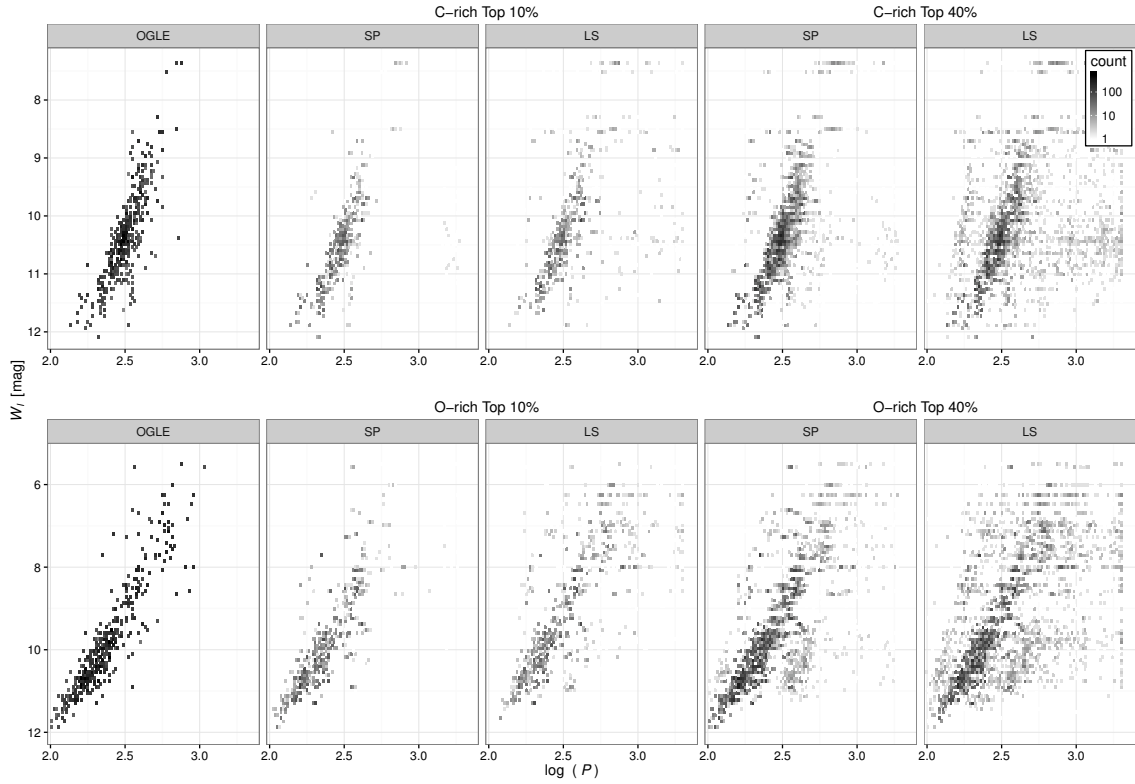
Figure 4.9: Actual and reconstructed Period-Luminosity relations for C-rich (top) and O-rich (bottom) Miras. The leftmost column shows the actual PLRs using periods and W_I magnitudes from Soszyński et al. (2009). The other columns use the same magnitudes but periods based on the SP or LS algorithms, as indicated in each panel. Recovered PLRs are plotted for various subsets selected according to the `conf` values obtained with the respective method.

the closed interval $[0, 1]$. Now, define

$$J = \left(\bigcup_i (s_i - l, s_i + l) \right) \cap [0, 1],$$

for a specific $l > 0$, the phase coverage can be measured by $\lambda(J)$ where $\lambda(\cdot)$ is the Lebesgue measure (we choose $l = 0.02$). $\lambda(J)$ describes the "length" of the union of the intervals J. For example, $\lambda(J) = 0.1$ for $J = (0.1, 0.2)$, and $\lambda(J) = 0.2$ for $J = (0.1, 0.2) \cup (0.5, 0.6)$.

We divide the light curves into 100 groups such that their $\lambda(J)$ is in one of the intervals $(k/100, (k+1)/100]$ for $k = 0, 1, \cdots, 99$ and compute the estimation accuracy for each subset. The results for the two models are plotted in the middle column of Fig. 4.8. Now the estimation accuracy is monotonically increasing as a function of phase coverage. The accuracy improvement of our method is highest when the phase coverage is around 0.5 for C-rich Miras. As the phase coverage approaches the extremes (0 or 1), the performance difference between the two methods diminishes. At $\lambda(J) \approx 0$, both methods will fail because this is a hopeless situation. At the other extreme, when $\lambda(J) \approx 1$ and abundant information is available for frequency estimation, both methods have an accuracy close to 1.

The periodogram $S_{\mathrm{SP}}(f)$ of our model defined in Eqn. 4.12 provides more information than just the optimal frequency. Suppose f_1 is the largest local maximal (global maximum) of $S_{\mathrm{SP}}(f)$, and f_2 is the second largest local maximal of $S_{\mathrm{SP}}(f)$. Now define `conf` $= S_{\mathrm{SP}}(f_1) - S_{\mathrm{SP}}(f_2) \geq 0$. The value of `conf` serves as a confidence measurement of the global optimal estimate in Eqn. 4.11. Larger values of `conf` indicate smaller uncertainty in our estimate, and thereby the estimate is more reliable. Now, let c_0 be the smallest value, and let $c_1, \ldots, c_{100}$ be the 1st–100th percentiles of all the `conf` values computed for all the light curves. Each light curve can be assigned to a percentile group if its `conf` is in

81

$(c_{k-1}, c_k]$ for some $k \in \{1, 2, \cdots, 100\}$. After assigning all light curves by `conf` to their corresponding percentile groups, the estimation accuracy in each group can be computed. The same procedure is applied to the GLS model, with the p-value of the F-statistics given in Eqn. 4.6 for the top peak being used as its `conf` measurement. The result is plotted in the right column of Fig. 4.8. The accuracy of our SP method is much higher than the LS model in the top 40 groups. In particular, the accuracy of our method is higher than 90% in the top 20 groups for both C- and O-rich Miras.
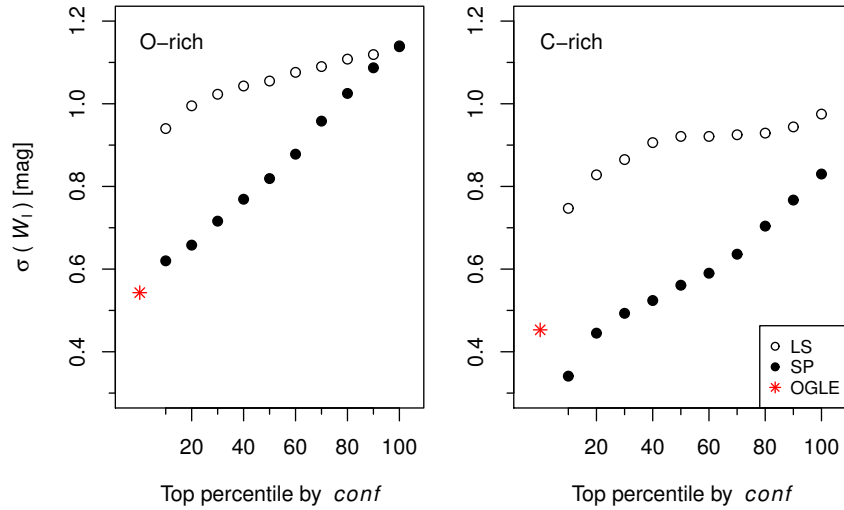
Figure 4.10: Dispersion of reconstructed Wesenheit PLRs for different sets of artificial light curves, based on their `conf` value. Top: O-rich PLRs. Bottom: C-rich PLRs. The starred symbols show the dispersion of the actual OGLE periods and W_I magnitudes. The filled symbols show the dispersion of the recovered PLRs based on SP-derived periods. The open symbols show the corresponding values for LS-derived periods.

Light curves with high values of `conf` are particularly reliable for constructing Period-Luminosity relations (hereafter, PLRs) based on the "Wesenheit" function (Madore, 1982). This function enables a simultaneous correction for the effects of dust attenuation and finite width of the instability strip by defining a new magnitude $W_I = I - 1.55(V - I)$, where V

Table 4.1: Comparison of estimation accuracies (%)

C_f (10^{-4})	Method	Class C-rich	O-rich	Both
1.0	SP	58.1	55.3	56.5
	LS	49.4	51.6	50.6
2.0	SP	69.6	63.8	66.3
	LS	60.1	60.6	60.4
2.7	SP	73.5	66.2	69.4
	LS	63.7	63.5	63.6
N light curves		43,116	56,884	10^5

and I are the mean magnitudes in those filters. Figure 4.9 compares PLRs based on W_I magnitudes and periods determined by OGLE and estimated with each of the two models. The top and bottom rows display the PLRs for C- and O-rich Miras, respectively. The leftmost column shows the PLRs based on the actual OGLE periods, while the next two sets of columns show the corresponding relations based on SP or LS periods for the simulated light curves with the top 10% and 40% values of conf.

In order to provide a quantitative comparison of the improvement obtained with our SP method, we calculated the dispersion of the actual W_I PLRs and their recovered counterparts as a function of conf value as follows, separately for C- and O-rich Miras. First, we selected all objects of a given class with $2 < \log P < 3$. If the Soszyński et al. (2009) catalog did not provide a V measurement for a given variable, the missing value was estimated through linear interpolation of the $(I, V - I)$ relation for objects of the same class within $|\Delta \log P| < 0.05$ dex. We fitted a quadratic PLR

$$m = a + b(\log P - 2.3) + c(\log P - 2.3)^2$$

with iterative 3σ clipping (removing $\sim 5\%$ of the data). We then computed the dispersion

of the initially selected OGLE sample about the best-fit relation, including outliers. This yielded "benchmark" dispersions of 0.45 & 0.54 mag for C- & O-rich variables, respectively. Keeping the best-fit relation fixed, we computed the dispersion of recovered PLRs using all artificial light curves within a certain range of `conf` (top 10%, top 20%, ...), using the periods and `conf` values derived by the SP or the LS method. As in the case of the OGLE samples, we only considered objects with $2 < \log P < 3$. The results are plotted in Fig. 4.10. The SP subsamples exhibit lower (or at worst, equal) dispersions than their LS counterparts for all percentiles and for both subtypes. As discussed previously, the improvement provided by our method is strongest for C-rich Miras and diminishes in significance as one includes light curves with progressively lower confidence values.

5. SUMMARY

In our work, we propose two light curve models for SNIa and Mira variables. They are developed in the framework of functional data analysis in statistics, and demonstrated to improve over existing methods. The Mira light curve model combines a dominant sinusoid component with Gaussian process. It has stronger power to estimate period, and is tested on a large-scale simulation.

In addition, we have presented an empirical model for SNIa light curves. Using this model, the entire light curve of a SNIa can be represented by a few scores. These scores characterize light curve shape, intrinsic color, and color excess for SNIa. Some light curve scores are even correlated with spectral features measured independently of SNIa light curves. In previous studies, the absorption features of SNIa spectra have been empirically compared with the color and the light curve width parameter. For example, Silverman et al. (2012a) showed the strength of Si II $\lambda 4000$ is anti-correlated with SALT II width parameter and no correlation with color. This anti-correlation only implies stronger Si II $\lambda 4000$ correlated with narrower light curve shape. However it is interesting to explore further how the light curve shape in multi-bands changes with the strength of this line. The score parameters from our model reveal more such information. Regarding this, we have presented a more detailed morphology analysis of light curve with respect to the feature strength.

Moreover, by examining the correlation among scores, especially the first two dominant scores, we find the SNIa light curve resides in a tight nonlinear subspace. A more dramatic dimension reduction is possible by nonlinear dimension reduction techniques. This tight nonlinear subspace specifies plausible parameter domains for SNIa light curves. A new light curve can be classified as SNIa if its scores are inside this subspace. On the other hand, a photometrically different SN can also be identified if its score vector is far

away from this subspace. Our algorithm may produce a quantitative photometric classi-fication scheme for supernovae. These classification and outlier detection task have been carried out previously by Hsiao et al. (2014), Ishida and de Souza (2013). In this paper, an initial analysis of nonlinear dimension reduction shows some promises. More thorough work is left for future study.

Beyond these empirical investigations, the proposed model embraces more potential in cosmology model fitting. Although the primary light curve shape parameter such as ΔM_{15}, the stretch parameter, or the parameter in SALT II is effective, it is still worthwhile to explore other constructions using the shape of the entire light curve. Estimation of ΔM_{15} is sensitive to local observations around the peak and around the +15 days in phase. If we lack enough observations to constraint light curve shapes around these days in phase, the estimated ΔM_{15} would have large uncertainty. Besides, the ΔM_{15} parameter only captures the declining part of the light curves, and fails to capture the light curve shape at the rising side. The stretch parameter may not be applicable for SN wavelength bands longer than I band, and may not always fit well for both the rising and falling part of a SNIa light curve. A product of our model is to replace the ΔM_{15} term in the existing distance prediction models by a functional linear term. This functional linear term is a more flexible, robust and data-driven way to adjust the light curve shape for distance prediction. By comparing with the previous distance models using ΔM_{15} adjustment, we have demonstrated that using the functional linear form of the entire light curve consistently gives smaller residual scattering and robust distance prediction.

Among the effort to reduce distance prediction scatter, one common conjecture is that SNIa is not a homogeneous group. There exist subclasses of their own characteristics. Each subclass has its own dust correction and K-correction. Picking out a more homo-geneous subclass help to further reduce the dispersion. Some works endeavor to identify subclasses based on spectral data (Benetti et al., 2005; Branch et al., 2009; Wang et al.,

2009). The dispersion reduction is more significant by pairing supernovae with identical spectral features and applying pairwise dust correction (Fakhouri et al., 2015). Our study finds that, when only light curve data is available, the scores extracted from the light curve can still help to determine spectral classes, although with limited precision. After that, the dispersion can be reduced by fitting the distance and dust correction model within a sub-class of SNIa observations. Another potential application of this result is to improve the precision of K-correction, as the spectral template from the corresponding spectral class can be applied for this subclass of observations.

REFERENCES

R. V. Baluev. The impact of red noise in radial velocity planet searches: Only three planets orbiting GJ 581? *Monthly Notices of the Royal Astronomical Society*, 429(3): 2052–2068, 2013.

S. Benetti, E. Cappellaro, P. A. Mazzali, M. Turatto, G. Altavilla, F. Bufano, N. Elias-Rosa, R. Kotak, G. Pignata, M. Salvo, et al. The diversity of type Ia supernovae: evidence for systematics? *The Astrophysical Journal*, 623(2):1011, 2005.

S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 3(1):1–122, 2011.

D. Branch, L. C. Dang, and E. Baron. Comparative direct analysis of type Ia supernova spectra. V. Insights from a larger sample and quantitative subclassification. *Publications of the Astronomical Society of the Pacific*, 121(877):238, 2009.

G. L. Bretthorst. *Bayesian spectrum analysis and parameter estimation*, volume 48 of *Lecture Notes in Statistics*. Springer New York, 1988.

C. R. Burns, M. Stritzinger, M. Phillips, S. Kattner, S. Persson, B. F. Madore, W. L. Freedman, L. Boldt, A. Campillay, C. Contreras, et al. The Carnegie Supernova Project: light curve fitting with SNooPy. *The Astronomical Journal*, 141(1):19, 2010.

J.-F. Cai, E. J. Candès, and Z. Shen. A singular value thresholding algorithm for matrix completion. *SIAM Journal on Optimization*, 20(4):1956–1982, 2010.

L. Campbell. A tentative classification of long period variables. *Harvard College Observatory Reprints*, 21:1, 1925.

Y. Chen, A. Jalali, S. Sanghavi, and H. Xu. Clustering partially observed graphs via convex optimization. *The Journal of Machine Learning Research*, 15(1):2213–2238, 2014.

A. Conley, M. Sullivan, E. Hsiao, J. Guy, P. Astier, D. Balam, C. Balland, S. Basa, R. Carlberg, D. Fouchez, et al. SiFTO: an empirical method for fitting SN Ia light curves. *The Astrophysical Journal*, 681(1):482, 2008.

C. Contreras, M. Hamuy, M. Phillips, G. Folatelli, N. B. Suntzeff, S. Persson, M. Stritzinger, L. Boldt, S. González, W. Krzeminski, et al. The Carnegie Supernova Project: first photometry data release of low-redshift type Ia supernovae. *The Astronomical Journal*, 139(2):519, 2010.

T. J. Deeming. Fourier analysis with unequally-spaced data. *Astrophysics and Space Science*, 36(1):137–158, 1975.

H. Fakhouri, K. Boone, G. Aldering, P. Antilogus, C. Aragon, S. Bailey, C. Baltay, K. Barbary, D. Baugh, S. Bongard, et al. Improving cosmological distance measurements using twin type Ia supernovae. *The Astrophysical Journal*, 815(1):58, 2015.

M. Ganeshalingam, W. Li, A. V. Filippenko, C. Anderson, G. Foster, E. L. Gates, C. V. Griffith, B. J. Grigsby, N. Joubert, J. Leja, et al. Results of the Lick Observatory Supernova Search Follow-up Photometry Program: BVRI light curves of 165 type Ia supernovae. *The Astrophysical Journal Supplement Series*, 190(2):418, 2010.

I. S. Glass and T. L. Evans. The calibrating stars of the Mira P-L relation. *Monthly Notices of the RAS*, 343:67–74, July 2003.

J. Guy, P. Astier, S. Baumont, D. Hardin, R. Pain, N. Regnault, S. Basa, R. Carlberg, A. Conley, S. Fabbro, et al. SALT2: using distant supernovae to improve the use of type Ia supernovae as distance indicators. *Astronomy & Astrophysics*, 466(1):11–21, 2007.

P. Hall, J. Reimann, and J. Rice. Nonparametric estimation of a periodic function. *Biometrika*, 87(3):545–557, 2000.

M. Hamuy, M. Phillips, N. B. Suntzeff, R. A. Schommer, J. Maza, R. Smith, P. Lira, and R. Aviles. The morphology of type Ia supernovae light curves. *arXiv preprint astro-ph/9609063*, 1996.

M. Hicken, P. Challis, S. Jha, R. P. Kirshner, T. Matheson, M. Modjaz, A. Rest, W. M. Wood-Vasey, G. Bakos, E. J. Barton, et al. CfA3: 185 type Ia supernova light curves from the CfA. *The Astrophysical Journal*, 700(1):331, 2009.

M. Hicken, P. Challis, R. P. Kirshner, A. Rest, C. E. Cramer, W. M. Wood-Vasey, G. Bakos, P. Berlind, W. R. Brown, N. Caldwell, et al. CfA4: Light Curves for 94 type Ia supernovae. *The Astrophysical Journal Supplement Series*, 200(2):12, 2012.

E. Hsiao, A. Conley, D. Howell, M. Sullivan, C. Pritchet, R. Carlberg, P. Nugent, and M. Phillips. K-corrections and spectral templates of Type Ia supernovae. *The Astrophysical Journal*, 663(2):1187, 2007.

E. Hsiao, G. Pignata, F. Bufano, L. Galbany, G. Folatelli, M. Phillips, M. Hamuy, J. Anderson, T. de Jaeger, et al. Defining photometric peculiar type Ia supernovae. *The Astrophysical Journal*, 795(2):142, 2014.

E. E. Ishida and R. S. de Souza. Kernel PCA for type Ia supernovae photometric classification. *Monthly Notices of the Royal Astronomical Society*, 430(1):509–532, 2013.

Y. Ita, T. Tanabé, N. Matsunaga, Y. Nakajima, C. Nagashima, T. Nagayama, D. Kato, M. Kurita, T. Nagata, S. Sato, M. Tamura, H. Nakaya, and Y. Nakada. Variable stars in the Magellanic Clouds - II. The data and infrared properties. *Monthly Notices of the Royal Astronomical Society*, 353:705–712, Sept. 2004. doi: 10.1111/j.1365-2966.2004. 08126.x.

G. M. James, T. J. Hastie, and C. A. Sugar. Principal component models for sparse functional data. *Biometrika*, 87(3):587–602, 2000.

S. Jha, A. G. Riess, and R. P. Kirshner. Improved distances to type Ia supernovae with multicolor light curve shapes: MLCS2k2. *The Astrophysical Journal*, 659(1):122, 2007.

S. Kattner, D. C. Leonard, C. R. Burns, M. Phillips, G. Folatelli, N. Morrell, M. D. Stritzinger, M. Hamuy, W. L. Freedman, S. E. Persson, et al. The standardizability of type Ia supernovae in the near-infrared: evidence for a peak-luminosity versus decline-

rate relation in the near-infrared. *Publications of the Astronomical Society of the Pacific*, 124(912):114, 2012.

M. Kayano and S. Konishi. Sparse functional principal component analysis via regularized basis expansions and its application. *Communications in Statistics-Simulation and Computation*, 39(7):1318–1333, 2010.

R. Kessler, B. Bassett, P. Belov, V. Bhatnagar, H. Campbell, A. Conley, J. A. Frieman, A. Glazov, S. González-Gaitán, R. Hlozek, et al. Results from the supernova photometric classification challenge. *Publications of the Astronomical Society of the Pacific*, 122 (898):1415, 2010.

G. Kimeldorf and G. Wahba. Some results on tchebycheffian spline functions. *Journal of Mathematical Analysis and Applications*, 33(1):82–95, 1971.

J. Lafler and T. D. Kinman. An RR Lyrae Star survey with Ihe Lick 20-INCH Astrograph II. The calculation of RR Lyrae periods by electronic computer. *The Astrophysical Journals*, 11:216, June 1965. doi: 10.1086/190116.

D. Liu, X. Lin, and D. Ghosh. Semiparametric regression of multidimensional genetic pathway data: least-squares kernel machines and linear mixed models. *Biometrics*, 63 (4):1079–1088, 2007.

N. R. Lomb. Least-squares frequency analysis of unequally spaced data. *Astrophysics and Space Science*, 39:447–462, Feb. 1976. doi: 10.1007/BF00648343.

H. Ludendorff. Die veränderlichen Sterne. *Handbuch der Astrophysik*, 6:49, 1928.

L. M. Macri, K. Z. Stanek, D. D. Sasselov, M. Krockenberger, and J. Kaluzny. The DIRECT Project: Catalogs of stellar objects in nearby galaxies. I. The central part of M33. *The Astronomical Journal*, 121:861–869, Feb. 2001. doi: 10.1086/318772.

L. M. Macri, C.-C. Ngeow, S. M. Kanbur, S. Mahzooni, and M. T. Smitka. Large Magellanic Cloud Near-Infrared Synoptic Survey. I. Cepheid Variables and the calibration of the Leavitt Law. *The Astronomical Journal*, 149:117, Apr. 2015. doi:

10.1088/0004-6256/149/4/117.

B. F. Madore. The Period-Luminosity relation. IV - intrinsic relations and reddenings for the Large Magellanic Cloud Cepheids. *The Astrophysical Journal*, 253:575–579, Feb. 1982. doi: 10.1086/159659.

K. S. Mandel, W. M. Wood-Vasey, A. S. Friedman, and R. P. Kirshner. Type Ia supernova light-curve inference: Hierarchical Bayesian analysis in the near-infrared. *The Astrophysical Journal*, 704(1):629, 2009.

K. S. Mandel, G. Narayan, and R. P. Kirshner. Type Ia supernova light curve inference: Hierarchical models in the optical and near-infrared. *The Astrophysical Journal*, 731 (2):120, 2011.

V. I. Marsakova. Fast variations of the mean brightness and Other light curve parameters of the carbon Mira-type star S Cep. *Journal of the American Association of Variable Star Observers (JAAVSO)*, 27:141–145, Oct. 1999.

M. C. Meyer. Constrained penalized splines. *Canadian Journal of Statistics*, 40(1):190–206, 2012.

H.-G. Müller and U. Stadtmüller. Generalized functional linear models. *Annals of Statistics*, pages 774–805, 2005.

K. P. Murphy. *Machine learning: a probabilistic perspective*. MIT press, 2012.

J. Newling, M. Varughese, B. Bassett, H. Campbell, R. Hlozek, M. Kunz, H. Lampeitl, B. Martin, R. Nichol, D. Parkinson, et al. Statistical classification techniques for photometric supernova typing. *Monthly Notices of the Royal Astronomical Society*, 414(3): 1987–2004, 2011.

J. Nocedal and S. Wright. *Numerical optimization*. Springer Science & Business Media, 2006.

P. Nugent, M. Phillips, E. Baron, D. Branch, and P. Hauschildt. Evidence for a spectroscopic sequence among type Ia supernovae. *The Astrophysical Journal Letters*, 455(2):

L147, 1995.

P. Nugent, A. Kim, and S. Perlmutter. K-corrections and extinction corrections for type Ia supernovae. *Publications of the Astronomical Society of the Pacific*, 114(798):803, 2002.

F. O'sullivan, B. S. Yandell, and W. J. Raynor Jr. Automatic smoothing of regression functions in generalized linear models. *Journal of the American Statistical Association*, 81(393):96–103, 1986.

U. Ozertem and D. Erdogmus. Locally defined principal curves and surfaces. *Journal of Machine Learning Research*, 12(Apr):1249–1286, 2011.

A. Pellerin and L. M. Macri. The M33 Synoptic Stellar Survey. I. Cepheid variables. *the Astrophysical Journals*, 193:26, Apr. 2011. doi: 10.1088/0067-0049/193/2/26.

J. Peng and D. Paul. A geometric approach to maximum likelihood estimation of the functional principal components from sparse longitudinal data. *Journal of Computational and Graphical Statistics*, 18(4):995–1015, 2009.

S. Perlmutter, G. Aldering, G. Goldhaber, R. Knop, P. Nugent, P. Castro, S. Deustua, S. Fabbro, A. Goobar, D. Groom, et al. Measurements of Ω and Λ from 42 high-redshift supernovae. *The Astrophysical Journal*, 517(2):565, 1999.

S. Pezzulli and B. Silverman. Some properties of smoothed principal components analysis for functional data. *Computational Statistics*, 8:1–1, 1993.

M. Phillips. The absolute magnitudes of type IA supernovae. *The Astrophysics Journal*, 413:L105–L108, 1993. doi: 10.1086/186970.

M. Phillips, P. Lira, N. B. Suntzeff, R. Schommer, M. Hamuy, and J. Maza. The reddening-free decline rate versus luminosity relationship for type Ia supernovae. *The Astronomical Journal*, 118(4):1766, 1999.

J. L. Prieto, A. Rest, and N. B. Suntzeff. A new method to calibrate the magnitudes of type Ia supernovae at maximum light. *The Astrophysical Journal*, 647(1):501, 2006.

B. Quinn and P. Thomson. Estimating the frequency of a periodic function. *Biometrika*, 78(1):65–74, 1991.

J. O. Ramsay and B. W. Silverman. *Functional data analysis*. Springer Science+ Business Media, 2010.

C. E. Rasmussen and C. K. I. Williams. *Gaussian processes for machine learning (adaptive computation and machine learning)*. The MIT Press, 2005. ISBN 026218253X.

J. D. Reimann. *Frequency estimation using unequally-spaced astronomical data.* PhD thesis, University of California, Berkeley, Jan. 1994.

J. W. Richards, D. Homrighausen, P. E. Freeman, C. M. Schafer, and D. Poznanski. Semi-supervised learning for photometric supernova classification. *Monthly Notices of the Royal Astronomical Society*, 419(2):1121–1135, 2012.

A. G. Riess, W. H. Press, and R. P. Kirshner. A precise distance indicator: type Ia supernova multicolor light-curve shapes. *The Astrophysical Journal*, 473(1):88, 1996.

A. G. Riess, A. V. Filippenko, P. Challis, A. Clocchiatti, A. Diercks, P. M. Garnavich, R. L. Gilliland, C. J. Hogan, S. Jha, R. P. Kirshner, et al. Observational evidence from supernovae for an accelerating universe and a cosmological constant. *The Astronomical Journal*, 116(3):1009, 1998.

A. Rohde, A. B. Tsybakov, et al. Estimation of high-dimensional low-rank matrices. *The Annals of Statistics*, 39(2):887–930, 2011.

J. D. Scargle. Studies in astronomical time series analysis. II - Statistical aspects of spectral analysis of unevenly spaced data. *the Astrophysical Journal*, 263:835–853, Dec. 1982. doi: 10.1086/160554.

E. F. Schlafly and D. P. Finkbeiner. Measuring reddening with Sloan Digital Sky Survey stellar spectra and recalibrating SFD. *The Astrophysical Journal*, 737(2):103, 2011.

A. Schwarzenberg-Czerny. Fast and statistically optimal period search in uneven sampled observations. *the Astrophysical Journall*, 460:L107, Apr. 1996. doi: 10.1086/309985.

H. L. Shang. A survey of functional principal component analysis. *AStA Advances in Statistical Analysis*, 98(2):121–142, 2014.

B. W. Silverman et al. Smoothed functional principal components analysis by choice of norm. *The Annals of Statistics*, 24(1):1–24, 1996.

J. M. Silverman, M. Ganeshalingam, W. Li, and A. V. Filippenko. Berkeley Supernova Ia Program–III. Spectra near maximum brightness improve the accuracy of derived distances to type Ia supernovae. *Monthly Notices of the Royal Astronomical Society*, 425 (3):1889–1916, 2012a.

J. M. Silverman, J. J. Kong, and A. V. Filippenko. Berkeley Supernova Ia Program–II. Initial analysis of spectra obtained near maximum brightness. *Monthly Notices of the Royal Astronomical Society*, 425(3):1819–1888, 2012b.

I. Soszynski, W. A. Dziembowski, A. Udalski, M. Kubiak, M. K. Szymanski, G. Pietrzynski, L. Wyrzykowski, O. Szewczyk, and K. Ulaczyk. The Optical Gravitational Lensing Experiment. Period–Luminosity relations of variable red giant stars. *Acta Astronomica*, 57:201–225, Sept. 2007.

I. Soszyński, A. Udalski, M. K. Szymański, M. Kubiak, G. Pietrzyński, Ł. Wyrzykowski, O. Szewczyk, K. Ulaczyk, and R. Poleski. The Optical Gravitational Lensing Experiment. The OGLE-III Catalog of variable stars. IV. long-period variables in the Large Magellanic Cloud. *Acta Astronomica*, 59:239–253, Sept. 2009.

D. Spergel, N. Gehrels, J. Breckinridge, M. Donahue, A. Dressler, B. Gaudi, T. Greene, O. Guyon, C. Hirata, J. Kalirai, et al. Wide-field infrared survey telescope-astrophysics focused telescope assets wfirst-afta final report. *arXiv preprint arXiv:1305.5422*, 2013.

M. D. Stritzinger, M. Phillips, L. N. Boldt, C. Burns, A. Campillay, C. Contreras, S. Gonzalez, G. Folatelli, N. Morrell, W. Krzeminski, et al. The Carnegie Supernova Project: second photometry data release of low-redshift type Ia supernovae. *The Astronomical Journal*, 142(5):156, 2011.

M. R. Templeton and M. Karovska. Long-period variability in o Ceti. *the Astrophysical Journal*, 691:1470–1478, Feb. 2009. doi: 10.1088/0004-637X/691/2/1470.

M. R. Templeton, J. A. Mattei, and L. A. Willson. Secular evolution in Mira variable pulsations. *The Astronomical Journal*, 130:776–788, Aug. 2005. doi: 10.1086/431740.

N. M. Tran. *An introduction to theoretical properties of functional principal component analysis*. PhD thesis, Department of Mathematics and Statistics, The University of Melbourne, 2008.

A. Udalski, M. K. Szymanski, I. Soszynski, and R. Poleski. The Optical Gravitational Lensing Experiment. Final reductions of the OGLE-III data. *Acta Astronomica*, 58: 69–87, June 2008.

A. Wagers, L. Wang, and S. Asztalos. Quantifying spectral features of Type Ia supernovae. *The Astrophysical Journal*, 711(2):711, 2010.

L. Wang, G. Goldhaber, G. Aldering, and S. Perlmutter. Multicolor light curves of type Ia supernovae on the color-magnitude diagram: A novel step toward more precise distance and extinction estimates. *The Astrophysical Journal*, 590(2):944, 2003.

X. Wang, A. Filippenko, M. Ganeshalingam, W. Li, J. Silverman, L. Wang, R. Chornock, R. Foley, E. Gates, B. Macomber, et al. Improved distances to type Ia supernovae with two spectroscopic subclasses. *The Astrophysical Journal Letters*, 699(2):L139, 2009.

Y. Wang, R. Khardon, and P. Protopapas. Nonparametric bayesian estimation of periodic light curves. *The Astrophysical Journal*, 756(1):67, 2012.

P. R. Wood, C. Alcock, R. A. Allsman, D. Alves, T. S. Axelrod, A. C. Becker, D. P. Bennett, K. H. Cook, A. J. Drake, K. C. Freeman, K. Griest, L. J. King, M. J. Lehner, S. L. Marshall, D. Minniti, B. A. Peterson, M. R. Pratt, P. J. Quinn, C. W. Stubbs, W. Sutherland, A. Tomaney, T. Vandehei, and D. L. Welch. MACHO observations of LMC red giants: Mira and semi-regular pulsators, and contact and semi-detached binaries. In T. Le Bertre, A. Lebre, and C. Waelkens, editors, *Asymptotic Giant Branch*

Stars, volume 191 of *IAU Symposium*, page 151, 1999.

J. Wright, A. Ganesh, S. Rao, Y. Peng, and Y. Ma. Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization. In *Advances in neural information processing systems*, pages 2080–2088, 2009.

F. Yao, H.-G. Müller, and J.-L. Wang. Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*, 100(470):577–590, 2005.

M. Zechmeister and M. Kürster. The generalised Lomb-Scargle periodogram. A new formalism for the floating-mean and Keplerian periodograms. *Astronomy and Astrophysics*, 496:577–584, Mar. 2009. doi: 10.1051/0004-6361:200811296.

X. Zhao, K. Maeda, X. Wang, L. Wang, H. Sai, J. Zhang, T. Zhang, F. Huang, and L. Rui. The xxygen features in Type Ia supernovae and the implications for the nature of thermonuclear explosions. *arXiv preprint arXiv:1605.07781*, 2016.

L. Zhou, J. Z. Huang, and R. J. Carroll. Joint modelling of paired sparse functional data using principal components. *Biometrika*, 95(3):601–619, 2008.