BayeSED: A GENERAL APPROACH TO FITTING THE SPECTRAL ENERGY DISTRIBUTION OF GALAXIES

YUNKUN HAN\textsuperscript{1,2} AND ZHANWEN HAN\textsuperscript{1,2}

\textsuperscript{1} Yunnan Observatories, Chinese Academy of Sciences, Kunming, 650011, China; hanyk@ynao.ac.cn, zhanwenhan@ynao.ac.cn
\textsuperscript{2} Key Laboratory for the Structure and Evolution of Celestial Objects, Chinese Academy of Sciences, Kunming, 650011, China

Received 2014 June 9; accepted 2014 August 25; published 2014 October 9

ABSTRACT

We present a newly developed version of BayeSED, a general Bayesian approach to the spectral energy distribution (SED) fitting of galaxies. The new BayeSED code has been systematically tested on a mock sample of galaxies. The comparison between the estimated and input values of the parameters shows that BayeSED can recover the physical parameters of galaxies reasonably well. We then applied BayeSED to interpret the SEDs of a large \(K_s\)-selected sample of galaxies in the COSMOS/UltraVISTA field with stellar population synthesis models. Using the new BayeSED code, a Bayesian model comparison of stellar population synthesis models has been performed for the first time. We found that the 2003 model by Bruzual & Charlot, statistically speaking, has greater Bayesian evidence than the 2005 model by Maraston for the \(K_s\)-selected sample. In addition, while setting the stellar metallicity as a free parameter obviously increases the Bayesian evidence of both models, varying the initial mass function has a notable effect only on the Maraston model. Meanwhile, the physical parameters estimated with BayeSED are found to be generally consistent with those obtained using the popular grid-based FAST code, while the former parameters exhibit more natural distributions. Based on the estimated physical parameters of the galaxies in the sample, we qualitatively classified the galaxies in the sample into five populations that may represent galaxies at different evolutionary stages or in different environments. We conclude that BayeSED could be a reliable and powerful tool for investigating the formation and evolution of galaxies from the rich multi-wavelength observations currently available. A binary version of the BayeSED code parallelized with Message Passing Interface is publicly available at https://bitbucket.org/hanyk/bayesed.

Key words: galaxies: evolution -- galaxies: fundamental parameters -- galaxies: statistics -- galaxies: stellar content -- methods: data analysis -- methods: statistical

Online-only material: color figures

1. INTRODUCTION

In the formation and evolution of galaxies, the physical processes involved in the formation and evolution of stars, the interstellar medium (ISM), and supermassive black holes are expected to be tightly interconnected. Our theoretical understanding of these processes and their interactions (Cole et al. 2000; Bower et al. 2006; Croton et al. 2006; Baugh 2006; Marulli et al. 2008; Somerville et al. 2008; Bonoli et al. 2009) (Di Matteo et al. 2005; Springel et al. 2005a, 2005b; Hopkins et al. 2006, 2008a, 2008b; Springel 2010) has been greatly advanced in the last few years using semi-analytical modeling (Cole et al. 2000; Bower et al. 2006; Croton et al. 2006; Baugh 2006; Marulli et al. 2008; Somerville et al. 2008; Bonoli et al. 2009) and numerical simulations (Di Matteo et al. 2005; Springel et al. 2005a, 2005b; Hopkins et al. 2006, 2008a, 2008b; Springel 2010). However, due to the high level of complexity present in these physical processes, our understanding of them and their mutual interactions is still far from complete.

Empirical clues from multi-wavelength observations of galaxies (Bell et al. 2003; Shen et al. 2003; Kauffmann et al. 2004; Baldry et al. 2004; Tremonti et al. 2004; Gallazzi et al. 2005) are very helpful in providing crucial insights for our further understanding of the formation and evolution of galaxies. Also, our theoretical understanding of the formation and evolution of galaxies needs to be tested, constrained, and someday confirmed by many observational results. Fortunately, all of these complex physical processes involved in the formation and evolution of galaxies have left some imprints in the integrated multi-wavelength spectral energy distributions (SEDs) of galaxies. Therefore, the multi-wavelength SEDs of galaxies are a very important source of information for our understanding of these complex physical processes. The advent of new observing facilities and large surveys at \(\gamma\)-ray to radio wavelengths (Lonsdale et al. 2003; Jansen et al. 2001; Martin et al. 2005; Giavalisco et al. 2004; Atwood et al. 2009; Abazajian et al. 2009; Scoville et al. 2007; Davis et al. 2007; Driver et al. 2009; Condon et al. 1998) now allow us to obtain the full SEDs of galaxies.

To extract physical information about galaxies from their observed multi-wavelength SEDs, we need some kind of theoretical model for the SEDs of galaxies. The SEDs of most galaxies can be thought of as the superposition of the SEDs of a population of stars with different masses, ages, metallicities, etc. that constitute the galaxy. Due to this, stellar population synthesis has been the main method of modeling the SEDs of galaxies, beginning with the pioneering works of Tinsley (1972), Searle et al. (1973), and Larson & Tinsley (1978). From then on, numerous efforts by different groups have been made to improve this technique (Bruzual 1983; Buzzoni 1989; Bruzual & Charlot 1993, 2003; Bruzual 2007; Leitherer & Heckman 1995; Fioc & Rocca-Volmerange 1997; Maraston 1998, 2005; Zhang et al. 2005; Li & Han 2008; Conroy et al. 2009). However, some important issues still remain. For example, some short lived but bright phases of stellar evolution, such as the thermally pulsing asymptotic giant branch, horizontal branch (Catelan 2009; Lei et al. 2013), and blue stragglers (Tian et al. 2006; Han et al. 2007; Chen & Han 2009), are still not well understood and they potentially have important effects on the resulting SEDs of galaxies. Furthermore, there are issues concerning the universality of the stellar initial mass function ( IMF; Padoan et al. 1997; Myers et al. 2011; Dutton et al. 2013), different parameterizations of the star formation history (SFH; Maraston et al. 2003; Jansen et al. 2001; Martin et al. 2005; Giavalisco et al. 2004; Atwood et al. 2009; Abazajian et al. 2009; Scoville et al. 2007; Davis et al. 2007; Driver et al. 2009; Condon et al. 1998) now allow us to obtain the full SEDs of galaxies.
The main method of extracting physical information from the multi-wavelength SEDs of galaxies is SED fitting. Using SED fitting, we try to derive one or several physical properties of galaxies by using certain fitting methods to compare SED models with observed SEDs. In other words, we need to solve the inverse problem: How can the physical properties of galaxies (e.g., stellar ages, stellar masses, SFHs, dust extinction) be reasonably derived from quantities that are directly observable (e.g., multi-wavelength photometric SEDs)? In the last decade, the technique of SED fitting has been significantly improved (see, e.g., Walcher et al. 2011, for a recent review). Numerous SED fitting methods and their corresponding software have been presented by many authors (Cid Fernandes et al. 2005; Ocvirk et al. 2006; Tojeiro et al. 2007; Koleva et al. 2008, 2009), and have been used by even more authors to derive valuable physical information about the formation and evolution of galaxies.

The modeling of a galaxy’s SED involves the convolution of the star formation and evolution history, the stellar IMF, and the formation and evolution of its dusty ISM, which has nonlinear effects on the resulting SEDs of galaxies, and many other properties that characterize the formation and evolution of the galaxy (Mo et al. 2010). Furthermore, apart from uncertainties in the observations, there are many uncertainties in the modeling of galaxy SEDs. Given these complexities and uncertainties, the problem of deriving the physical properties of galaxies from their directly observable properties is generally not invertible in a strict mathematical sense. However, from the perspective of statistical inversion theory (Kaipio & Somersalo 2004; Tarantola 2005), the inverse problems can be solved by means of Bayesian statistical inference. This is very different from the commonly used approach of solving an optimization problem in the least \( \chi^2 \) sense (Tinsley & Gunn 1976; Bolzonella et al. 2000; Walcher et al. 2006; Maraston et al. 2006; Koleva et al. 2009; Pforr et al. 2012, 2013; Mitchell et al. 2013; Li et al. 2013) where the main purpose is finding the best fitting result. In Bayesian statistical inference, all quantities (e.g., parameters and fluxes) are modeled as random variables. Therefore, instead of finding a specific value of a parameter that best matches the observations, the solution to an inverse problem is the posterior probability distribution of the quantity of interest, which describes the degree of confidence about the quantity given the available observations. The posterior probability distributions represent our full knowledge about the parameters with the degree of uncertainty and the degeneracies between them manifesting as easily noticeable broad or multi-peaked distributions. Besides, in Bayesian inference, additional information about the problem can be incorporated as priors to constrain even further the solution. The Bayesian inference methods have been successfully used in many fields of physics (Gregory 2005), and especially cosmology, to derive the cosmology parameters of the universe (Lewis & Bridle 2002; Verde et al. 2003; Cole et al. 2005; Dunkley et al. 2009; Hinshaw et al. 2013).

In recent years, Bayesian methods have been used in the field of SED fitting of galaxies by more and more authors. Benítez (2000) systematically applied Bayesian inference to estimate the photometric redshift of galaxies. It is important to note that they used Bayesian priors as information in addition to the observed photometric SEDs to give better redshift estimates for the first time. Kauffmann et al. (2003) have used the Bayesian technique to estimate the stellar mass-to-light ratios, dust attenuation corrections, and burst mass fractions for a sample of \(10^5\) galaxies from the Sloan Digital Sky Survey. They also presented a rigorous mathematical description of the method which currently forms the basis for Bayesian SED fitting. Salim et al. (2005, 2007) have used the combination of SDSS and Galaxy Evolution Explorer (GALEX) photometry to obtain dust-corrected star formation rates (SFRs) of galaxies in the local universe. da Cunha et al. (2008) presented an empirical but physically motivated model to interpret the SEDs of galaxies from UV to far-IR consistently, as well as the corresponding MAGPHYS (Multi-wavelength Analysis of Galaxy Physical Properties) package. Similarly, Noll et al. (2009) presented the code CIGALE (Code Investigating GALaxy Emission) for a Bayesian-like analysis of galaxy SEDs from far-UV to far-IR by fitting the attenuated stellar emission and the related dust emission simultaneously.

Recently, the Markov Chain Monte Carlo (MCMC) algorithm has been employed by different authors (Acquaviva et al. 2011; Serra et al. 2011; Johnson et al. 2013) to allow a much more efficient and complete sampling of the parameter space of an SED model than in grid-based methods such as CIGALE and MAGPHYS. In Han & Han (2012), we described our BayeSED code where the multimodal nested (MultiNest) sampling algorithm has been employed, and applied it to a sample of hyperluminous infrared galaxies as a demonstration. The use of MultiNest instead of MCMC allows us to obtain not only the posterior distribution of all model parameters, but also the Bayesian evidence of the model that can be used as a generalization of Occam’s razor for quantitative model comparison. Meanwhile, the principal component analysis and artificial neural networks (ANNs) techniques have been employed to significantly speed up the generation of model SEDs, a major bottleneck for efficient sampling of the parameter space of an SED model.

Since its first description in Han & Han (2012), the BayeSED code has been improved significantly, with the improvements including but not limited to the following. First, the new MultiNest algorithm, which is improved by importance nested sampling (INS) and which allows a more efficient exploration of higher-dimensional parameter spaces and more accurate calculation of Bayesian evidence, has been employed. Second, besides the ANNs algorithm, the K-nearest neighbors (KNNs) algorithm has been added as another method for efficient interpolation of model SED libraries. Third, the redshift of a galaxy can be set as a free parameter and the effect of the intergalactic medium (IGM) and Galactic extinction have been considered. Fourth, the main body of BayeSED has been completely rewritten in C++ in an object-oriented programing fashion, and parallelized with Message Passing Interface (MPI) to be able to interpret the SEDs of multiple galaxies simultaneously. In this paper, we systematically test this new version of the BayeSED code with a mock sample of galaxies and apply it to interpret the observed SEDs of a \(K_s\)-selected sample of galaxies in the COSMOS/UltraVISTA field with evolutionary population synthesis models.

This paper is organized as follows. In Section 2, we describe our BayeSED code and recent improvements to it. We begin in
Section 2.1 by introducing the basic idea of Bayesian inference and its application to the problem of SED fitting. In Section 2.2, we introduce the implementation of Bayesian inference with the efficient and robust Bayesian inference tool—MultiNest, which is capable of calculating the Bayesian evidence of a model and exploring its parameter space, which could be very complex and of a moderately high dimension. To use sampling methods like MCMC and MultiNest, we must be able to evaluate an SED model at any point in the allowed parameter space. Therefore, in Section 2.3, we present the methods for interpolating a model SED library while using the evolutionary population synthesis model as an example. In Section 2.4, we introduce how the MultiNest algorithm and the interpolating algorithm are combined to build up our BayeSED code. To test the ability of BayeSED to recover the physical parameters of galaxies from their multi-wavelength photometry, we employ the method of using a mock sample of galaxies in Section 3. In Section 4, we systematically apply our BayeSED code to interpret the SEDs of a Ks-selected sample of galaxies in the COSMOS/UltraVISTA field given by Muzzin et al. (2013). Finally, a summary of our BayeSED method and the results obtained with its application are presented in Section 5.

2. BayeSED—BAYESIAN SPECTRAL ENERGY DISTRIBUTION FITTING OF GALAXIES

2.1. Bayesian Inference

In BayeSED, we have employed the Bayesian inference methods to interpret SEDs of galaxies. Bayesian methods have been widely used in astrophysics and cosmology (see, e.g., Trotta 2008, for a recent review). They provide a more consistent conceptual basis for dealing with problems of inference in the presence of uncertainties than traditional statistical methods. For a set of experimental or observational data \(d\), and a model (or hypothesis) \(M\) with some parameters \(\theta\) that are employed to explain them, the Bayes’ theorem states that

\[
P(\theta|d, M) = \frac{P(d|\theta, M)P(\theta|M)}{P(d|M)}. \tag{1}
\]

For SED fitting, \(d\) represents the observed SED of a galaxy while \(\theta\) represents the parameters of an SED model \(M\). In Equation (1), \(P(d|\theta, M)\) is the posterior probability of parameters \(\theta\) given the data \(d\) and model \(M\). \(P(d|\theta, M)\) is the probability of \(d\) given the model \(M\) and its parameters \(\theta\). It is also known as the likelihood \(L(\theta)\), which describes how the degree of plausibility of the parameter \(\theta\) changes when new data \(d\) is considered. \(P(\theta|M)\) is the prior, which describes knowledge about the parameters irrespective of the data. Finally, \(P(d|M)\) is a normalization constant called the marginal likelihood, also known as Bayesian evidence.

Bayesian inference is generally divided into two categories: parameter estimation and model comparison. In parameter estimation, the Bayesian evidence \(P(d|M)\) is usually ignored, as it is a normalizing factor independent of the parameters \(\theta\). The posterior includes all information that can be used for the complete Bayesian inference of the parameter values. It can be marginalized over each parameter to obtain individual parameter constraints. Therefore, the posterior probability density function (PDF) of a parameter \(\theta_i\) could be obtained as

\[
P(\theta_i|d, M) = \int d\theta_1 \cdots d\theta_{i-1} d\theta_{i+1} \cdots d\theta_N P(\theta|d, M). \tag{2}
\]

The Bayesian evidence of a model, which is not important for parameter estimation but is critical for model comparison, is given by

\[
P(d|M) = \int_{\Omega_M} P(d|\theta, M)P(\theta|M)d\theta, \tag{3}
\]

where \(\Omega_M\) represents the whole \(N\)-dimensional parameter space of the model \(M\). It is clear that the Bayesian evidence of a model is simply the average of the likelihood weighted by the priors. However, this simple definition automatically implements the principle of Occam’s razor: a simpler theory with compact parameter space should be better than a more complicated one, unless the latter is significantly better for the explanation of observational data. Generally, the Bayesian evidence is simply larger for a model with a better fit to observations, while and smaller for a more complicated model with more free parameters or larger parameter space. The comparison between two models \(M_2\) and \(M_1\) can be formally expressed as the ratio of their respective posterior probabilities given the observational data set \(d\):

\[
\frac{P(M_2|d)}{P(M_1|d)} = \frac{P(d|M_2)P(M_2)}{P(d|M_1)P(M_1)} \tag{4}
\]

where \(P(M_2|d)/P(M_1)\) is the prior odds ratio of the two models, which is often set to 1 if none of the two is of special interest. If so, then the Bayes factor, which is defined as

\[
B_{2,1} = \frac{P(d|M_2)}{P(d|M_1)}. \tag{5}
\]

can be directly used for Bayesian model comparison. According to the empirically calibrated Jeffreys’s scale (Jeffreys 1961; Trotta 2008), \(\ln(B_{2,1}) > 0, 1, 1.5, \text{ and } 5\) (corresponding to odds of about 1:1, 3:1, 12:1, and 150:1), represent inconclusive, weak, moderate, and strong evidence in favor of \(M_2\), respectively.

2.2. Sampling of High-dimensional Parameter Space with MultiNest

Commonly, for the problem of Bayesian parameter estimation, we need to solve the \(N - 1\) dimensional integration of Equation (2). However, in most cases, it is very hard to obtain an accurate analytical solution for this equation. Moreover, for many problems in astrophysics, we cannot even write down the analytical form of this equation, since the mathematical expression of the priors and/or likelihood function may simply not exist. In practice, Bayesian parameter estimation could be achieved more conveniently by taking a set of samples from the parameter space that are distributed according to the posterior \(P(\theta|d, M)\), where the posterior might be unnormalized. Then, the estimation of the parameters could be obtained by some simple statistics of these samples.

The most widely used sampling method for this is the MCMC method. The MCMC technique, which is often based on the Metropolis-Hastings algorithm (Metropolis et al. 1953; Hastings 1970), provides an efficient way to explore the parameter space of a model and ensures that the number density of samples is asymptotically proportional to the posterior probability density. However, the commonly used MCMC methods are very computationally intensive when the posterior distribution is multimodal or has large degeneracies between parameters, particularly in high dimensions. On the other hand, the calculation of Bayesian evidence, which is critical for Bayesian model
the works of Mukherjee et al. (2006) and Shaw et al. (2007) have become more and more popular and has been successfully applied to numerous inference problems in particle physics, with some further development of this algorithm, the resulting replacement for traditional MCMC sampling techniques. With some further development of this algorithm, the resulting Bayesian inference tool was announced to be publicly released to increase the acceptance ratio and the sampling efficiency. Building on these works and further pursuing the notion of detecting and characterizing multiple modes in the posterior from the distribution of nested samples, Feroz & Hobson (2008) introduced the MultiNest algorithm as a viable, general replacement for traditional MCMC sampling techniques. With some further development of this algorithm, the resulting Bayesian inference tool was announced to be publicly released in Feroz et al. (2009b). From then on, the MultiNest algorithm has become more and more popular and has been successfully applied to numerous inference problems in particle physics, cosmology, and astrophysics (Trotta et al. 2009; Feroz et al. 2009a, 2010; Martin et al. 2011; Graff et al. 2012; Karpenka et al. 2013; Kavanagh 2014). In Han & Han (2012), we have employed the MultiNest algorithm to build our BayeSED code for the SED fitting of galaxies. In the current version of BayeSED, we have employed the most recent version of MultiNest as described in Feroz et al. (2013). The newly developed MultiNest algorithm was largely improved by the technique known as INS, which increases the accuracy of the calculation of Bayesian evidence by up to an order of magnitude. To achieve the same level of accuracy, the higher evidence accuracy from INS could potentially speed up MultiNest by a factor of a few, if fewer live points or higher target efficiency are used. Comparison, cannot be easily obtained using most MCMC techniques. This is because the evaluation of the multidimensional integral in Equation (3) is a challenging numerical task.

The nested sampling, first introduced by Skilling (2004), provides an efficient method to calculate the Bayesian evidence while also producing posterior inferences as a by-product. So, by using the nested sampling method, we may achieve efficient Bayesian parameter estimation and model comparison simultaneously. This method has been further improved by the works of Mukherjee et al. (2006) and Shaw et al. (2007) to increase the acceptance ratio and the sampling efficiency. Building on these works and further pursuing the notion of detecting and characterizing multiple modes in the posterior from the distribution of nested samples, Feroz & Hobson (2008) introduced the MultiNest algorithm as a viable, general replacement for traditional MCMC sampling techniques. With some further development of this algorithm, the resulting Bayesian inference tool was announced to be publicly released in Feroz et al. (2009b). From then on, the MultiNest algorithm has become more and more popular and has been successfully applied to numerous inference problems in particle physics, cosmology, and astrophysics (Trotta et al. 2009; Feroz et al. 2009a, 2010; Martin et al. 2011; Graff et al. 2012; Karpenka et al. 2013; Kavanagh 2014). In Han & Han (2012), we have employed the MultiNest algorithm to build our BayeSED code for the SED fitting of galaxies. In the current version of BayeSED, we have employed the most recent version of MultiNest as described in Feroz et al. (2013). The newly developed MultiNest algorithm was largely improved by the technique known as INS, which increases the accuracy of the calculation of Bayesian evidence by up to an order of magnitude. To achieve the same level of accuracy, the higher evidence accuracy from INS could potentially speed up MultiNest by a factor of a few, if fewer live points or higher target efficiency are used.

### 2.3. Interpolation of Model SED Library

For the Bayesian inference of the SEDs of galaxies using an extensive sampling method such as MultiNest (Section 2.2), the SED model needs to be able to be evaluated at any point of its parameter space. However, it would be very computationally expensive to employ a detailed SED model, such as an evolutionary population synthesis model, to generate SEDs during the sampling of a high-dimensional and complex parameter space. Besides, for many SED models, only a precomputed library of model SEDs is available. Therefore, it is often very necessary to interpolate a model SED library. In this subsection, we introduce the interpolation method that we have used in the BayeSED code, taking the evolutionary population synthesis model as an example.

### 2.3.1. Building Evolutionary Population Synthesis Model SED Libraries

Currently, the evolutionary population synthesis model is the standard method for modeling the SEDs of galaxies. It is based on the theory of star formation and evolution, the empirical or theoretical stellar spectral library, and the chemical evolution theory of galaxies, and it models the SED of a galaxy as a sum of the contribution from individual stars. As mentioned in Section 1, there are still many uncertainties in these ingredients of an evolutionary population synthesis model. So, due to different treatments of these issues, there are many competing evolutionary population synthesis models and many possible parameterizations of the model.

In this paper, we have employed the model of Bruzual & Charlot (2003)(bc03) and Maraston (2005)(ma05), two of the most widely used evolutionary population synthesis models. For the bc03 model, the IMFs of Chabrier (2003) and Salpeter (1955) are used, while for thema05 model the IMFs of Kroupa (2001) and Salpeter (1955) are used. The SFHs of galaxies are assumed to be exponentially declining in the form of star formation rate (SFR) ∝ exp(−t/τ), where t is the time since the onset of star formation and τ is the e-folding star formation timescale. To consider the effect of dust attenuation, a uniform dust screen geometry with a Calzetti et al. (2000) dust extinction law is assumed.

To build SED libraries of the two evolutionary population synthesis models with different assumptions, we have employed a modified version of a grid-based SED fitting code—FAST (Kriek et al. 2009)—to actually generate the model SEDs. We have built four SED libraries, which we refer to as "bc03_sa," "bc03_ch," "ma05_sa," and "ma05_kr", respectively. The four SED libraries cover a parameter space with \( \log(t/yr) \) in the range of \([6.5, 11]\) in steps of 0.1, \( \log(Age/yr) \) in the range of \([7.0, 10.1]\) in steps of 0.05, and visual attenuation \( A_v \) in the range of \([0, 5]\) in steps of 0.2. For the bc03 model, the metallicity Z could be 0.004, 0.008, 0.02, or 0.05 while for the ma05 model, Z could be 0.001, 0.01, 0.02, or 0.04. In total, there are 243,434 model SEDs for each of the four libraries. A summary of these libraries is presented in Table 1.

### 2.3.2. Dimension Reduction of Model SED

with Principal Component Analysis

An SED model can be considered a mapping from parameter space \( X(x_1, x_2, \ldots, x_n) \) to the corresponding SED \( S(f_1, f_2, \ldots, f_n) \), where \( x_i \) represents a parameter and \( f_i \) represents a flux at a wavelength. Depending on the resolution of the SED, \( n \) could be equal to hundreds or even thousands. Therefore, for an SED model with many free parameters, the size of the required library

### Notes

1. In FAST code, “pr” indicates a photometric resolution SED and the number of wavelengths for an SED is 460.

2. The parameter is in the range of \([6.5, 11]\) in steps of 0.1.

### Table 1

Summary of Model SED Libraries

| Name   | Resolution | SFH | IMF       | Dust Extinction Law | \( \log(t/yr) \) | \( \log(Age/yr) \) | \( A_v/mag \) | Z          |
|--------|------------|-----|-----------|---------------------|-----------------|-----------------|--------------|------------|
| bc03_sa| pr(460)    | exp | Salpeter (1955) | Calzetti et al. (2000) | [6.5, 0.1, 11]  | [7.0, 0.05, 10.1] | [0, 0.2, 4]  | 0.004, 0.008, 0.02, 0.05 |
| bc03_ch| pr(460)    | exp | Chabrier (2003) | Calzetti et al. (2000) | [6.5, 0.1, 11]  | [7.0, 0.05, 10.1] | [0, 0.2, 4]  | 0.004, 0.008, 0.02, 0.05 |
| ma05_sa| pr(460)    | exp | Salpeter (1955) | Calzetti et al. (2000) | [6.5, 0.1, 11]  | [7.0, 0.05, 10.1] | [0, 0.2, 4]  | 0.001, 0.01, 0.02, 0.04 |
| ma05_kr| pr(460)    | exp | Kroupa (2001)  | Calzetti et al. (2000) | [6.5, 0.1, 11]  | [7.0, 0.05, 10.1] | [0, 0.2, 4]  | 0.001, 0.01, 0.02, 0.04 |
could be very huge. For the four SED libraries that we have built in Section 2.3.1, $n$ is equal to 460 which results in a size of 1.3 Gbyte for each library. However, due to the continuity of an SED when the flux at a given wavelength is changed, the fluxes at nearby wavelengths will be changed in a very similar way. This means that the fluxes at different wavelengths are not completely independent of each other and the actual number of dimensions of the SED could be much less than $n$. It is thus possible to apply some kind of dimensionality reduction technique to efficiently compress an SED.

One such technique is known as principal component analysis (PCA), also known as the Karhunen–Loeve transform. It is mathematically defined as an orthogonal linear transformation that transforms the original data to a new coordinate system. The goal of PCA for an SED library is to find an $m$-dimensional ($m < n$) linear model of the $n$-dimensional SED library that represents the original SEDs as accurately as possible in a least-squares sense. In the current version of our BayeSED code, we have employed the PCA algorithm in the SHARK machine learning library.5

We have applied PCA to the four SED libraries that we built in Section 2.3.1. It should be noted that we take the logarithm of SEDs before applying PCA. The PCA algorithm in Shark provides two linear models as its outputs. The first linear model, called the “encoder,” is a linear transformation from an $n$-dimensional SED $S$ to an $m$-dimensional vector $A(A_1, A_2, \ldots, A_m)$, where $A_i$ is the amplitude of the $i$th principal component. The second linear model, called the “decoder,” is the inverse transformation of the “encoder.” Thus, it is a mathematical defined as an orthogonal linear transformation (PCA), also known as the Karhunen–Loeve transform. It is also known as the principal component analysis (PCA), also known as the Karhunen–Loeve transform. It is

5 http://image.diku.dk/shark/sphinx_pages/build/html/index.html

differ in the IMF have almost identical principal components but a slightly different mean spectrum. However, it is worthwhile to note that the SEDs from the bc03 and ma05 models follow different distributions in the space of principal components as shown in Figures 3 and 4. While the SEDs from the two models conform to somewhat similar distributions in the PC1–PC2 space, they have very different distributions in the PC2–PC3 space. Meanwhile, the SEDs from the ma05 model show more complex distributions than those from the bc03 model. Furthermore, the relationship between the amplitudes of the principal components and physical parameters are different for the two models. Generally, the SEDs from each evolutionary population synthesis model show a unique distribution in principal component space, a trend which is more obvious in three-dimensional space, as shown in Figure 5. These differences may reflect the consequence of different methodologies and treatments of the TP-AGB phase in the two models.

The total number of principal components $m$ is equal to the dimension of the SED $n$, which is 460 in our case. However, to compress the original SED libraries, we need to ignore those principal components with much less contribution to the overall variance of the SEDs in the library. In this work, we choose to ignore those principal components with a contribution of $<10^{-6}$. This results in 23 principal components for the two SED libraries of the bc03 model and 26 principal components for the two SED libraries of the ma05 model. To check the reliability of the PCA method, we have compared the original SEDs in the “bc03_sa” SED library with those reconstructed from the first 23 principal components. As is clearly shown in Figure 6, the reconstructed SEDs are almost identical to the original SEDs in most cases. However, with the application of the PCA method, the size of the SED library is reduced to only 23/460 = 5% of the original library.

2.3.3. Interpolation with Artificial Neural Networks

With the application of the PCA method, the size of a model SED library will be significantly reduced, since an SED can now be represented by the amplitude of only a few principal components instead of the luminosity at many more wavelengths. On the other hand, the mapping from parameters $X$ to the corresponding SED $S$ can now be divided into the
mapping from $X$ to the amplitudes of the principal component $A$ and the mapping from $A$ to the final $S$. The latter mapping is actually the “decoder” provided in Section 2.3.2, which is a linear transformation from $A$ to $S$. So if we can map $X$ to $A$ very quickly, then we should be able to evaluate the original SED model at any point in its parameter space very efficiently. However, as shown in Figures 3–5, the relationship between $X$ and $A$ could be very complex.
One method of achieving such an efficient mapping from $X$ to $A$ is the ANN algorithm. ANNs are mathematical constructs originally designed to simulate some intellectual behaviors of the human brain. Just like a human brain, an ANN tries to understand the underlying relationship between two set of things (e.g., $X$ and $A$) by learning some instances which obey this relationship. When the learning procedure is successfully finished, the ANN could be used to predict the corresponding $A$ from any instance of $X$, including those that have not been learned before. In the last decade, ANN methods have been successfully used in many problems in astrophysics (Firth et al. 2003; Collister & Lahav 2004; Carballo et al. 2008; Yèche et al. 2010; Almeida et al. 2010; Silva et al. 2011).

Currently, there are many kinds and implements of ANNs using different programming languages. In Han & Han (2012), we modified the widely used ANNz code (Collister & Lahav 2004), which was originally built for estimating photometric redshifts, to be suited to interpolated SED models. However, to be able to control more freely every component of an ANN algorithm, we have employed a more general and configurable neural network library—the Fast Artificial Neural Network (FANN) Library—in the current version of BayeSED. Similar to ANNz, the ANN algorithm implemented in FANN is the most widely used multi-layer perceptron (MLP) feed-forward network. An MLP network consists of a number of layers of neurons. Basically, there are three types layers, referred to as the input layer, hidden layer, and output layer, respectively. In a feed-forward network, information propagates through the input layer, the hidden layer, and the output layer sequentially without any internal feedback. Commonly, the network architecture of such an ANN is denoted as $N_{\text{in}}:N_1:N_2:...:N_{\text{out}}$, where $N_{\text{in}}$ is the number of neurons in the input layer, $N_i$ is the number of neurons in $i$th hidden layer, and $N_{\text{out}}$ is the number of neurons in the output layer.

In Figure 7, we show the network architecture of an ANN used for the interpolation of the evolutionary population synthesis model of bc03 and ma05. The input layer of this ANN has four neurons, corresponding to the four parameters of the evolutionary population synthesis model. These neurons emit the value of the corresponding parameters to the neurons in the next layer. It is worth mentioning that all parameters have been normalized to have a zero mean and a standard deviation of 1 for a better performance of the ANN. An additional neuron, which is referred to as $b_1$, is a bias neuron. The bias neuron plays as an offset and always emits one. The capability of an ANN is mainly determined by the structure of its hidden layers. According to the universal approximation theorem (Cybenko 1989; Hornik 1991; Haykin 1999), a multilayer feed-forward network with only one hidden layer can approximate any continuous function to an arbitrary precision. However, the neurons in the hidden layer must have a continuous, bounded, and nonconstant activation function. Meanwhile, the number of neurons in the hidden layer needs to be chosen carefully according to the complexity of the problem under consideration. As the input layer, an additional bias neuron, $b_2$, is also needed. Finally, the last layer gives the output of an ANN. In our case, there is only one output which is the amplitude of a principal component.

In principle, more neurons in the hidden layer can provide a better result, but at the expense of much more training time. Practically, we found that a hidden layer with 30 neurons is good enough for the libraries of the bc03 and ma05 models. The choice of activation function for the neurons in the hidden layer and output layer is also crucial. The FANN library provides us with many possible choices for the activation function. For the neurons in the hidden layer, an activation function defined by Elliott (1993) was chosen. This activation function is similar to the commonly used sigmoid activation function but is easier to compute and therefore faster. For the neuron in the output layer,
a linear activation function was chosen to make sure that the output could be scaled to any value. Since one ANN gives only the amplitude of one principal component, 23 and 26 ANNs with the structure as shown in Figure 7 need to be trained for the libraries of the bc03 and ma05 models, respectively.

The training of an ANN is an optimization problem where we adjust the weights in the ANN to minimize the difference between the outputs of the ANN and that given by the instances in the training data. The universal approximation theorem of an MLP network states the existence of the solution of such an optimization problem. However, it told us nothing about how to actually find the solution. So the effectiveness of an ANN method largely depends on the algorithm used for training. The most widely used algorithm for the training of an ANN is the backpropagation algorithm (Rumelhart et al. 1986). As the name suggests, the error obtained by propagating an input through the network is then propagated back through the network while the weights are adjusted in such a way that the error becomes smaller. The original backpropagation algorithm has been further improved by more advanced algorithms, such as quickprop (Fahlman 1988) and RPROP (Riedmiller & Braun 1993; Igel & Hsken 2000). All of these training algorithms have been implemented in the FANN library. As for our cases, we found that the RPROP training algorithm performs best.

In addition to the RPROP algorithm, another strategy has been used for the training of ANNs. The set of input and output data, which in our case are the parameters of a stellar population synthesis model and the corresponding amplitudes of the principal components of SEDs, are sorted randomly and then split into three groups. The first group of data, called “training data,” is used by the RPROP algorithm to adjust the weights in the ANN. During the training of an ANN, the RPROP algorithm tries to minimize the difference between the output of the ANN and the results in the “training data.” However, if too much training is applied to these data, then the ANN will eventually over fit them. Over fitting means that the ANN will be able to fit the “training data” very precisely, but will lose the generalization for other data that have not been used by the RPROP algorithm during the training. Therefore, another group of data, referred to as “validation data,” is used to avoid over fitting. During the training, we trace the difference between the ANN output and the results in the “validation data,“ known as the error. The RPROP algorithm will be stopped when the normalized error begins to increase. Finally, the third group of data, called the “testing data,” will be used for an ultimate test of the ANN when the training is finished.

We have trained 23 and 26 ANNs with the structures shown in Figure 7 for the libraries of the bc03 and ma05 models, respectively. Here, we use the training of 23 ANNs for the “bc03_sa” SED library as an example. The SED library has a total of 243, 432 SEDs. We first apply the “decoder,” which is obtained using the PCA method as presented in Section 2.3.2, to the SEDs in this library to obtain the amplitudes of the their first 23 principal components. Then, the parameters and the corresponding amplitudes of the principal components are split into “training data” (50%), “validation data” (20%), and “testing data” (30%). These data are used as instances for the training of ANNs. In practice, we actually train the 23 ANNs for the “bc03_sa” SED library simultaneously. Since the code used for ANN training has been parallelized using OpenMP, the whole process of training 23 ANNs for this SED library can be finished in about 40 minutes with 23 Intel 2.20 GHz CPU cores. Finally, all of these ANNs and other information about this SED library are saved to one file, which is only 566 Kbytes in size. This single file will be used to replace the original “bc03_sa” SED library, which is about 1.3 Gbytes in size, while using BayeSED to interpret the observed photometric SEDs of galaxies.

In Figure 8, we provide an example of tracing the normalized error of three ANNs on the “validation data” as a function of the number of iteration during the training. As shown in the figure, the normalized errors of these ANNs have decreased dramatically during the first few hundred iterations. This demonstrated the power of the RPROP algorithm for ANN training. After about 1000 iterations, the error begins to decrease very slowly, which means the training tends to converge. As mentioned before, the training will be stopped when the error begins to increase. It is worth noting that the final error for the amplitudes of different principal components are slightly different. It seems that the lower-order principal component can be “learned” better by ANN. This is good for us, since the lower-order principal component is more important for the reconstruction of SEDs. We can test the effectiveness of the ANN method by comparing the amplitudes of the principal components of the original SEDs in the “bc03_sa” SED library, which is obtained with “encoder,” with those generated by ANNs. This is shown in Figure 9 for the “testing data” set. As mentioned before, these data have not been used in any way during the training of these ANNs. So, this kind of test should be very rigorous. As shown in the figure, it is clear that the amplitudes of the principal components of the original SEDs can be generated fairly well by ANNs.

Finally, the amplitudes of the principal components generated by ANNs can be used to reconstruct the SEDs. This is the ultimate goal of using the ANN interpolation method. In Figure 10, we test the effectiveness of this method by comparing the original SEDs in the “bc03_sa” SED library with those reconstructed by “decoder” from the amplitudes of the principal components generated by ANNs. As shown in the figure, the errors for SEDs reconstructed employing the ANN method are slightly larger than those obtained using only the PCA method (Figure 6). This is because ANNs cannot predict the amplitudes of the principal components without any error. However, the errors are still very small. In Figure 11, we provide an example...
Hence, the method of ANN interpolation of an SED library is used to obtain the amplitudes of the principal components generated with ANNs. It is clear that the original SED can be reconstructed pretty well using the amplitudes of the principal components generated with ANNs. The error induced by PCA is negligible, while that induced by ANNs is larger and seems to be wavelength dependent. However, in most cases the error is within 0.1 dex.

![Figure 9](image)

**Figure 9.** Test of the ANN interpolation method by comparing the amplitudes of the principal components of the original SEDs in the “bc03_sa” SED library, obtained with “encoder,” with those generated by ANN. The results are for the “testing data” set, which were not used in any way during the training of the ANNs. (A color version of this figure is available in the online journal.)

![Figure 10](image)

**Figure 10.** Test of the ANN interpolation method by comparing the original SEDs in the “bc03_sa” SED library with those reconstructed with “decoder” from the amplitudes of the principal components generated with ANNs. The probability density distribution of the errors is shown in the lower panel. The mean, standard deviation, and percentage of outliers (error > 3σ) for the distribution are shown at the top of this figure. (A color version of this figure is available in the online journal.)

with not too many neurons in one hidden layer has been found to be enough for most problems, there are no simple and general rules for this determination. Second, the training of ANNs may require too much time if the relationship between the parameters and the corresponding SEDs is too complex (highly nonlinear and/or uncontinuous). Third, although the trained ANNs may have generally perform well even for those instances that have not been used during the training, they cannot exactly reproduce those instances that have been used for the training. One of the methods that can overcome these shortcomings of the ANN method is the interpolation method with KNNs. We will use the KNN method as a complement to the ANN method.

The basic idea of the KNN method is very simple. To evaluate the distance between two points is defined as a Euclidean distance: $D(X, Y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$.

To find efficiently the first KNNs of any point, the known points need to be preprocessed into a data structure. Then, a look-up algorithm must be applied to this structure to find the nearest neighbors. For this purpose, we have employed a modified version of the Nearest-Neighbor-Regression algorithm in the Shark machine learning library. The library provides two algorithms for the look-up of the KNNs in a possibly high-dimensional parameter space. They are referred to as “SimpleNearestNeighbors” and “TreeNearestNeighbors,” respectively.

The “SimpleNearestNeighbors” algorithm is a brute force algorithm which simply evaluates the distance between pairs of points one by one. Therefore, the organization of known points is not important for this algorithm. The “TreeNearestNeighbors” algorithm, however, requires that the known points be organized in a specific way.
organized into some kind of tree structure in advance. Generally, the “TreeNearestNeighbors” algorithm is much faster than the “SimpleNearestNeighbors” algorithm as long as the dimension of the data points is not too high. The Shark machine learning library provides three choices for the tree structure: KDTree (k-dimensional tree), KHCTree (Kernel Hierarchical Clustering tree), and LCTree (Linear Cut tree), respectively. All three belong to the binary space-partitioning tree. The KDTree (Friedman et al. 1977) is the most widely used algorithm for nearest-neighbor search. It works well in low-dimensional data, but quickly loses its effectiveness as the dimensionality increases. KHCTree and LCTree are more advanced tree structures (see the document in the Shark library) for nearest-neighbor searches. It is often used in low-dimensional data, but quickly loses its effectiveness as the dimensionality increases. KHCTree and LCTree are more advanced tree structures (see the document in the Shark library) for nearest-neighbor searches. We found that the LCTree structure practically has the best performance for our case.

Similar to the ANN interpolation method in Section 2.3.3, we have applied the KNN interpolation method to predict the amplitudes of the principal components. It is possible for us to predict the corresponding SED directly from the given values of the input parameters. However, for KNN interpolation, we need to save all of the instances of an SED library to a file in the disk and reload it into the memory of the computer during the sampling of the parameter space. It is clear that the applicability of this method is largely limited by the size of the SED library. As shown in Section 2.3.2, the size of an SED library can be reduced to only 5% of the original by using the PCA method. By combining it with the PCA method, the KNN interpolation method can be more useful in practice. In Figures 12 and 13, we show a test of the KNN interpolation method, which is similar to that for the ANN method in Figures 9 and 10. We should note that the instances used for this test were not used to build the LCTree that was used for interpolation. Therefore, similar to the test of the ANN method using “testing data,” this should be a rigorous test. It is clear that for both the amplitudes of the principal components and the reconstructed SEDs, the KNN interpolation method could be even better than the ANN interpolation method.

There is no doubt that the KNN interpolation method has some advantages over the ANN interpolation method. For KNN interpolation, the intensive training process which is crucial for ANN is unnecessary. We only need to store properly the known instances of a model in a data structure that is convenient for searching. However, this does not mean that the KNN method is better than the ANN method in all respects. For example, the ANN interpolation is much faster than the KNN interpolation during the sampling of the parameter space. Besides, the size of the data that must be stored for KNN interpolation (e.g., 46 Mbyte for the “bc03_sa” SED library) is much larger than that for ANN (566 Kbytes). Furthermore, the KNN method is much more sensitive to outliers and the local structure of the data than the ANN method. Therefore, in practice, we use the KNN method as a complement to the ANN method. Thus, it would be worth checking whether the results could be different using the two methods.

2.4. Building the BayeSED Code

As the last part of this section, we introduce how the MultiNest (Section 2.2) and ANN (Section 2.3.3) or KNN (Section 2.3.4) algorithms are combined to build our BayeSED code for interpreting the multi-wavelength SEDs of galaxies. In Figure 14, we show the flowchart of the BayeSED code. The main input for BayeSED is the observed multi-wavelength photometric SED, including measurement errors, of a galaxy that needs to be interpreted. On the other hand, the priors for the SED model being used to explain the observations are considered to be additional inputs. This includes the allowed ranges for all of the free parameters of the model and their corresponding distributions. Currently, we only allow a uniform distribution of a free parameter. So for parameters with a large dynamic range, their logarithms should be used as free parameters instead. In the future, we plan to provide more choices for the distribution of priors, including those that are physically more informative.

The MultiNest sampling of the parameter space of the SED model being used to explain the observations lies at the heart of BayeSED. During the sampling, the MultiNest
The computation of the likelihood at a given point involves determining the model SED at that point. This is achieved by using the ANN algorithm (Section 2.3.3) or KNN algorithm (Section 2.3.4) to interpolate a pre-computed SED library. As mentioned before, the computation of a model SED is a major bottleneck for an efficient sampling of the possibly high-dimensional and complex parameter space of an SED model. Thanks to the ANN and KNN algorithms, this can be accomplished very quickly in our BayeSED code. The huge SED libraries are only used to build the final ANNs or KNNS. Therefore, they are no longer needed during the sampling of the parameter space and a small file including all of the necessary information is used instead.

When a model SED is generated with ANNs or KNNS, the effects of cosmological redshift and IGM on the SED are further considered. In the current version of BayeSED, we provide two options for considering the effects of IGM. These options are based on the prescriptions of Madau et al. (1996) and Meiksin (2006), respectively. There is the option to consider the effect of Galactic dust reddening and extinction by setting the value of $E(B-V)$ at the position of the object, where we use the $R$-dependent Galactic extinction curve of Fitzpatrick (1999) with a ratio of total to selective extinction of $R(V) = A(V)/E(B-V) = 3.1$. The galaxy redshift is considered as an optional free parameter for the fitting of SED. Then, the redshift and other physical parameters of a galaxy can be obtained simultaneously with our BayeSED code. However, while many publicly available codes have been designed for redshift determination, BayeSED is not optimized for that purpose. Therefore, other codes could be used to determine the redshifts of galaxies which could then be used in BayeSED. We will test the reliability of using BayeSED to determine the redshifts of galaxies in Section 4. Finally, the model SED is convolved with the transmission function of filters to obtain model fluxes that are directly comparable with multi-wavelength observations.

The likelihood value at a specific point of the parameter space as requested by the MultiNest sampler is obtained by comparing the model fluxes and the corresponding multi-wavelength observations. The distribution of observational errors are usually assumed to be Gaussian. Then, the normalized likelihood function is defined as

$$
    L(\theta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi} \sigma_{o,i}} \exp \left( -\frac{1}{2} \frac{(F_{o,i} - F_{m(\theta),i})^2}{\sigma_{o,i}^2} \right),
$$

where $\sigma_{o,i}$ is the observational error in the $i$th band, and $F_{o,i}$ and $F_{m(\theta),i}$ are the observed flux and model flux in the $i$th band, respectively. In practice, the term $1/\sqrt{2\pi} \sigma_{o,i}$ is usually omitted, since it is independent of the shape of the likelihood function. Therefore, in most Bayesian SED fitting, the definition of the likelihood function is simplified as

$$
    L(\theta) = \prod_{i=1}^{n} \exp \left( -\frac{1}{2} \frac{(F_{o,i} - F_{m(\theta),i})^2}{\sigma_{o,i}^2} \right).
$$

In the above definition of the likelihood, observations at different wavelength bands are assumed to be independent and only observational error have been considered. The possible systematic errors of the SED model, which could be important, especially for population synthesis models (Conroy et al. 2009; Cerviño 2013), have not yet been considered. The systematic error of an SED model is likely wavelength and model dependent, and so is not easy to properly consider. In the EAZY code (Brammer et al. 2008), this has been considered as a template error function. However, it is not clear how universal this kind of error function may be.

On the other hand, thanks to the application of PCA, an SED can be described by the amplitudes of the principal components. Then, the likelihood function could be defined as

$$
    L(\theta) = \prod_{i=1}^{N} \exp \left( -\frac{1}{2} \frac{(A_{o,i} - A_{m(\theta),i})^2}{\sigma_{o,i}^2} \right),
$$

where $N$ is the number of principal components. In this approach, the model SEDs need not be reconstructed from the amplitudes of the principal components over and over again, while the observed SED needs to be projected to the principal components only once. This is especially useful for the analysis of spectroscopic data (Chen et al. 2012) where the dimensions of the data are much larger than the number of necessary principal components. However, this is less helpful for the analysis.

---

**Figure 14.** Flowchart for interpreting SEDs of galaxies with BayeSED. Since BayeSED has been fully parallelized with MPI, this kind of analysis can be done simultaneously for multiple galaxies.
of photometric data where the dimensions of the data are comparable to the number of necessary principal components. Besides, it is not as straightforward to project the sparsely sampled photometric SED to the principal components (see also Wild et al. 2014).

Many outputs could be obtained by using the BayeSED code to interpret the multi-wavelength SED of a galaxy. First, we can obtain the Bayesian evidence of the model which is used to explain the observed SED. Second, we can obtain an estimate of all of the model parameters. As mentioned in Section 2.1, the posterior PDF includes all of the information about the parameters. However, in practice, it is not possible to report the results of Bayesian parameter estimation with a full PDF, especially for a large sample of galaxies. Therefore, it is very necessary to use some summary statistics instead. In BayeSED, we provide many summary statistics about a parameter, including the mean, median, maximum-a-likelihood (MAL, or best fit), and maximum-a-posteriori. The corresponding error of a parameter is estimated with the standard deviation or percentiles of the PDFs. In addition, the best-fit model SED and the corresponding amplitudes of the principal components, rest-frame absolute magnitudes, and observation-frame apparent magnitudes could be optionally obtained. Finally, it is worth noting that BayeSED has been fully parallelized with MPI. So it is possible to interpret simultaneously the multi-wavelength SEDs of a large sample of galaxies, and all results are saved into a single file.

3. APPLICATION TO A MOCK SAMPLE OF GALAXIES

As mentioned in Section 1, by interpreting the SEDs of galaxies, we try to solve the inverse problem of SED modeling: deriving the physical parameters of galaxies from their observed multi-wavelength photometric SEDs. The ability of an SED-fitting code to solve this problem can be properly tested by using mock samples of galaxies. Before applying the BayeSED code to interpret the SEDs of galaxies in a real sample, we will thus test its reliability in this section.

3.1. Building a Mock Sample of Galaxies

The starting point for building a mock sample of galaxies is a set of model SEDs of galaxies which has been obtained using a model of a galaxy SED, such as the evolutionary population synthesis model. Then, these model SEDs are transformed according to what real galaxies would experience in order to obtain mock fluxes at multiple wavelength bands. Since the true value of all the parameters of the mock observations are known in advance, it is easy to check if they can be recovered properly by an SED-fitting code.

In our case, we started from the SED libraries constructed in Section 2.3.1 to make four mock samples of galaxies. We have taken the bc03 model with a Salpeter (1955) IMF as an example. From the SED library, a total of 10,000 model SEDs are randomly selected to make a mock sample with 10,000 galaxies. These model SEDs are shifted to a random redshift $z$ ranging from 0 to 6, while also considering the effect of the IGM. We demand that the age of a galaxy must be smaller than the age of the universe at that redshift. Then, the model SEDs are convolved with the transmission function of filters to obtain model fluxes. Finally, some random noises with a Gaussian distribution are added to these model fluxes. The Gaussian distribution has a zero mean and a dispersion equal to 10% (signal-to-noise ratio (S/N)) of the model flux. The filters and corresponding errors are selected to mimic the $K_s$-selected sample of galaxies in the COSMOS/UltraVISTA field that will be studied in the next section.

We should note that there are other methods to build more realistic mock galaxy samples. For example, the distribution of luminosity and redshift could be drawn from the luminosity function of the galaxies, or the distributions of the physical parameters of galaxies could be predicted by a model for the formation and evolution of galaxies. However, these more realistic mock samples are not really necessary for a reasonable test of an SED-fitting code. A good SED-fitting code should be able to properly recover the physical parameters of galaxies from their multi-wavelength observations, regardless of how these parameters are distributed. Nevertheless, we should keep in mind that not all physical parameters of galaxies can be recovered equally well, even though the best possible SED-fitting code has been used. Besides the code itself, there are many other factors that can take part in determining the possibility of recovery, for example, the number of available filters and corresponding S/N, the relative importance of a parameter for determining the shape of SED, the degeneracies between parameters, etc. Therefore, the mock sample of galaxies is only used to check the internal consistency of BayeSED and the effects of intrinsic degeneracies between the parameters of an SED model (see, e.g., Walcher et al. 2008).

3.2. Interpretation and Results

We have applied our BayeSED code to interpret the mock sample of galaxies built in Section 3.1. For the interpolation of model SEDs, both the ANN and KNN methods have been used and the results obtained will be compared here. It is worth noting that the results presented here represent an overall verification of everything that is involved in the BayeSED code. Since the mock sample is built with the original SED library, all of the potential errors that are hidden in the code programming, or the PCA, ANN, KNN and MultiNest algorithms, could propagate into the final results.

Since the ANN and KNN methods are used to approximate the original bc03 model, they can be considered as two special versions of that model. Thus, it is meaningful to check their Bayesian evidence for the mock observations that are built from the original bc03 model. In Figure 15, we show the probability density distribution function (PDF) and cumulative distribution function (CDF) of the Bayes factor $ln(B_{knn,ann})$ for the mock sample of galaxies. It is clear from the figure that the results obtained with the KNN method have more Bayesian evidence than those obtained with ANN method. Except for the interpolation method, everything else is the same for the two approaches. So this indicates that the KNN method should be a better approximation of the original bc03 model.

In Figures 16–20, we compare the recovered values of redshift, age, metallicity, $e$-folding time, and dust extinction with their true values. Among these, the redshift can be best recovered. For both the ANN and KNN methods, the mean distribution of errors is almost zero, while the dispersion and the fraction of outliers with an error larger than $3\sigma$ are also very small. The age and dust extinction of galaxies can be recovered moderately well. This can be better understood from their posterior PDFs for a mock galaxy, as shown in Figure 21.

---

9 The MPI implemented in the MultiNest algorithm itself is switched off, since we found that it is not efficient for multiple SED fitting.

10 The PDFs in this paper are obtained kernel density estimation (KDE) method.
The Astrophysical Journal Supplement Series, 215:2 (24pp), 2014 November

Han & Han

Figure 15. PDF and CDF of the Bayes factor \( \ln(B_{\text{knn,ann}}) \) for the mock sample of galaxies. The bc03 model with a Salpeter (1955) IMF was used in this example. The KNN method gives a much better approximation to the original bc03 model than the ANN method, since only 26% of the mock galaxies favor the latter.

(A color version of this figure is available in the online journal.)

Figure 16. Recovered values of redshift (\( z_{\text{mean}} \)) compared with the true values (\( z \)) of the mock sample of galaxies. The redshifts and corresponding errors are estimated from the mean and standard deviation of the posterior distribution. The bc03 model with a Salpeter (1955) IMF was used in this example. At the top of this figure, we show the mean, the dispersion, and the fraction of outliers with an error larger than 3\( \sigma \) of the error distribution.

(A color version of this figure is available in the online journal.)

The best-fitting results for this mock galaxy are shown in Figure 22. The PDFs of log(age/yr) and \( A_v/\text{mag} \) show a weak second peak, indicating a degeneracy between them. Finally, the metallicity and e-folding time cannot be properly recovered. This seems to be the case because these two parameters are less important for shaping the SED, and thus are more easily affected by errors in the observations and the degeneracies with other parameters. As can be noted in Figure 21, the PDFs of log(Z/Z\(_\odot\)) and log(tau/yr) show many peaks, indicating serious degeneracies with other parameters. Meanwhile, it is clear from Figures 18 and 19 that their error distributions show a clear anti-correlation with the true value. This kind of trend simply implies that the two parameters cannot be effectively constrained by the observations, but are mainly constrained by their allowed range. Since a flat prior is assumed and the parameters are estimated with the mean of the PDF, the parameters tend to be overestimated at the lower end and underestimated at the higher end. In practice, we found that if the redshift is fixed to the right value, the two parameters can be recovered much better. However, it still seems difficult to recover them properly with only photometric data.

With the estimated values of the free parameters of an evolutionary population model, we are able to derive other parameters, such as the stellar mass and SFR. However, it would be more convenient to be able to estimate the free parameters and other derived parameters simultaneously. This is allowed in our BayeSED code. We achieved this by building another set of ANNs or KNNs to derive other parameters from the free parameters of the SED model. This process is similar to that for the SED, except that the output of ANNs or KNNs are the derived parameters instead of the SED. With this method, we are able to simultaneously estimate any number of parameters that are derived from the free parameters of the SED model.

As examples, in Figures 23–25, we compare the derived values of the stellar mass, stellar bolometric luminosity, and SFR with their true values. Generally, these parameters can...
be recovered properly, except for a small fraction of extreme outliers. Besides, the recovered values of these parameters seem to be biased for the outliers. This is especially clear for the estimation of the SFR. We found that this bias mainly depends on the method used for summary statistics. In Figure 26, we show the results obtained with the MAL estimation. We can see that the distribution of errors is more symmetric with the MAL estimation, and the mean of errors is much closer to zero. However, the dispersion and the fraction of outliers are not changed much. So for a population of galaxies, the Mean estimation and MAL estimation should be equally good.

In this section, we have systematically tested the reliability of BayeSED for recovering the free and derived parameters of a stellar population synthesis model from its observed multiwavelength photometric SEDs by employing a mock sample of galaxies. Generally, we believe that the results obtained with the BayeSED code are acceptable. Indeed, there are some extreme outliers in the recovered values which could partly be due to the BayeSED code itself. However, we believe that the errors in the recovered values of the parameters are dominated by the nature of these parameters themselves and the limited information about them in the photometric data.

4. APPLICATION TO A $K_S$-SELECTED SAMPLE IN THE COSMOS/ULTRAVISTA FIELD

With the systematic tests of the BayeSED code in Section 3, we believe it should be possible to obtain reliable results for real galaxies. In this section, we apply BayeSED to interpret the SEDs of galaxies in a $K_S$-selected sample in the COSMOS/ULTRA VISTA field. Since the tests in Section 3 showed that the results obtained with the KNN method have greater Bayesian evidence than those obtained with ANN, we only show the results obtained with the KNN method in this section.

4.1. A $K_S$-selected Catalog in the COSMOS/ULTRA VISTA Field

The UltraVISTA survey (McCracken et al. 2012) is a NIR sky survey with a unique combination of area and depth. When fully complete, the survey will cover an area of 1.8 deg$^2$ down to $K_s \sim 24.0$. Meanwhile, the survey field of UltraVISTA is the COSMOS field (Scoville et al. 2007), which has the most extensive multi-wavelength coverage and is an attractive field for the study of distant galaxies. Muzzin et al. (2013) presented a catalog covering 1.62 deg$^2$ of the COSMOS/ULTRA VISTA field. The catalog provides photometry in 30 photometric bands including the available GALEX, Subaru, Canada–France–Hawaii Telescope, VISTA, and Spitzer data. The sources in the catalog have been selected from the Muzzin et al. (2013) catalog with $<z<2$. A more comprehensive study of all the galaxies in the Muzzin et al. (2013) catalog with BayeSED will be presented in a future work. In addition to the photometric catalog, Muzzin et al. (2013) also provide a
show a sharp peak around the true values. The PDFs of log($Z/Z_\odot$) for all of the galaxies in the survey. We population parameters determined using the FAST SED-fitting (Brammer et al. 2008) and a catalog of stellar masses and stellar catalog of photometric redshifts computed with the EAZY code (Kriek et al. 2009) for all of the galaxies in the survey. We will compare our results obtained with the BayeSED code with those of Muzzin et al. (2013) obtained with the FAST code.

4.2. Interpretation and Results

For comparison with the results of Muzzin et al. (2013), we have used the Bruzual & Charlot (2003) or Maraston (2005) SED models with solar metallicity, Calzetti et al. (2000) dust extinction law, and an exponentially declining SFH to interpret the SEDs of galaxies in our sample. Either the Chabrier (2003) or Salpeter (1955) IMF was used for the bc03 model, while the Kroupa (2001) or Salpeter (1955) IMF was used for the ma05 model. In total, four free parameters are involved, including log($\tau$/yr) in the range of [0, 4]. During the sampling with MultiNest, the four parameters may be uniformly and continuously selected from the allowed parameter space. As another prior, the age of a galaxy is restricted to be less than the age of the universe at that redshift of the galaxy. In BayeSED, the scale factors of the model SEDs are not considered as free parameters during the sampling of the parameter space with MultiNest. Instead, they are uniquely determined using the efficient iterative algorithm of Sha et al. (2007) as in the EAZY code. With this algorithm, a linear combination of multiple SED models can be used to interpret the observed SEDs of galaxies. In this paper, only the stellar
Figure 23. Derived stellar mass ($M_{*\text{mean}}$) compared with the true values ($M_*$) of the mock sample of galaxies. The stellar mass is estimated with the mean of the posterior distribution of the stellar mass. The bc03 model with a Salpeter (1955) IMF was used in this example. At the top of this figure, we show the mean, dispersion, and the fraction of outliers with an error larger than $3\sigma$ of the distribution of errors.

(A color version of this figure is available in the online journal.)

Figure 24. Similar to Figure 23, but for the bolometric luminosity ($L_{bol}$).

(A color version of this figure is available in the online journal.)

Figure 25. Similar to Figure 23, but for the star formation rate (SFR).

(A color version of this figure is available in the online journal.)

Figure 26. Similar to Figure 25, but now the MAL estimations of the star formation rates of galaxies have been used.

(A color version of this figure is available in the online journal.)

Figure 27. Age of galaxies obtained with our BayeSED code compared with that obtained with the FAST code by Muzzin et al. (2013). The parameter is estimated with the mean of the posterior distribution. The probability density distribution of the difference between the results of the two codes is shown in the lower panel. At the top of the figure, the mean, the dispersion, and the fraction of outliers with a difference larger than $3\sigma$ are shown.

(A color version of this figure is available in the online journal.)

population synthesis model of bc03 or the ma05 model is employed to interpret the observed SEDs. Therefore, to be consistent with this SED model selection, the Spitzer bands with wavelengths longer than 4.5 $\mu$m have not been used during the fitting of SEDs, since no model for dust emission has been used.

4.2.1. Comparison with the Results of Muzzin et al. (2013)

In Section 3, we verified the internal consistency of BayeSED using a mock sample of galaxies. Here, we instead check the external consistency of BayeSED with the widely used FAST code as employed by Muzzin et al. (2013). Due to the very different methodologies employed by the two codes, we expect some differences for the results obtained by them. However, the results obtained by the two codes should be generally consistent with each other.

In Figures 27–31, we compare the values of age, $e$-folding time, dust extinction, stellar mass, and SFR obtained using the BayeSED code with those obtained with the FAST code by Muzzin et al. (2013). For each parameter, the mean, the dispersion, and the fraction of outliers with a difference larger than $3\sigma$ are shown in the lower panel.
BayeSED with those obtained using FAST by Muzzin et al. (2013). Only the results obtained with the Bruzual & Charlot (2003) model and the Chabrier (2003) IMF have been shown. As shown in the figures, except for a small fraction of extreme outliers, our results are generally consistent with those of Muzzin et al. (2013). This is more clear for those parameters that are more likely to be well constrained, such as the stellar mass.

In addition, the estimation of the e-folding time (tau) seems very similar for the two codes. It is much better than the case for a mock sample of galaxies as shown in Figure 19. This is mainly because the redshift of a galaxy is fixed to the spectroscopic redshift, instead of as an additional free parameter that needs to be estimated from the observations.

An important advantage of BayeSED over traditional grid-based SED-fitting methods is that the parameter space of SED models can be sampled extensively and continuously. With the grid-base methods, the estimation of parameters is only allowed within the precomputed set of grid points. This is a very unnatural restriction when interpreting the SEDs of real galaxies and can result in biased results. In Figure 32, we have shown the distribution of stellar mass versus SFR with the results obtained with BayeSED and those obtained with the grid-based FAST SED-fitting code. Generally, the two sets of results are consistent with each other, especially for the star-forming main sequence. However, the results obtained with the grid-based code show some unnatural parallel groups which do not exist in the results obtained with BayeSED. Actually, this kind of issue is even more clear for the free parameters of an SED model.

As mentioned in Section 1, there are many uncertainties in the population synthesis modeling of galaxy SEDs. Therefore,
with the ability of the BayeSED code to efficiently compute the Bayesian evidence in hand, it would be very interesting to perform a Bayesian model comparison for different population synthesis modeling. In Figure 33, we show the PDF and CDF of the Bayes factor ln(B) for 5467 galaxies in the $K_s$-selected sample with spectroscopic redshift. Four different combinations of the population synthesis model and IMF are considered as four different models for Bayesian model comparison. The Bruzual & Charlot (2003) model and solar metallicity, the one used in the work of Muzzin et al. (2013), is used as the base model for the computation of all the Bayes factors hereafter. The distributions show that the base model statistically has greater Bayesian evidence than the other three models, since only 45%, 40%, and 24% of galaxies support them, respectively. In general, the Bruzual & Charlot (2003) model statistically has greater Bayesian evidence than the Maraston (2005) model, regardless of which IMF has been used.

(A color version of this figure is available in the online journal.)

For many works of SED-fitting of galaxies, the solar metallicity of the stellar population is commonly assumed even for high-redshift galaxies. Apparently, this is not a very reasonable assumption for galaxies in the real universe. An excuse for this assumption is that it is usually very hard to determine the stellar population metallicity of galaxies with photometric data only. On the other hand, the SED-fitting of galaxies would be much more time-consuming, especially for grid-base methods. However, biased results could be obtained if solar metallicity is assumed for all galaxies. To test the importance of the assumption of metallicity, we have employed the BayeSED code to interpret the SEDs of the same sample of galaxies, but with metallicity as an additional free parameter ranging from $0.2 Z_\odot$ to $2 Z_\odot$. In practice, we found that this will not obviously increase the time of computation for our BayeSED code. In Figure 34, we show the PDF and CDF of the Bayes factor ln(B) for the same sample of galaxies in this case. Compared with the results in Figure 33, the Bayesian evidence for all models is clearly increased. The two bc03 models statistically have larger Bayesian evidence than the base model. Meanwhile, for the two ma05 models, only the one with a Kroupa (2001) IMF has a slightly larger support rate than the base model. Besides, it is worth noting that the Maraston (2005) model seems more sensitive to the different IMF selection than the Bruzual & Charlot (2003) model. Generally, with metallicity as an additional free parameter, it becomes more clear that the Bruzual & Charlot (2003) model has statistically greater Bayesian evidence than the Maraston (2005) model.

As mentioned in Section 2.4, the redshift of a galaxy can also be set as a free parameter in the BayeSED code. Therefore, it is possible to simultaneously obtain the photometric redshift and stellar population parameters of a galaxy with BayeSED while using the same set of SED models, and thus it is more self-consistent. Here, we test the reliability of BayeSED for the determination of the photometric redshifts of galaxies. The distribution of Bayesian evidence for this case is shown in Figure 35. Compared with the results shown in Figure 34 where the spectroscopic redshifts have been used, the Bayesian evidence of all the models decreases a little in this case. In Figure 36, we compared the estimated photometric redshifts with the spectroscopic redshifts for the galaxies in our sample. The performance of a code for photometric redshift estimation is usually judged by the root mean square (rms) of $(z_p - z_s)/(1 + z_s)$. In our case, $\sigma_{\text{rms}} = 0.0449$. When outliers with errors larger than $3 \sigma$ (48/5467 = 0.88%) are removed, $\sigma_{\text{rms}} = 0.0254$. This outperforms the result of Benítez (2000) (0.06), who also employed a Bayesian approach and employed more informative physical priors. Our results are not as good as those of Muzzin et al. (2013) ($\sigma_{\text{rms}} = 0.013$), who have employed the EAZY code and considered the effect of a zero-point offset from an iterative procedure. However, they only considered 5119 galaxies with high-quality spectroscopic redshifts and uncontaminated photometry. A better judgment for the performance of photometric redshift estimation can be achieved using the normalized median absolute deviation.
Figure 34. Similar to Figure 33, but now the metallicity is set to be an additional free parameter ranging from 0.2 \( Z_\odot \) to 2 \( Z_\odot \). It is clear that the Bayesian evidence of all the models has increased significantly. The two bc03 models have a much higher support rate than the base model. Meanwhile, for the two ma05 models, only the one with a Kroupa (2001) IMF has a slightly higher support rate than the base model. Therefore, it becomes even clearer that the Bruzual & Charlot (2003) model statistically has greater Bayesian evidence than the Maraston (2005) model. Meanwhile, it seems that the latter model is more sensitive to the different choice of IMF than the former model.

(A color version of this figure is available in the online journal.)

\[ \Delta z = z_p - z_s, \] which is defined (Brammer et al. 2008) as

\[ \sigma_{\text{NMAD}} = 1.48 \times \text{median} \left( \frac{|\Delta z - \text{median}(\Delta z)|}{1 + z_p} \right). \] (10)

\( \sigma_{\text{NMAD}} \) is less sensitive than \( \sigma_{\text{rms}} \) to the outliers. In our case, \( \sigma_{\text{NMAD}} = 0.0185 \), while the median of the error is \(-0.0135\) and the fraction of outliers is 1.45%. Except for the larger median value, this is better than the results of Brammer et al. (2008) using the EAZY code, which is applied to a smaller sample with 1989 galaxies but spans a wider redshift range. Given the difficulty of using a population synthesis model instead of using carefully selected templates or employing some kind of empirical training procedure to estimate the photometric redshift of galaxies, the degree of accuracy achieved by BayeSED is acceptable. In the future version of BayeSED, we plan to add more optimizations that have been adopted by many codes for a better estimation of the photometric redshift.

4.2.3. The Distribution of Galaxy Color and Stellar Population Parameters

The observed distributions of the galaxy properties provide important clues for understanding the formation and evolution of galaxies, and benchmarks for discriminations between different semi-analytic modelings and hydrodynamical simulations of galaxy formation. However, a detailed mathematical characterization of the distributions of these properties and a full theoretical understanding of them in the context of galaxy formation and evolution (Somerville et al. 2008; Schaye et al. 2010; Buzzoni 2011) is beyond the scope of this paper. Here, we show some well-known features of the distribution of galaxy properties obtained with the BayeSED code. For the results presented in this subsection, the spectroscopic redshifts of galaxies have been used and the metallicity is set to be a free parameter ranging from 0.2 \( Z_\odot \) to 2 \( Z_\odot \). In addition, only the results obtained with the Bruzual & Charlot (2003) model and the Chabrier (2003) IMF have been presented, since it has the largest Bayesian evidence as shown in Figure 34.
In Figure 37, we show the distribution of the rest-frame $U-B$ color of galaxies in our sample. The well-known bimodal distribution of the colors of galaxies is clearly shown in the figure. These galaxies can be divided into the red ones with $U - B > 1.1$ and the blue ones with $U - B < 1.1$. The red and blue galaxies are thought to be fundamentally different populations of galaxies. This can be more clearly noted in the $M_\star - SFR$ diagram, as shown in Figure 38. Generally, blue galaxies are mainly the star-forming galaxies in the “main sequence,” while red galaxies are primarily quiescent galaxies with negligible, if any, star formation. Meanwhile, there are a few galaxies with SFR $> 10$ above the “main sequence.” These galaxies have the bluest color in our sample and should be the starburst galaxies. On the other hand, a portion of red galaxies are actually star-forming galaxies with $M_\star \gtrsim 3 \times 10^{10} M_\odot$.

It is well known that many physical parameters, such as metallicity, dust attenuation, and age, can affect the color of galaxies. These parameters degenerate with each other to a certain extent, preventing them from being determined accurately simultaneously. However, they should be accurate enough to perform some qualitative analysis. Moreover, we can check whether the distributions of these parameters are reasonable in the context of galaxy evolution.

As shown in Figure 18, it is usually hard to determine the stellar metallicity of galaxies with photometric data only. However, since the spectroscopic redshifts of galaxies have been used, we expect that the estimated metallicities of galaxies here should be more accurate. In Figure 39(a), we show the distribution of the stellar metallicity of galaxies in the $M_\star - SFR$ diagram. As is clearly shown in the figure, the blue galaxies all have low metallicity while the red galaxies have much higher, near-solar metallicity. This is reasonable in the context of galaxy evolution. Since the blue galaxies are mainly star-forming and starburst galaxies, the stars in these galaxies should have been formed recently from low-metallicity gas. On the other hand, red galaxies are fully evolved galaxies where many stars are formed from recycled gas enriched by the last generations of star formation. Meanwhile, in Figure 39(b), we show the distribution of dust extinction of the same galaxies in the $M_\star - SFR$ diagram. The starburst galaxies have the highest dust extinction, while the more massive star-forming galaxies show less but still important dust extinction. On the contrary, there is very little dust extinction in either the quiescent or low mass star-forming galaxies.

Usually, the stellar population age of galaxies is thought to be the main parameter for the color of galaxies, since galaxies become redder when they age. However, as shown in Figure 39(c), it seems difficult to separate galaxies into red and blue based on age alone. The starburst galaxies which are the youngest in our sample are indeed blue galaxies. However, some blue low mass star-forming galaxies, instead of the reddest quiescent galaxies, are the oldest galaxies in our sample. As related to age, in Figure 39(d), we show the distribution of the $e$-folding time of SFH in the $M_\star - SFR$ diagram. Interestingly, the galaxies in our sample can be clearly divided into at least five groups with this distribution. The blue star-forming galaxies in the main sequence (G1) have the longest SFH $e$-folding times. On the other hand, the red quiescent galaxies (G2) have much shorter SFH $e$-folding times. Most galaxies in our sample belong to these two groups. The galaxies with the shortest SFH $e$-folding times (G3) are located between the last two groups. These galaxies are also unique in the distribution of age, as shown in Figure 39(c). Meanwhile, the galaxies at the high-mass end of the main sequence (G4) constitute the fourth group...
group. Finally, those starburst galaxies with very young ages, the highest dust extinction, and much shorter e-folding times than normal star-forming galaxies are considered to be the fifth group of galaxies (G5).

The galaxies in the G3 and G4 groups are likely in the transition stages between red and blue galaxies. To check if they are related to AGN activities, we show the distribution of the apparent magnitude at 24 μm in Figure 39(e). Meanwhile, we show the distribution of the redshifts of these galaxies in Figure 39(f). For most of these galaxies, the fluxes at 24 μm correspond to the rest-frame mid-IR fluxes, which are thought to be responsible by dust heated by AGNs (Nenkova et al. 2002, 2008a, 2008b; Elitzur 2006; Fritz et al. 2006; Han et al. 2012). As shown clearly in Figure 39(e), galaxies at the

Figure 39. Distribution of (a) age, (b) metallicity, (c) dust extinction, (d) e-folding time of SFH, (e) apparent magnitude at 24 μm, and (f) redshift in the $M^*-$SFR diagram. With these distributions, five populations of galaxies can be easily recognized and are most obvious in (d).

(A color version of this figure is available in the online journal.)
high-mass end of the star-forming main sequence have the strongest mid-IR emissions. These galaxies should be the composite galaxies. Most galaxies in the G4 group show important mid-IR emission, while most galaxies in the G3 group do not. The galaxies in the G3 group could be galaxies at an earlier evolution stage, when the AGN activity is still weak or not yet launched.

5. SUMMARY AND CONCLUSIONS

In this paper, we have described an updated version of BayeSED. Based on previous work (Han & Han 2012, 2013), we have presented a major update to this code in this paper. First, the most up-to-date version of the MultiNest Bayesian inference (Section 2.1) tool has been employed. The MultiNest algorithm (Section 2.2) has recently been improved through INS to allow a more efficient sampling of high-dimensional parameter space and more accurate calculation of Bayesian evidence. Second, besides the ANN method, we have added the KNN method as another method for the interpolation of model SED libraries in BayeSED (Section 2.3). For both ANN and KNN, we have defined a file format to store all of the necessary information about an SED model such that almost any SED model can be easily used by BayeSED to interpret the observed SEDs of galaxies. Third, the redshift of a galaxy can be set as a free parameter and the effects of IGM and Galactic extinction have been considered. Therefore, it is now possible to obtain the redshift and other physical parameters of a galaxy, both simultaneously and self-consistently. Fourth, the main body of BayeSED has been completely rewritten in C++ in an object-oriented programming fashion and parallelized with MPI. With BayeSED, it is now possible to analyze the SEDs of a large sample of galaxies with a detailed Bayesian approach that is based on an intensive sampling of the parameter space of SED models.

We have systematically tested the reliability of the BayeSED code to recover the physical parameters of galaxies from their multi-wavelength photometric SEDs with a mock sample of galaxies (Section 3). For both the ANN and KNN methods, the tests show that BayeSED can recover the redshift and stellar population parameters reasonably well. We found that different parameters can be recovered with a varying degree of accuracy. While the redshift, age, dust extinction, stellar mass, bolometric luminosity, and SFR can be properly recovered, it is usually hard to recover the metallicity and ε-folding time of SFH with photometric data, especially when the spectroscopic redshift is not available. We believe that this is due to the nature of these parameters themselves and the limited information about them in the photometric data. Meanwhile, the tests showed that the KNN interpolation method, although more memory and time-consuming than the ANN method, may lead to more accurate results.

We systematically applied BayeSED to interpret the observed SEDs of a large sample of galaxies in the COSMOS/ULTRA-VISTA field with different evolutionary population synthesis models (Section 4). A Bayesian model comparison of evolutionary population synthesis models has been accomplished for the first time. Using a Ks-selected sample of galaxies with spectroscopic redshifts and mostly less than one, we found that the Bruzual & Charlot (2003) model, statistically speaking, has greater Bayesian evidence than the Maraston (2005) model. In addition, we found that the Maraston (2005) model is more sensitive to the different choice of IMF than the Bruzual & Charlot (2003) model. However, the conclusion is drawn in a statistical sense and could be different for samples of galaxies at higher redshift (Maraston et al. 2006, 2010; Henriques et al. 2011), which is worthy of further investigation. Meanwhile, we found that by using stellar metallicity as an additional free parameter, the Bayesian evidence for stellar population synthesis models can be increased significantly. Therefore, we conclude that it is important to set metallicity as a free parameter to obtain unbiased results, even if this parameter cannot be estimated very accurately with photometric data only.

We have compared our results obtained using our BayeSED code with that obtained using the widely used FAST code, and found good agreement. Nevertheless, we found that the parameters estimated with BayeSED show more natural distributions than more conventional grid-based SED-fitting codes such as FAST. Also, based on the rest-frame color and stellar population parameters obtained with BayeSED, we recognized five distinct populations of galaxies in the Ks-selected sample of galaxies. They may represent galaxies at different stages of evolution or in different environments.

With the systematic tests using a mock sample of galaxies and the comparison with a popular grid-based SED-fitting code for a real sample of galaxies, we conclude that the BayeSED code can be reliably applied to interpret the SEDs of large samples of galaxies. Based on the MultiNest algorithm allowing intensive sampling of parameter space and efficient computation of the Bayesian evidence of the SED models, BayeSED could be a powerful tool for investigating the formation and evolution of galaxies from the rich multi-wavelength observations currently available. In particular, with the efficient computation of Bayesian evidence for SED models, BayeSED could be useful for further developing evolutionary population synthesis models and other SED models for galaxies. Besides, while we have only applied BayeSED to the photometric data so far, it should be straightforward to apply similar methods to the spectroscopic data in the future.

We thank an anonymous referee for valuable comments that helped to improve this paper. The authors gratefully acknowledge the computing time granted by the Yunnan Observatories and provided by the facilities at the Yunnan Observatories Supercomputing Platform. We thank Professor Houjun Mo for helpful discussions about Bayesian evidence and Lulu Fan for helpful discussions about Bayesian SED-fitting of galaxies. This work is supported by the National Natural Science Foundation of China (grant No. 11303084, U1331117, 11390374), the Science and Technology Innovation Talent Programme of the Yunnan Province (grant No. 2013HA005), and the Chinese Academy of Sciences (grant No. XDB09010202).

REFERENCES

Abazajian, K. N., Adelman-McCarthy, J. K., Agüeros, M. A., et al. 2009, ApJS, 182, 543
Acquaviva, V., Gawiser, E., & Guaita, L. 2011, ApJ, 737, 47
Almeida, C., Baugh, C. M., Lacey, C. G., et al. 2010, MNRAS, 402, 544
Atwood, W. B., Abdol, A. A., Ackermann, M., et al. 2009, ApJ, 697, 1071
Baldry, I. K., Glazebrook, K., Brinkmann, J., et al. 2004, ApJ, 600, 681
Baugh, C. M. 2006, RPPh, 69, 3101
Bell, E. F., McIntosh, D. H., Katz, N., & Weinberg, M. D. 2003, ApJS, 149, 289
Benitez, N. 2000, ApJ, 536, 571
Bolzonella, M., Miralles, J.-M., & Pelló, R. 2000, A&A, 363, 476
Bonoli, S., Marulli, F., Springel, V., et al. 2009, MNRAS, 396, 423
Bower, R. G., Benson, A. J., Malbon, R., et al. 2006, MNRAS, 370, 645
Brammer, G. B., van Dokkum, P. G., & Coppi, P. 2008, ApJ, 686, 1503
Bruzual, A. G. 1983, ApJ, 273, 105
Springel, V., White, S. D. M., Jenkins, A., et al. 2005b, *Natur*, 435, 629
Tarantola, A. 2005, *Inverse Problem Theory and Methods for Model Parameter Estimation* (Philadelphia, PA: Society for Industrial and Applied Mathematics)

Tian, B., Deng, L., Han, Z., & Zhang, X. B. 2006, *A&A*, 455, 247
Tinsley, B. M. 1972, *A&A*, 20, 383
Tinsley, B. M., & Gunn, J. E. 1976, *ApJ*, 203, 52
Tojeiro, R., Heavens, A. F., Jimenez, R., & Panter, B. 2007, *MNRAS*, 381, 1252
Tremonti, C. A., Heckman, T. M., Kauffmann, G., et al. 2004, *ApJ*, 613, 898
Trotta, R. 2008, *ConPh*, 49, 71

Verde, L., Peiris, H. V., Spergel, D. N., et al. 2003, *ApJS*, 148, 195
Walcher, C. J., Böker, T., Charlot, S., et al. 2006, *ApJ*, 649, 692
Walcher, C. J., Lamareille, F., Vergani, D., et al. 2008, *A&A*, 491, 713
Walcher, J., Groves, B., Budavári, T., & Dale, D. 2011, *Ap&SS*, 331, 1
Wild, V., Almaini, O., Cirasuolo, M., et al. 2014, *MNRAS*, 440, 1880
Yèche, C., Petitjean, P., Rich, J., et al. 2010, *A&A*, 523, A14
Zhang, F., Han, Z., Li, L., & Hurley, J. R. 2005, *MNRAS*, 357, 1088
Zibetti, S., Gallazzi, A., Charlot, S., Pierini, D., & Pasquali, A. 2013, *MNRAS*, 428, 1479