

This is the accepted manuscript made available via CHORUS. The article has been published as:

## Artificial neural network for constructing type Ia supernovae spectrum evolution model

Qiao-Bin Cheng, Chao-Jun Feng, Xiang-Hua Zhai, and Xin-Zhou Li

Phys. Rev. D **97**, 123530 — Published 20 June 2018

DOI: [10.1103/PhysRevD.97.123530](https://doi.org/10.1103/PhysRevD.97.123530)

# Artificial Neural Network for Constructing Type Ia Supernovae Spectrum Evolution Model

Qiao-Bin Cheng,<sup>1</sup> Chao-Jun Feng,<sup>1,\*</sup> Xiang-Hua Zhai,<sup>1,†</sup> and Xin-Zhou Li<sup>1,‡</sup>

<sup>1</sup>*Shanghai United Center for Astrophysics (SUCA),  
Shanghai Normal University, 100 Guilin Road, Shanghai 200234, P.R.China*

We construct and train an artificial neural network called the back-propagation neural network to describe the evolution of the type Ia supernova spectrum by using the data from the CfA Supernova Program. This network method has many attractive features, and one of them is that the constructed model is differentiable. Benefitting from this, we calculate the absorption velocity and its variation. The model we constructed can well describe not only the spectrum of SNe Ia with wavelength range from 3500Å to 8000Å, but also the light-curve evolution with phase time from -15 to 50 with different colors. Moreover, the number of parameters needed during the training process is much less than the usual methods.

PACS numbers:

## I. INTRODUCTION

Type Ia supernova (SN Ia) is regarded as the standard candle to measure the distance on cosmological scales, since all of them have almost the same intrinsic brightness. By using the SN Ia distance indicator, the accelerating expansion of the universe was discovered in 1998 [1, 2]. The standard model of SN Ia involves the thermonuclear disruption of a carbon-oxygen white dwarf star as it approaches the Chandrasekhar mass, and most type Ia supernovae are very similar in their photometric and spectroscopic properties. The SN Ia may be generated through the merging of components in close binaries [3]. A correlation between the peak luminosity and the shape of the early light curve was also found, with brighter objects having a lower rate of decline than dimmer ones [4], which is matched by a spectroscopic sequence.

The model of the spectroscopic sequence is constructed by a training process, such as the SALT2 model[5], in which the mean evolution of the spectral energy distribution (SED) sequence of SN Ia and its variation with color are modeled as a functional form, see Eq.(1) in[5]. During the training process of SALT2, one ends up with more than 3000 parameters[5] to fit, due to the obvious non-linearities of the SED.

In the past decades, among the various machine intelligence procedures, artificial neural network (ANN) methods have been established as powerful techniques to solve a variety of real-world problems because of its excellent learning capacity [6, 7]. ANN is one of the popular areas of artificial intelligence research and also an abstract computational model based on the organizational structure of the human brain [6]. In this paper, we will construct and train an ANN called the back-propagation

neural network (BPNN) to describe the evolution of the SN Ia spectrum. The inputs of our ANN are the phase (or time), wavelength and also the color from the data that will be described later, while the output is just the corresponding flux of the SN Ia. The SED we trained is almost the same as that in the SALT2 model. However, the employment of a neural architecture adds many attractive features:

- One does't need to assume a functional form of the SED model during the ANN's training process.
- The SED via ANNs is differentiable, then it is easily used in any subsequent calculations, e.g. the calculation of the absorption velocity gradients of Si II  $\lambda$ 6355 line of SN Ia in the following.
- The required number of model parameters is far less than the traditional methods.
- The method can be realized in hardware, using neuroprocessors, and hence offer the opportunity to handle the real-time analysis of the SN Ia's spectra.
- The method can also be efficiently implemented on parallel architectures, such as the data parallelism, in which the data set is split into small ones and each of them is feed to a processor.

Si II  $\lambda$ 6355 is one of the strongest features in optical/near-infrared spectra of SN Ia; the blueshift of its absorption minimum has often been used to diagnose the diversity among SN Ia [8]. Different subclass evolutions of Si II  $\lambda$ 6355 absorption velocity ( $v_{abs}$ ) have been compared in Ref.[9], while the relation between its gradient and the parameter  $\Delta m_{15}$  of SN Ia has been studied in [10]. In this paper, we have studied the properties of  $v_{abs}$  and its gradients by using the spectrum evolution model after ANN's training. The results are fully consistent with those in [11, 12].

The structure of this paper is as follows. In Section II, we construct the neural network, and then train it

---

\*Electronic address: fengcj@shnu.edu.cn

†Electronic address: zhaixh@shnu.edu.cn

‡Electronic address: kyhz@shnu.edu.cn

in Section III. The spectra data set we used for training is described in Section IV, while the training results are presented in Section V, and the  $v_{abs}$  and its gradient are also computed in this section. In Section VI the relation between  $v_{abs}$  and color is shown. Finally, discussions and conclusions are given in Section VII.

## II. NETWORK CONSTRUCTION

In the following, the neural network will be constructed, which is called the back-propagation neural network. This kind of ANN has already been used in astronomy and physics, for example, to classify the type of a supernova (e.g. Ia, Ib, II, etc.) by using the neural network from the multi-frequency observations of its light curve, see [13] and references therein. The structure (or topology) of the BPNN could be described as Fig.1: where a

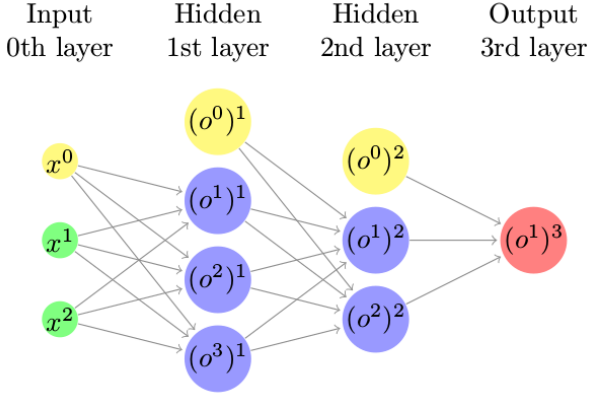


FIG. 1: Typical structure of an ANN.

few hidden layers and a few neural neurons in each layer are presented for limited space. In general, there could be totally  $L + 2$  layers in the network, and we will use  $l = 0, 1, 2, \dots, L, L + 1$  to denote them, where  $l = 0$  is called the input layer,  $l = L + 1$  the output layer, and others are all called the hidden levels.

Let  $N^l$  denote the number of neurons at level  $l$ , with  $N^0$  and  $N^{L+1}$  being also called the dimensions of the input and the output, respectively. In this paper, we consider two cases with  $N^0 = 2$  and  $3$ , and  $N^{L+1} = 1$  for both.

Between the  $l$ -th and the  $(l-1)$ -th level, we have the weights :

$$\mathbf{W}^l \equiv (W_j^i)^l \in \mathcal{R}^{N^L \times (N^{L-1}+1)} \quad (1)$$

for  $l \geq 1$ , where  $i = 1, 2, \dots, N^l, j = 0, 1, 2, \dots, N^{l-1}$ , and here we have included the bias vector in  $j = 0$ , namely

$$(W_0^i)^l = (b^i)^l. \quad (2)$$

By construction, the input of the  $l$ -th level is just the output of the  $(l-1)$ -th level plus the bias, so we set the output of the  $l$ -th level  $(o_m^i)^l$  as the following:

$$(o_m^i)^0 = (n_m^i)^0 = X_m^i, \quad l = 0, \quad (3)$$

$$(n_m^i)^l = \sum_{j=0}^{N^{l-1}} (W_j^i)^l (o_m^j)^{l-1}, \quad l \geq 1, \quad (4)$$

$$(o_m^i)^l = g_l((n_m^i)^l), \quad l \geq 1, \quad (5)$$

where  $i = 1, 2, \dots, N^l, m = 1, 2, \dots, M$  and  $M$  denotes the total number of the samples. Here, the function  $g_l(x)$  is the activation function, which is often taken as a sigmoid function or the tangent hyperbolic function

$$g(x) = \begin{cases} \frac{1}{1+e^{-x}} & \text{for sigmoid function,} \\ \tanh(x) & \text{for tanh(x),} \end{cases} \quad (6)$$

between the hidden layers, i.e.  $l \leq L$ . Straightforward, we also have:

$$\dot{g} = \frac{dg}{dx} = \begin{cases} g(1-g) & \text{for sigmoid function,} \\ 1-g^2 & \text{for tanh(x).} \end{cases} \quad (7)$$

For the output layer, we take the activation function as a linear function  $g(x) = x, \dot{g} = 1$ .

## III. NETWORK TRAINING

To train the network, one needs to minimize a cost function after feeding training samples. The cost function is also called the error function, which describes the error between the output and training samples. In our cases, the cost function is given by  $E = \mathbf{e}^T \mathcal{C}^{-1} \mathbf{e} / 2$ , or more specifically:

$$E = \frac{1}{2} \left( \vec{F}^{\text{obs}} - \vec{F}^{\text{ANN}} \right)^T \mathcal{C}^{-1} \left( \vec{F}^{\text{obs}} - \vec{F}^{\text{ANN}} \right), \quad (8)$$

where  $\mathcal{C}$  is the covariance matrix of the observational flux  $\vec{F}^{\text{obs}}$  and the flux  $\vec{F}^{\text{ANN}}$  is the output of BPNN for a supernova at the redshift  $z$  that depends on the phase  $p \equiv (t - t_{\text{max}}^B) / (1 + z)$ , the wavelength  $\lambda$  and maybe the color  $C$ . In the rest frame of a supernova,  $p$  is just the days after ( $p > 0$ ) or before ( $p < 0$ ) the date on which it has a maximum luminosity. Furthermore, the error for a given sample  $m$  in Eq.(8) reads:

$$e_m = F_m^{\text{obs}} - F_m^{\text{ANN}}. \quad (9)$$

In the following, we will perform the Levenberg-Marquardt (LM) algorithm to train the BPNN. In each step  $s$ , the following weighted normal equations will be solved

$$(\mathbf{J}^T \mathcal{C}^{-1} \mathbf{J} + \mu \mathbf{I}) \left( W_{s+1} - W_s \right) = -\mathbf{J}^T \mathcal{C}^{-1} \mathbf{e}, \quad (10)$$

to update each weight between the layers of the network for next step  $s + 1$ . Here  $\mathbf{I}$  is the identity matrix,  $\mu$  is the combination coefficient that could be changed during the training procedure as the following: If the updated parameters  $\Delta W = W_{s+1} - W_s$  computing from Eq.(10) leads to a reduction of the error, the updated is accepted and  $\mu$  is decreased by factor of 2 or so in the next. Otherwise,  $\mu$  is increased by a factor of 2 or so, and Eq.(10) is solved again to process until the decreased error is found. This process repeats until reaching the required precision.

The Jacobi matrix  $\mathbf{J} \in \mathcal{R}^{M \times N}$  is defined as

$$\begin{pmatrix} \frac{\partial e_1}{\partial W_1} & \frac{\partial e_1}{\partial W_2} & \cdots & \frac{\partial e_1}{\partial W_N} \\ \frac{\partial e_2}{\partial W_1} & \frac{\partial e_2}{\partial W_2} & \cdots & \frac{\partial e_2}{\partial W_N} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial e_M}{\partial W_1} & \frac{\partial e_M}{\partial W_2} & \cdots & \frac{\partial e_M}{\partial W_N} \end{pmatrix}, \quad (11)$$

where  $N = \sum_{l=1}^{L+1} N^l(N^{l-1} + 1)$  is the total number of weights ( including bias).

For a given sample  $m$  we also have the following recurrence relation

$$\frac{\partial e_m}{\partial \mathbf{W}^l} = \frac{\partial(F_m^{obs} - F_m^{ANN})}{\partial \mathbf{W}^l} = \frac{-\partial(o_m^{L+1})}{\partial \mathbf{W}^l}, \quad (12)$$

$$\frac{\partial(o_m^{L+1})}{\partial \mathbf{W}^l} = \delta^l (\mathbf{o}_m^{l-1})^T, \quad (13)$$

$$\delta^{L+1} = 1, \quad (14)$$

$$\delta^l = \mathbf{G}^l(\mathbf{n}_m^l)(\bar{\mathbf{W}}^l)^T \delta^{l+1}, \quad l \leq L, \quad (15)$$

where  $(\bar{\mathbf{W}}^l)$  are the weights not including the bias vector, i.e.  $(\bar{W}_j^l)^l = (\bar{W}_j^l)^l, j \neq 0$ , and

$$\mathbf{G}^l(\mathbf{n}_m^l) = \begin{pmatrix} \dot{g}^l((n_m^1)^l) & 0 & \cdots & 0 \\ 0 & \dot{g}^l((n_m^2)^l) & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \dot{g}^l((n_m^{N^l})^l) \end{pmatrix}. \quad (16)$$

In summary, by using LM algorithm, the update rule of weights can be presented as

$$W_{s+1} = W_s - (\mathbf{J}^T \mathcal{C}^{-1} \mathbf{J} + \mu \mathbf{I})^{-1} \mathbf{J}^T \mathcal{C}^{-1} \mathbf{e}. \quad (17)$$

Actually, we also have tried the famous RPROP algorithm[14] to train BPNN, but it turns out that the LM one is much better, especially when the input dimension is higher than 2.

#### IV. DATA ANALYSIS

The data used in this paper are obtained through the CfA Supernova Program, and there are 2603 spectra of 462 low- $z$  SNIa in total, see Refs.[9, 15–21]. It should be noticed that some of the spectra in the CfA data have no flux error information and in this case we estimate the flux error by averaging the errors of the other data. For the training, the recalibration (or re-normalization)

of spectra from the corresponding supernova needs the knowledge of  $t_{max}^B$ , i.e. the time when the flux has a maximum luminosity, so that spectra without the information of  $t_{max}^B$  will be disregarded. We only focus on  $-15 < p < 50$ , which is the most important phase range when dealing with the spectra. Beyond this range, we have either few spectra or slowly varying ones. The spectra of SN Ia with color  $C > 0.8$  and those without a spectrum in range  $-10 < p < 15$  are ignored, since the supernova has strong extinctions for large color. Actually, the impact is limited on training of including these outside range in the sense of that the trained weights of the network do not change much, but we do not think that the network could well describe the range outside of  $[-15, 50]$  in which there are only a few data points, and the color range  $C > 0.8$  in which the data have too strong extinctions to deal with when the covariance matrices are unknown. Here, the color of a SN Ia could be defined through the spectra:

$$C = 2.5 \log \left[ \frac{\int_{\lambda} \lambda T_V(\lambda) S_{SN}(\lambda) d\lambda}{\int_{\lambda} \lambda T_B(\lambda) S_{SN}(\lambda) d\lambda} \right], \quad (18)$$

where  $S_{SN}(\lambda)$  is the de-redshift spectrum nearest  $t_{max}^B$ , and  $T_B(\lambda)$  and  $T_V(\lambda)$  are the effective instrument transmissions in photometric band B and V, respectively. The color defined from spectroscopy in the equation above is almost the same properties as that defined from the photometry, see Ref.[15].

Finally, there are 1787 spectra of 238 SN Ia (about 4,600 thousand data points) will be left for the BPNN training. We have shown the number and redshift distributions of these samples in Fig.2 and Fig.3. From the number distribution of the spectra in Fig.2 (unfilled histogram), we find that many spectra were obtained near the max date  $t_{max}^B$  ( $p = 0$ ), but the average phase is about  $\bar{p} \sim 10.5$  because the distribution has a long tail. The height of each column of the filled histogram in Fig.2 indicates the number of SN Ia whose earliest spectrum is obtained at the phase  $p$ . It shows that more than half of the supernovae have the spectra earlier than the max date. From Fig.3, one can see that almost all the SN Ia are below  $z < 0.05$ , and the average value is  $\bar{z} \sim 0.02$ .

The  $\Delta m_{15}$  and color  $C$  distributions of the spectra are also plotted in Fig.4. The filled histogram tells us the average value of  $\Delta m_{15}$  is about  $\bar{\Delta m}_{15} \sim 1.15$  mag, while the unfilled one shows the average value of  $\bar{C} \sim 0.34$  mag.

To make best use of the spectroscopic data, one needs to recalibrate or re-normalize the data by using the corresponding SN Ia photometry, since in general the photometry is much better calibrated and uniform than its spectra. The recalibration is performed by multiplying each input spectrum by the factor:

$$f_{SN}^p = \frac{\bar{F}_R^{obs}(p)}{\int_{\lambda} \lambda T_R(\lambda) S_{SN}(\lambda)_p d\lambda} \quad (19)$$

where  $\bar{F}_R^{obs}(p)$  is the average value of the flux in the R

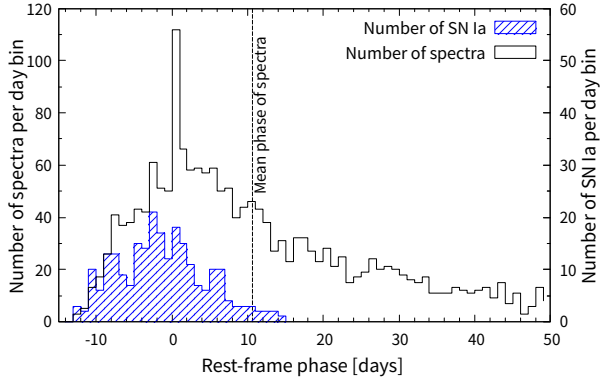


FIG. 2: Number distributions of the spectra and SN Ia.

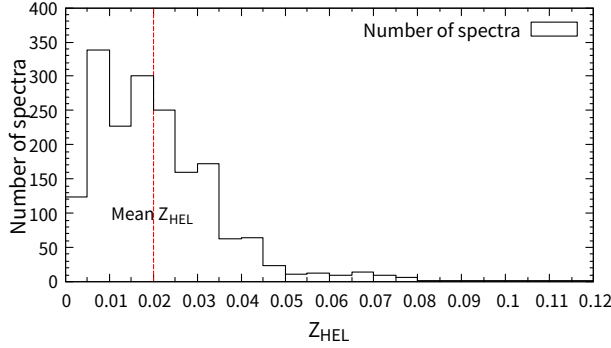


FIG. 3: Redshift distribution of the spectra.

band filter.

## V. TRAINING RESULTS

The output flux of BPNN  $\vec{F}^{ANN}$  in Eq.(8) could either depend on two inputs, i.e. phase  $p$  and wavelength  $\lambda$ , or three inputs, phase  $p$ , wavelength  $\lambda$  and color  $C$ . In the following, the networks of these two cases will be trained. For simplicity, we assume each hidden level of the network has the same number of neurons in both cases.

### A. Case I: $\vec{F}^{ANN} = \vec{F}^{ANN}(p, \lambda)$

In this case, the network will be trained by using the spectra in the wavelength range between  $3500\text{\AA}$  and  $8000\text{\AA}$  with a bin interval of  $10\text{\AA}$ . In each interval, we average the flux weighted by its variation. There are about 57,000 data points in total.

Fig.5 indicates the variation of the root mean squared error (RMSE) and the mean absolute error (MAE) vs the neuron number, where the BPNN only has one hidden level. One can clearly see that both RMSE and MAE

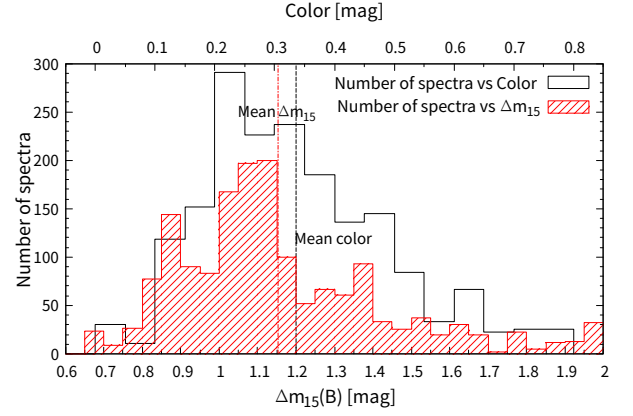


FIG. 4: Parameters distribution of the spectra.

decrease with the increasing of the neuron number. We have also compared different structure of network with three or four hidden levels, see Tab.I. The total number

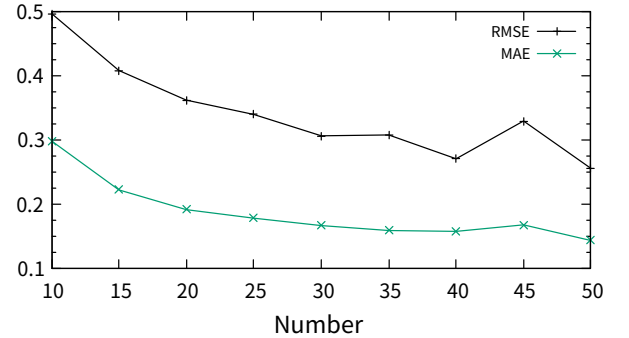


FIG. 5: The RMSE and MAE vs the neuron number in the hidden level for the BPNN with only one hidden level.

of parameters including weights and bias in the training network is less than 1000 except the last one, which has the most complicated structure. However, during the training process of SALT2, one usually ends up with more than 3000 parameters[5] to fit. To illustrate the ability of the BPNN method, a best-trained spectrum evolution model is presented in Figs.6-8.

To see whether the training is over-fit, we split the data into training and test set by randomly choosing the training points from the all data set. A typical value of RMSE and MAE between these two set is presented in the following table:

From Tab.II, one can see that the difference between the training and test set is small. Therefore, it seems that there is no over-fit issue for the networks we have trained.

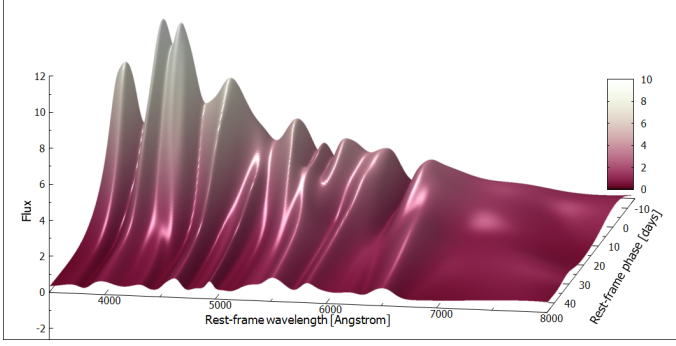


FIG. 6: 3D vision: A best-trained spectrum evolution model.

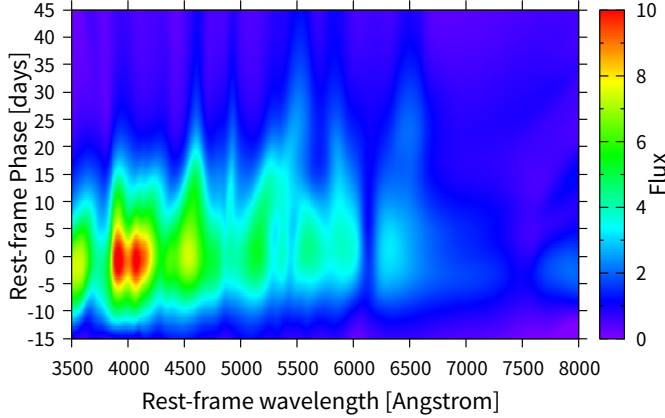


FIG. 7: Contour: A best-trained spectrum evolution model.

### B. Case II: $\vec{F}^{ANN} = \vec{F}^{ANN}(p, \lambda, C)$

It is well known that the K-correction of light-curve mainly depends on spectral color[22, 23], so in this case,

Structure	#	RMSE	MAE
2-10-10-1	151	0.4964	0.2981
2-15-15-1	301	0.4077	0.2220
2-20-20-1	501	0.3549	0.1871
2-10-10-10-1	261	0.4063	0.2206
2-15-15-15-1	541	0.3491	0.1866
2-20-20-20-1	921	0.3059	0.1690
2-10-10-10-10-1	371	0.3769	0.1988
2-15-15-15-15-1	781	0.3016	0.1705
2-20-20-20-20-1	1341	0.2780	0.1665

TABLE I: Results in the Case I with different network structure. The first column is the topology of the network, the second column is the total number of the weights (including the bias), and the third and forth columns are the final RMSE and MAE respectively. Note that, the values of RMSE and MAE are already divided by the number of data points.

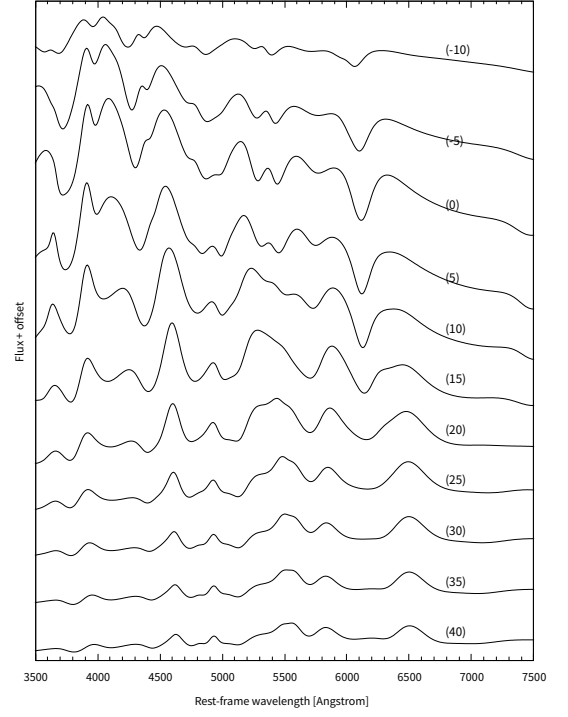


FIG. 8: Evolution of spectrum with phase, where the number in parentheses denotes the phase.

the color will be included as one input of the network and the full set of spectra data will be used instead of the bin one.

After training, we find the behavior of the RMSE and MAE is the same as that in the Case I. It is interesting that the color is almost a constant before the maximum luminosity day, after which it is proportional to the phase time, see Fig.9. We also compared the trained spectra with those in SALT2 at different phases, see Fig.10. One can see that the difference between them is very small at long wavelength, whereas at the short wavelength, the difference mainly depends on the color value; when  $C \sim 0.15$ , the model we trained is almost the same as that in SALT2. The RMSE/MAEs over the training and test set for the SALT2 and BPNN model are also given in Tab.III.

From Tab.III, one can see that the values of RMSE/MAE are slightly smaller for the BPNN model

	percent	number	RMSE	MAE
Training Set	80%	1416	0.4471	0.2699
Test Set	20%	371	0.4297	0.2643
Training Set	70%	1240	0.4303	0.2657
Test Set	30%	547	0.4778	0.2790

TABLE II: The data is split randomly to the training and test set according ( 80% to 20% or 70% to 30% split).

RMSE/MAE	SALT2	BPNN Model
Training Set (80%)	0.7098/0.4656	0.4471/0.2699
Test Set (20%)	0.6685/0.4489	0.4297/0.2643
Training Set (70%)	0.6900/0.4612	0.4303/0.2657
Test Set (30%)	0.7270/0.4642	0.4778/0.2790

TABLE III: The RMSE/MAEs over the training and test set for the SALT2 and BPNN model.

than that for the SALT2.

## VI. APPLICATION: ABSORPTION VELOCITY

As we stated in the introduction section, one of the attractive features of applying the ANN method is that the output of the network is differentiable. One can benefit from this property to perform the consequence calculation, e.g. to calculate the absorption velocity gradients of Si II  $\lambda 6355$  line of SN Ia.

The absorption velocity is defined by using the relativistic Doppler formula:

$$\frac{v_{abs}}{c} = \frac{(\lambda_{abs}/\lambda_0)^2 - 1}{(\lambda_{abs}/\lambda_0)^2 + 1}, \quad (20)$$

where  $\lambda_0$  is the wavelength of the corresponding transition, and  $c$  is the speed of light. The wavelength corresponding to the maximum absorption  $\lambda_{abs}(p)$  can be obtained from the trained network, i.e. the value of wavelength minimizes  $\vec{F}^{ANN}$  at phase  $p$ , see Fig.11. And then the changes of  $v_{abs}$  can be more efficiently obtained than those in Ref.[24], see Fig.12. The evolution is much faster at early times of the supernova's explosion, which is about  $v_{abs} \sim 13,000$  km/s at phase  $p \sim -8$ . Later,

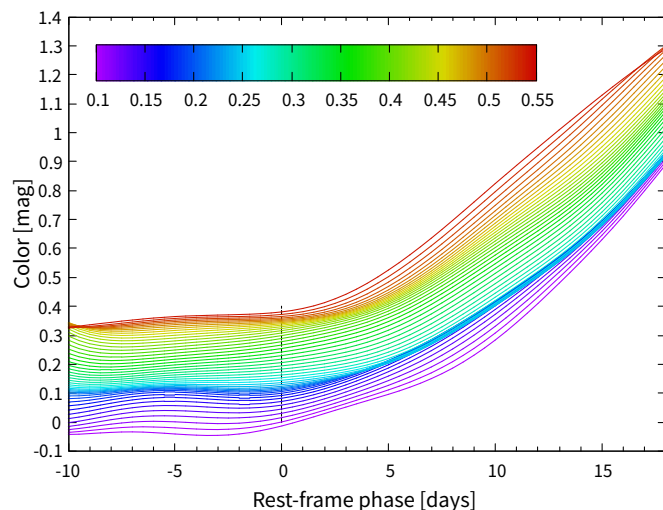


FIG. 9: Color vs phase.

the speed decreases to  $v_{abs} \sim 11,000$  km/s at the maximum date  $p \sim 0$  with range  $\sim 2000$  km/s for different colors, and it is nearly a constant during these phases. Finally, it will continue to decrease to  $v_{abs} \sim 9,000$  km/s at  $p \sim 25$  with range 1,000 km/s. The speed is almost a constant during  $0 < p < 5$ , i.e.  $\partial_t v_{abs} \sim 0$ , while  $\partial_t v_{abs} \sim 110$  km/s/day after phase  $p \sim 10$ .

## VII. CONCLUSION AND FUTURE WORK

We have constructed an SN Ia spectrum evolution model by training a back-propagation neural network with the observed nearby spectra from the CfA SN Pro-

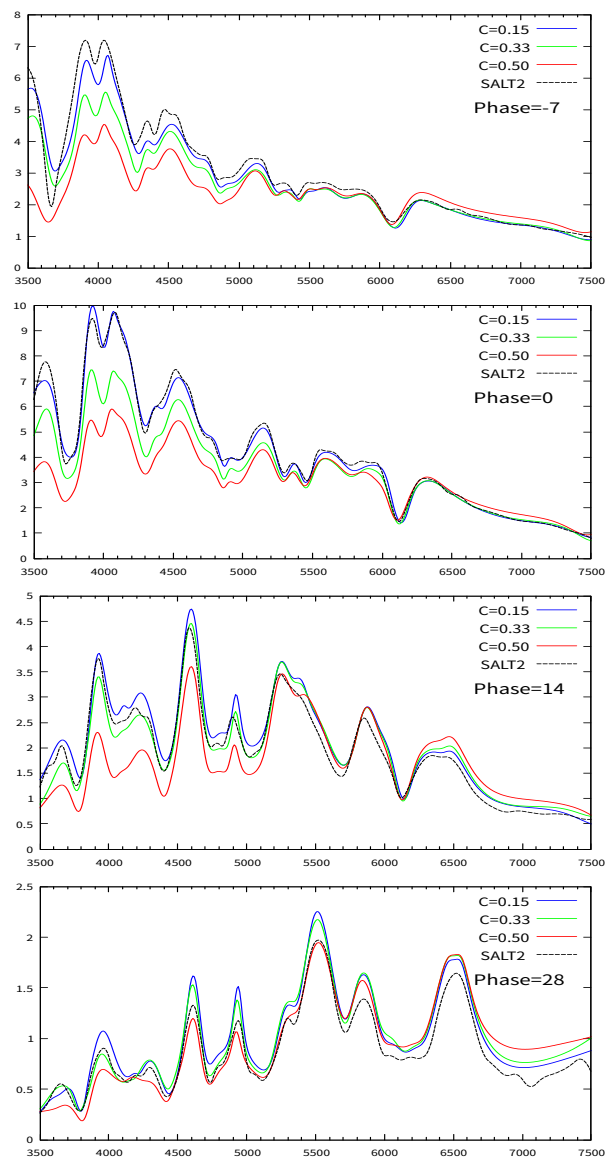


FIG. 10: Comparison of different color spectra. The black dash line is SALT2 template spectra.



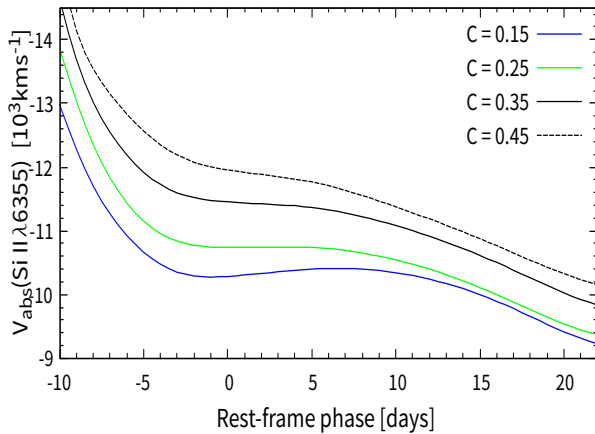


FIG. 11: Evolution of the Si II  $\lambda 6355$  absorption velocity with time for the spectra of different color.

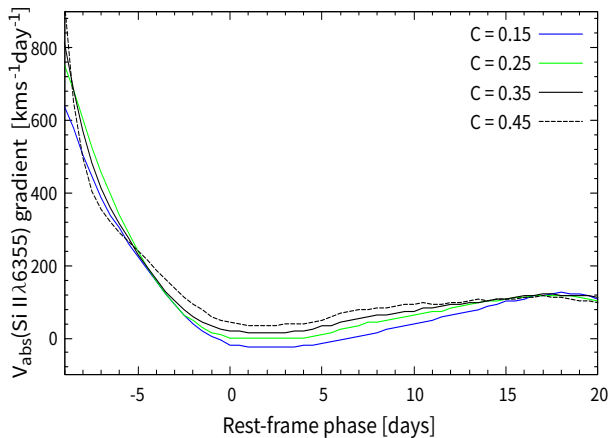


FIG. 12:  $\partial V_{abs}/\partial t$  as a function of phase.

gram. The model can well describe not only the spectrum of SNe Ia with wavelength range from  $3500\text{\AA}$  to  $8000\text{\AA}$ , but also the light-curve evolution with phase time from  $-15$  to  $50$  for different colors. We also compare the results with the SALT2 template, and find that the model we constructed by using ANN is almost the same as the SALT2 template but with much less parameters during

the training process.

From Fig.5, one can see that the RMSE and MAE appear to be still decreasing with increasingly network complexity. Actually, we have tried a few more complicated network with more layers and hidden units, but the RMSE and MAE do not decrease very much, something like  $\lesssim 0.1\%$ . However, the computing time increases very much. It costs even a few months to train a network with the most complicated network during our research. Besides, the number of parameters will increase when the network become complicated and it may lead to a over-fit issue.

By taking the advantage of network method, we calculated Si II  $\lambda 6355$  absorption velocity and its gradient in different phase and color, and our results are consistent with the previous works in the literatures.

It also shows that the Levenberg-Marquardt algorithm for training the network is much faster than others such as the RPROP algorithm.

With the artificial neural network, the evolution of the universe could be obtained with the trained spectra model, which we will tackle in our future work. Actually, the history of the universe history could be observed without assumption of any cosmological model, for example, the model given in Ref.[25, 26]. Such a model-independent attempt has been put forward in Ref.[27]. We will try to achieve this goal using the network method directly.

## Acknowledgments

This work is supported by National Science Foundation of China grant Nos. 11105091, 10671128 and 11047138, the Key Project of Chinese Ministry of Education grant No. 211059, “Chen Guang” project supported by Shanghai Municipal Education Commission and Shanghai Education Development Foundation Grant No. 12CG51, and Shanghai Natural Science Foundation, China grant No. 10ZR1422000. This research has made use of the CfA Supernova Archive, which is funded in part by the National Science Foundation through grant AST 0907903.

---

[1] A. G. Riess *et al.* [Supernova Search Team], *Astron. J.* **116**, 1009 (1998) doi:10.1086/300499 [astro-ph/9805201].  
[2] S. Perlmutter *et al.* [Supernova Cosmology Project Collaboration], *Astrophys. J.* **517**, 565 (1999) doi:10.1086/307221 [astro-ph/9812133].  
[3] A. V. Bogomazov and A. I. Tutukov, 2009, *Astronomy Reports*, **53**, 214  
[4] M. M. Phillips, *Astrophys. J.* **413**, L105 (1993). doi:10.1086/186970  
[5] J. Guy *et al.* [SNLS Collaboration], *Astron. Astrophys.* **466**, 11 (2007) doi:10.1051/0004-6361:20066930 [astro-

ph/0701828 [ASTRO-PH]].  
[6] W.S. McCulloch and W. Pitts. *Bulletin of Mathematical Biophysics*, 5(4): 115133, December 1943.  
[7] M. Minsky and S. Papert. *Perceptrons*. MIT Press, Cambridge, MA, 1969.  
[8] X. Wang *et al.*, *Astrophys. J.* **699**, L139 (2009) doi:10.1088/0004-637X/699/2/L139 [arXiv:0906.1616 [astro-ph.CO]].  
[9] S. Blondin *et al.*, *Astron. J.* **143**, 126 (2012) doi:10.1088/0004-6256/143/5/126 [arXiv:1203.4832 [astro-ph.SR]].



- [10] S. Benetti *et al.*, *Astrophys. J.* **623**, 1011 (2005) doi:10.1086/428608 [astro-ph/0411059].
- [11] D. Branch *et al.*, *Publ. Astron. Soc. Pac.* **118**, 560 (2006) doi:10.1086/502778 [astro-ph/0601048].
- [12] S. Blondin *et al.*, *Astron. J.* **131**, 1648 (2006) doi:10.1086/498724 [astro-ph/0510089].
- [13] P. Graff, F. Feroz, M. P. Hobson and A. N. Lasenby, *Mon. Not. Roy. Astron. Soc.* **441**, no. 2, 1741 (2014) doi:10.1093/mnras/stu642 [arXiv:1309.0790 [astro-ph.IM]].
- [14] M. Riedmiller and H. Braun, *Proceedings of the IEEE International Conference on Neural Networks*, 586-591, IEEE Press, 1993
- [15] T. Matheson *et al.*, *Astron. J.* **135**, 1598 (2008) doi:10.1088/0004-6256/135/4/1598 [arXiv:0803.1705 [astro-ph]].
- [16] D. Branch *et al.*, *Astron. J.* **126**, 1489 (2003) doi:10.1086/377016 [astro-ph/0305321].
- [17] S. Jha *et al.*, *Astrophys. J. Suppl.* **125**, 73 (1999) doi:10.1086/313275 [astro-ph/9906220].
- [18] K. Krisciunas *et al.*, *Astron. J.* **142**, 74 (2011) doi:10.1088/0004-6256/142/3/74 [arXiv:1106.3968 [astro-ph.CO]].
- [19] W. Li *et al.*, *Publ. Astron. Soc. Pac.* **115**, 453 (2003) doi:10.1086/374200 [astro-ph/0301428].
- [20] R. J. Foley, G. Narayan, P. J. Challis, A. V. Filippenko, R. P. Kirshner, J. M. Silverman and T. N. Steele, *Astrophys. J.* **708**, 1748 (2010) doi:10.1088/0004-637X/708/2/1748 [arXiv:0912.0263 [astro-ph.CO]].
- [21] M. Hicken, P. M. Garnavich, J. L. Prieto, S. Blondin, D. L. DePoy, R. P. Kirshner and J. Parrent, *Astrophys. J.* **669**, L17 (2007) doi:10.1086/523301 [arXiv:0709.1501 [astro-ph]].
- [22] A. Kim, A. Goobar and S. Perlmutter, *Publ. Astron. Soc. Pac.* **108**, 190 (1996) doi:10.1086/133709 [astro-ph/9505024].
- [23] P. Nugent, A. Kim and S. Perlmutter, *Publ. Astron. Soc. Pac.* **114**, 803 (2002) doi:10.1086/341707 [astro-ph/0205351].
- [24] R. J. Foley, N. E. Sanders and R. P. Kirshner, *Astrophys. J.* **742**, 89 (2011) doi:10.1088/0004-637X/742/2/89 [arXiv:1107.3555 [astro-ph.CO]].
- [25] C. J. Feng, X. Y. Shen, P. Li and X. Z. Li, *JCAP* **1209**, 023 (2012) doi:10.1088/1475-7516/2012/09/023 [arXiv:1206.0063 [astro-ph.CO]].
- [26] C. J. Feng, *Phys. Lett. B* **672**, 94 (2009) doi:10.1016/j.physletb.2009.01.022 [arXiv:0810.2594 [hep-th]].
- [27] C. J. Feng and X. Z. Li, *Astrophys. J.* **821**, 30 (2016) doi:10.3847/0004-637X/821/1/30 [arXiv:1604.01930 [astro-ph.CO]].