# Binary Stellar Population Spectral Fitting Based on Strategy-Improved Genetic Algorithm: A Postprint

**Authors:** Chi Huanbin, Li Zhongmu, Wang Feng

**Date:** 2020-07-29T16:00:58+00:00

## Abstract

The essence of binary star population spectral fitting lies in searching the parameter space corresponding to the binary star population theoretical spectral library, and identifying the theoretical spectrum that best matches the observed spectrum via least squares minimization. Rapid and accurate fitting represents both the challenge and the crucial element in effectively employing population synthesis methods to process massive galaxy spectral datasets. To enhance the speed of binary star population spectral fitting and address the issues of substantial computational cost and low fitting efficiency, this work employs a strategically improved genetic algorithm to solve the optimization model for binary star population spectral fitting, and compares its performance against the BS2fit algorithm and conventional genetic algorithms. Experimental results demonstrate that the strategically improved genetic algorithm enhances the speed of binary star population spectral fitting by an average of 43.5%, thereby advancing the application of evolutionary population synthesis methods in astronomical research to a certain extent.

## Full Text

### Binary-Star Population Spectral Fitting Based on Strategy-Improved Genetic Algorithm

**Chi Huanbin[1], Li Zhongmu[2], Wang Feng[1],[3]**
[1]School of Management and Economics, Kunming University of Science and Technology, Kunming 650093, China
[2]Institute of Astronomy, Dali University, Dali 671000, China
[3]Center for Astrophysics, Guangzhou University, Guangzhou 510006, China

**Abstract:** The essence of binary-star population spectral fitting is to search the parameter space corresponding to the binary-star population theoretical spectral library, finding the theoretical spectrum most similar to the observed spectrum through least-squares methods. Fast and accurate fitting is the key and difficult point for effectively utilizing population synthesis methods to process massive galaxy spectra. To improve the speed of binary-star population spectral fitting, and to address the problems of enormous computational requirements and low fitting efficiency, this paper builds upon the optimization model for binary-star population spectral fitting and employs a strategy-improved genetic algorithm to solve the model, comparing it with the BS2fit algorithm and conventional genetic algorithm. Experimental results demonstrate that the strategy-improved genetic algorithm can increase the speed of binary-star population spectral fitting by an average of 43.5%, promoting to some extent the application of evolutionary population synthesis methods in astronomical research.

**Keywords:** spectral fitting; galaxies; genetic algorithm
**Classification Codes:** P1-092; TP391
**Document Code:** A

## 0 Introduction

Binary-star populations are stellar assemblies containing co-evolving binary stars and single stars with identical initial metallicity and evolutionary age, representing an extension and enrichment of single-star populations. Parameters derived from binary-star population models for star clusters and galaxies are typically more reasonable [1], consequently attracting increasing attention to binary-star population research.

Spectral fitting is an important method for obtaining physical information about galaxies [2]. With the development of various sky survey projects and astronomical information technology, increasingly high-quality spectral data have been obtained, creating conditions for stellar spectral fitting and effectively promoting its advancement. Spectral fitting has achieved a series of results in recent years. For instance, reference [2] successfully used Bayesian methods to estimate galaxy parameter information from single-star population spectral fitting. Reference [3] analyzed binary-star and single-star population models based on Bayesian evidence using Bayesian model selection methods, finding that binary-star population models achieve much better spectral fitting results than models that do not consider binary interactions. References [4] and [5] found through spectral fitting of early-type galaxies from ultraviolet to visible bands that binary-star population models provide good explanations for blue straggler observations in globular clusters. Reference [6] (Han, Z. 2007) first discovered through binary-star population spectral fitting that it can well explain the UV-upturn phenomenon in early-type galaxies.

Since binary-star population evolution is more complex than traditional single-star populations and considers more physical inputs, its corresponding theoretical spectral library is large, and the spectral sampling rate required for certain fitting precision is also higher. Consequently, binary-star population spectral fitting requires massive computations. The ability to perform binary-star population spectral fitting quickly and efficiently directly affects the application of evolutionary population synthesis methods. Therefore, exploring fast and effective spectral fitting methods based on binary-star population models is extremely important.

To address the problem that binary-star population spectral fitting speed cannot meet the needs of galaxy evolution research and to accelerate the promotion and application of evolutionary population synthesis methods in astronomical research, this paper makes an attempt based on management strategies. We transform spectral fitting into an optimization problem, establishing an optimization model with the objective of minimizing the mean square error between theoretical and observed spectra, and propose a strategy-improved genetic algorithm to solve the model, comparing it with the BS2fit algorithm [1] and conventional genetic algorithm to verify the algorithm's effectiveness.

*Funding: National Key R&D Program (2018YFA0404603), NSFC-CAS Joint Fund for Astronomy (U1831204, U1931141), National Natural Science Foundation of China (1196114001, 11903009, 11863002, GZ1284), Yunnan Provincial Key R&D Program (2018IA054), Yunnan Wang Jingxiu Academician Workstation*

*Received: 2020-06-15; Revised: 2020-07-13*

*Author Introduction: Chi Huanbin, male, Ph.D. candidate, research interests: data processing and intelligent computing. Email: 563372780@qq.com*

*Corresponding Author: Wang Feng, male, professor. Research interests: massive astronomical data processing. Email: fengwang@gzhu.edu.cn*

## 1 Introduction to Binary-Star Population Models

A primary objective of binary-star population models is to construct high-precision and complete binary-star population theoretical spectral libraries. The construction of binary-star population theoretical spectra includes two components: assumptions about model input parameters and calculation of theoretical spectra.

As described in reference [1], the binary-star population model employs Hurley's [7] rapid stellar evolution code instead of stellar evolutionary tracks, uses the BaSeL3.1 stellar spectral library [8] and the Chabrier initial mass function [9], and assumes that the mass ratio q of primary and secondary stars in each binary system follows a uniform distribution from 0 to 1 [9]. The star formation rate is assumed to take the form ( ) ( ) - [8], where and represent the galaxy's age and the duration timescale of star formation, respectively.

Using the evolutionary population synthesis method, the number of stars formed at different times in a galaxy can be calculated given the initial mass function and star formation rate. Subsequently, Hurley's rapid stellar evolution code is employed for stellar evolution calculations. Finally, the BaSeL3.1 stellar sample library converts stellar evolutionary parameters into spectral energy distributions. After completing stellar evolution calculations, the spectra of individual included stars are synthesized to calculate the spectral flux of a population of age at wavelength .

Each population's spectral flux is determined by three parameters: metallicity Z, old population age , and young population age . Considering the effects of stellar velocity dispersion , dust attenuation (using the attenuation theory proposed by Calzetti et al. in 2000 [8]), redshift, and Milky Way extinction, the observed spectral flux at wavelength in binary-star population synthesis can be expressed as:

[ ( ) ( )] ( ) ( ) ( )

where is the population's spectral flux determined by the three parameters (Z, , ), G( ) is a Gaussian distribution with mean and standard deviation of and , p( ) is the percentage of energy passing through dust with attenuation , r( ) is the energy ratio after Milky Way extinction , and is the convolution operator.

To simplify the binary-star population theoretical spectral library, this paper sets the values of and to 0, considering only Z, , and . However, this does not affect the experimental purpose, as it similarly allows comparative analysis of the performance of the proposed improved genetic algorithm in spectral fitting, since our comparative experiments are also based on the same simplified binary-star population theoretical spectral library.

## 2.1 Binary-Star Population Spectral Fitting Model

The purpose of spectral fitting is to obtain physical parameter information about galaxy star clusters. As described in the previous section, in binary-star population models, spectra are jointly determined by eight parameters: , , , and stellar population mass. Among these, a galaxy's stellar mass can be calculated using the fitting formula from Thomas et al. (2005) [10]:

where is the galaxy's stellar mass and is the stellar velocity dispersion. Other parameters can be obtained through spectral fitting, which provides quantitative information about population composition by comparing theoretical spectra with observed spectra. The specific task is to find the theoretical spectrum from the theoretical spectral library that most closely matches the observed spectrum. In spectral fitting, by comparing observed spectra with theoretical spectra and minimizing the difference between theoretical and observed spectral data, the optimal theoretical spectrum is determined, thereby inferring the main parameters of stellar composition.

The binary-star population theoretical spectral library has a dense parameter grid: stellar metallicity (Z) ranges from 0.0003 to 0.03, with parameter intervals of 0.0001 when Z $<$ 0.001, and intervals of 0.001 for higher metallicities (Z $>$ 0.001). The old population age ( ) of star clusters ranges from 0 to 15 Gyr with intervals of 0.1 Gyr; the young population age ( ) decreases from to 0 Gyr with intervals of 0.1 Gyr. The stellar velocity dispersion $\sigma*$ parameter changes from 0 to 350 km $s^{-1}$ in steps of 10 km $s^{-1}$; dust attenuation ( ) ranges from 0 to 0.5 with intervals of 0.01. Consequently, the binary-star population theoretical spectral library encompasses a huge parameter space, and spectral fitting inherently involves massive computations, for which we establish the following fitting model.

Let the observed spectrum be and the binary-star population synthetic theoretical spectrum be . The spectral fitting goodness can be characterized by:

[( ) ] (3)

where is the weight parameter for wavelength $\lambda$. The best fitting result corresponds to min( ). In most cases, uses observational errors but can also be assigned a specific threshold according to special requirements. The essence of population spectral fitting is to search the parameter space corresponding to the theoretical spectral library to find the theoretical spectrum that minimizes for the observed spectrum. When is minimized, the binary-star population synthetic spectral fitting model satisfies the physical constraint: 0< <1, and its essence is an optimization problem that can be formalized as:

where equation (3) is the objective function.

## 2.2 Binary-Star Population Spectral Fitting Algorithm

The Genetic Algorithm [11] (GA) is a bionic heuristic algorithm that simulates natural biological genetic evolution behavior, proposing selection, crossover, and mutation operators. Through multi-point parallel iterative search of the solution space, it can quickly and effectively converge to the global optimal solution. GA is a population-based search algorithm where individuals represent samples of possible solutions. As the search iterates, individuals produce better offspring by sharing and exchanging information among themselves. The conventional genetic algorithm (CGA) contains crucial operations including selection, crossover, and mutation, which can select winning individuals, maintain population diversity, and exchange information among individuals within the population.

However, traditional improved strategy genetic algorithms exhibit poor local search capability and are very sensitive to initial values. After analyzing the operational mechanism of the GA algorithm and inspired by the inertia adaptation of particle swarm optimization, and to overcome this deficiency, this paper designs a self-adaptive genetic algorithm (SGA) based on conventional genetic algorithm improvement strategies to solve the binary-star population spectral fitting problem. The conventional genetic algorithm (CGA) uses fixed crossover probability (PC, Probability of Cross) and fixed mutation probability (PM, Probability of Mutation). This causes the algorithm to proceed according to predetermined rules, resulting in insufficient genetic variation flexibility. The performance of the algorithm, such as convergence speed and global optimization capability, is sensitive to the selection of evolutionary parameters. If an individual's fitness value is greater than the population's average fitness value, the assigned crossover and mutation probabilities should be smaller. Furthermore, for individuals with larger fitness values, indicating they are closer to the target solution, smaller crossover and mutation probabilities should be assigned to protect their information from being passed to the next generation. Conversely, for individuals with smaller fitness values, disturbance factors should be opportunistically introduced to enhance their mutation capability, and individuals should be assigned larger crossover and mutation probabilities to increase their probability of mutation. Thus, individuals with smaller fitness values have greater opportunities for mutation. To reduce dependence on evolutionary parameter selection and improve algorithm convergence speed and global optimization capability, the SGA improvements are as follows:

1) Adaptive PC update:
2) Adaptive PM update:

where iteration is the current iteration number, is the population's average fitness value, is the fitness value of the individual selected for crossover operation, is the fitness value of the individual selected for mutation operation, fmax is the maximum fitness value, is a random parameter, and based on experience is set to

3) The main purpose of the disturbance factor is to enhance the algorithm's exploration capability of the search space. To this end, the following disturbance operator (DO, Disruption Operator) is introduced:

where: is a random number between (-2,2), s is the Hamming distance between the i-th chromosome and the j-th chromosome in the neighborhood, and s is the Hamming distance between the current optimal chromosome and the i-th chromosome.

The specific flow of the SGA algorithm is shown in Figure 1 [Figure 1: see original paper] and mainly includes the following steps:

Step 1: Spectral data preprocessing, including emission line removal, normalization, and redshift processing;

Step 2: Problem encoding and population initialization;

Step 3: Calculate fitness values according to the objective function and retain the current optimal population. Determine population size including chromosome length and number, initialize population according to strategy, and set evolutionary termination conditions;

Step 4: Determine whether convergence or stopping conditions are met; if satisfied, output search results, otherwise proceed to iteration;

Step 5: Selection operator: randomly select a group of better individuals based on fitness values;

Step 6: Update crossover probability according to equation (5);

Step 7: Crossover operator: randomly select two chromosomes from the set to exchange one or more bits, produce two new individuals according to crossover probability, and add disturbance according to equation (7);

Step 8: Update mutation operator according to equation (6): randomly select a chromosome from the set and change one or more bits according to mutation probability to produce a new chromosome;

Step 9: Return to Step 4 for termination judgment, loop through the above steps until termination conditions are met.

## 3.1 Experiments with the Strategy-Improved Genetic Algorithm

To verify the optimization efficiency of SGA in binary-star population spectral fitting, this paper selects the CGA and the BS2fit algorithm from reference [1] for comparison. The experimental code was developed using Python 3.6, with Ubuntu 18.04 64-bit as the operating system, 8 GB memory capacity, and an i7-4790 CPU with a clock speed of 3.60 GHz. The experimental theoretical spectral data uses the binary-star population theoretical spectral library based on the Chabrier initial function for fitting[1]. The observed spectral data were selected from ultraviolet-visible band spectral data of 99 local galaxies, which can be downloaded from the internet[2].

These data are characterized by covering a wide wavelength range from ultraviolet to visible light. Moreover, since binary stars have obvious effects on ultraviolet spectra, these data are also ideal for exploring the impact of binary-star evolution on spectral fitting. Each observed spectrum has = 3871 sampling points, with 5500 Å used as the normalization factor for spectral preprocessing. The SGA algorithm parameters are set as: crossover probability upper limit C2 = 1, mutation probability upper limit M2 = 1, random parameter .5, individual encoding length n = 20, population size Pop = 100. The time consumption indicator takes the average of 20 repetitions for each algorithm.

The experiment first randomly selects one spectrum from the observed spectral data (NGC 1399 galaxy) as experimental data, referencing literature [12] for data preprocessing. The observed spectrum is then masked for bad points, emission lines, sky lines, etc., to obtain spectral data numbered spec-0266-51630-0010. Subsequently, drawing on related research [13], / is adopted as the evaluation criterion, where when / < 1, the theoretical spectrum is considered to fit the observed spectrum well, the population synthesis parameters are accurately estimated, and the fitting result is acceptable. The fitting iteration process is shown in Figure 2 [Figure 2: see original paper] (left). The algorithm converges after 20 population iterations, with a corresponding fitness value of 1308 and fitting accuracy of = 0.3433. The theoretical spectrum fits the observed spectrum well, as shown in the right panel of Figure 2, where the gray curve represents the actual observed spectrum and the red curve represents the optimal fitted theoretical spectrum.

[1] http://cluster.shao.ac.cn/~zhongmu/ssp_{{{chab}}{{hrseds}}}/
[2] *ftp://ftp.stsci.edu/pub/catalogs/nearby*{gal}/sed.html

## 3.2 Comparison with Conventional Genetic Algorithm (CGA)

CGA algorithm parameters: crossover probability PC = 0.8, mutation probability PM = 0.1, chromosome length of 20, Pop = 100. The fitting results using the CGA algorithm on the binary-star population theoretical spectral library for the test spectrum are shown in Figure 3 [Figure 3: see original paper]. In the CGA algorithm, the population falls into a local optimum after 16 iterations, begins to converge after 15 iterations (iteration), and finds the optimal theoretical spectrum.

The experimental iteration 20 times with an average time consumption of 3201.3 s, which is slightly better than the BS2fit algorithm, with a speed improvement of 19.2%. Compared with the CGA algorithm, SGA converges after 20 iterations, can effectively jump out of local optimum traps, and shows significant convergence speed improvement, with an average time consumption of 1304.4 s for 20 experiments, representing a speed improvement of 19.1%.

## 3.3 Comparison with BS2fit Algorithm

The BS2fit algorithm was implemented in Python as described in reference [1]. The fitting traversal process is shown in Figure 4 [Figure 4: see original paper], where the horizontal axis represents the number of traversal calculations and the vertical axis represents the chi-square statistic value obtained from each calculation. Due to full grid search, the binary-star population model's theoretical library covers long wavelength bands with dense evolutionary parameter grids (the model spectral library covers metallicity Z from 0.0003 to 0.03 with precision of 0.0001; t1, t2 range from 0 to 15 Gyr with precision of 0.1 Gyr), resulting in a huge sample parameter space.

Although we simplified the model parameters, fitting one observed spectrum still requires massive one-by-one comparisons, which can find the optimal theoretical spectrum. The average time consumption for 20 experiments is 3954.2 s. Compared with the BS2fit algorithm, the SGA algorithm has stronger global search capability, higher theoretical spectral search efficiency, and significantly improved convergence speed, with average time consumption reduced by 67.2% over 20 runs.

As described above, all three algorithms can find the theoretical spectrum corresponding to the minimum . According to the theoretical spectrum's parameter grid settings, the fitting parameters given in Table 1 are retrieved.

**Table 1: bsSP-fitted Parameters of NGC 1399 Galaxy**

| Parameter | This Paper's Fitted Value | Reference [4] bsSP-fitted Value |
|---|---|---|
| Z (metallicity) | $0.030 \pm 0.00$ | (0.035, discussed in text) |
| Old population age | $13.50 \pm 0.00$ | (~27 Gyr, discussed in text) |
| Young population age | $12.27 \pm 0.81$ | (~24.5 Gyr, discussed in text) |

The reasons for errors are that we simplified the experimental model. First, the theoretical spectral library only considered (Z, , ) without considering the effects of parameters like and . Second, we did not consider observational spectral errors or process the uncertainties in observed spectra, i.e., the ideal case with large signal-to-noise ratio (S/N) values. The metallicity obtained in this paper is 0.005 smaller than the fitted value in reference [4], which aligns with our expectations because model input uncertainties typically lead to underestimated metallicity values [4]. The main reason why the fitted values of population ages ( , ) are 50% lower than those in reference [4] is that the simplified model did not consider velocity dispersion, leading to underestimated population ages. This fitting result also demonstrates the "age-metallicity degeneracy effect."

In summary, the improved genetic algorithm with adaptive strategies increases fitting speed by an average of 43.5% compared with conventional genetic algorithm and BS2fit algorithm, with results meeting expectations.

## 4 Summary and Discussion

Using a self-adaptive strategy-improved genetic algorithm for binary-star population spectral fitting, the results indicate: (1) It is feasible to use swarm intelligence algorithms (SIA) to solve binary-star spectral fitting problems. (2) Compared with conventional genetic algorithm and BS2fit algorithm, the improved genetic algorithm can increase computational speed by 19.1% and 67.2% respectively in binary-star population spectral fitting, with an average speed improvement of 43.5% for binary-star population spectral fitting, demonstrating significant improvement. The improved genetic algorithm builds upon the conventional genetic algorithm, leveraging its strong global search capability and achieving faster convergence compared with conventional genetic algorithm. Moreover, the improved genetic algorithm is more intelligent in searching, can effectively jump out of local optimum traps, improve the global search capability of genetic algorithm, and avoid premature convergence. Compared with the full grid search of the BS2fit algorithm, the SGA algorithm makes tradeoffs in the search space and is therefore more efficient. Although the improved genetic algorithm improves speed in binary-star population fitting, there remains considerable room for improvement. On one hand, encoding methods also affect algorithm exploration capability and convergence, and improved algorithm encoding such as quantum encoding is worth exploring. On the other hand, swarm intelligence algorithms are developing rapidly, and exploring more effective search strategies, such as integration with reinforcement learning Q-learning, is also necessary. Considering that genetic algorithms are easy to implement in parallel, the next step could explore the improvement of spectral fitting speed using improved genetic algorithms under parallel strategies.

## References

[1] Li Z, Zhang L, Liu J, et al. Integrated spectral energy distributions of binary star composite stellar populations[J]. Monthly Notices of the Royal Astronomical Society, 2012, 424(2): 874-883.

[2] Han Y, Han Z. BayeSED: A GENERAL APPROACH TO FITTING THE SPECTRAL ENERGY DISTRIBUTION OF GALAXIES[J]. Astrophysical Journal Supplement Series, 2014, 215(1).

[3] Han Y , Han Z . A Comprehensive Bayesian Discrimination of the Simple Stellar Population Model, Star Formation History, and Dust Attenuation Law in the Spectral Energy Distribution Modeling of Galaxies[J]. The Astrophysical Journal Supplement Series, 2018, 240(1).

[4] Li, Zhongmu , et al. Potential Importance of Binary Evolution in UV-Optical Spectral Fitting of Early-Type Galaxies." Astrophysical Journal 776.1(2013):37.

[5] Zhang F , Li L , Han Z , et al. Yunnan-III models for Evolutionary population synthesis[J]. Monthly Notices of the Royal Astronomical Society(4):3390-3408.

[6] Han Z, Podsiadlowski P, Lynasgray A E, et al. A binary model for the UV-upturn of elliptical galaxies[J]. Monthly Notices of the Royal Astronomical Society, 2007, 380(3): 1098-1118.

[7] Hurley J R, Tout C A, Pols O R, et al. Evolution of binary stars and the effect of tides on binary populations[J]. Monthly Notices of the Royal Astronomical Society, 2002, 329(4): 897-928.

[8] Calzetti D , Armus L , Bohlin R C , et al. The Dust Content and Opacity of Actively Star-Forming Galaxies[J]. Astrophysical Journal, 1999, 533(2):682.

[9] Chabrier G. The Galactic Disk Mass Function: Reconciliation of the Hubble Space Telescope and Nearby Determinations[J]. The Astrophysical Journal, 2003, 586(2).

[10] Thomas D, Maraston C, Bender R, et al. The Epochs of Early-Type Galaxy Formation as a Function of Environment[J]. The Astrophysical Journal, 2005, 621(2): 673-694.

[11] David E. Goldberg. Genetic Algorithm in Search, Optimization and Machine Learning[J]. addison wesley, 1989, xiii(7):2104–2116.

[12] 张茜，张健楠，赵永恒．基于聚类的星系光谱分析．天文研究与技术，2020，17(2): 233-243.

Zhang Xi, Zhang Jiannan, Zhao Yongheng. Spectral Classification of Galaxies Based on Clustering Analysis. Astronomical Research and Technology, 2020, 17(2): 233-243.

[13] 韩金姝. 基于遗传算法的星族合成参数估计 [J]. 天文学报,2015,56(02):93-101.

Han Jinshu. Parameter Estimation of Stellar Population Synthesis Using Genetic Algorithm [J]. Astronomy Letters, 2015,56(02):93-101.