CIGALEMC: GALAXY PARAMETER ESTIMATION USING A MARKOV CHAIN MONTE CARLO APPROACH WITH CIGALE

PAOLO SERRA1, ALEXANDRE AMBLARD3, PASQUALE TEMI1, DENIS BURGARELLA2, ELODIE GIOVANNOLI2, VERONIQUE BUAT2, STEFAN NOLL3, AND STEPHEN IM1
1 Astrophysics Branch, NASA Ames Research Center, MS 245-6, Moffett Field, CA 94035, USA
2 Observatoire Astronomique de Marseille-Provence, 38 rue Frederic Joliot-Curie, 13388 Marseille Cedex 13, France
3 Institut für Astro- und Teilchenphysik, Universität Innsbruck, Technikerstr.258, 6020 Innsbruck, Austria

ABSTRACT
We introduce a fast Markov Chain Monte Carlo (MCMC) exploration of the astrophysical parameter space using a modified version of the publicly available code Code Investigating GALaxy Emission (CIGALE). The original CIGALE builds a grid of theoretical spectral energy distribution (SED) models and fits to photometric fluxes from ultraviolet to infrared to put constraints on parameters related to both formation and evolution of galaxies. Such a grid-based method can lead to a long and challenging parameter extraction since the computation time increases exponentially with the number of parameters considered and results can be dependent on the density of sampling points, which must be chosen in advance for each parameter. MCMC methods, on the other hand, scale approximately linearly with the number of parameters, allowing a faster and more accurate exploration of the parameter space by using a smaller number of efficiently chosen samples. We test our MCMC version of the code CIGALE (called CIGALEMC) with simulated data. After checking the ability of the code to retrieve the input parameters used to build the mock sample, we fit theoretical SEDs to real data from the well-known and -studied Spitzer Infrared Nearby Galaxy Survey sample. We discuss constraints on the parameters and show the advantages of our MCMC sampling method in terms of accuracy of the results and optimization of CPU time.

Key words: galaxies: fundamental parameters – methods: data analysis

Online-only material: color figures

1. INTRODUCTION
The spectral energy distribution (SED) of galaxies depends on many physical processes related to the emission from different stellar populations, absorption, and re-emission from dust and gas as well as the possible presence of active galactic nuclei (AGNs). Each process has been studied by many authors; libraries of stellar population models (Fioc & Rocca-Volmerange 1997; Bruzual & Charlot 2003; Maraston 2005), fitting curves for dust emission (Calzetti et al. 1994, 2000; Witt & Gordon 2000), and studies of emission of dust grains (Chary & Elbaz 2001; Dale & Helou 2002; Lagache et al. 2003, 2004; Siebenmorgen & Krügel 2007; Silva et al. 1998; Dopita et al. 2005; da Cunha et al. 2008) are the basis of sophisticated fitting codes which derive physical parameters such as stellar mass, star formation rate (SFR), dust luminosity, and so on.

Many parameters are usually necessary to describe these processes and model theoretical SEDs of galaxies. A grid of theoretical SED models is usually built and fitted to the data and statistical properties are derived for the parameters of interest. A big drawback of any grid-based method is that, for any fitting process, the time to build models grows linearly with the number of models and then about exponentially with the number of parameters involved: such approaches are difficult to implement for complex models involving a sufficiently large number of parameters or when a fine sampling of the parameter space is necessary in order to retrieve statistically robust results. In the past few years, Markov Chain Monte Carlo (MCMC) techniques have started being widely used in physics. In cosmology, parameter estimation from cosmic microwave background data with MCMC methods has been introduced in Christensen et al. (2001) and has been implemented in the publicly available code COSMOMC (Cosmological Monte Carlo; Lewis & Bridle 2002) in astrophysics, an MCMC approach to the stellar population synthesis modeling has been introduced in Conroy et al. (2009).

Here we use COSMOMC as a generic sampler and interface it to the publicly available code Code Investigating GALaxy Emission (CIGALE; Noll et al. 2009) in order to allow a fast and accurate evaluation of the multidimensional parameter space probed by this code. The main advantage of this method is that the computing time to fit the data scales approximately linearly with the number of parameters involved, allowing the user to consider complex models with many parameters for only small additional computational time. MCMC techniques also allow us to probe the shape of the probability distribution, giving far more information than just best fit and marginalized values for the parameters.

The paper is organized as follows; in the next section we briefly describe CIGALE, introducing the main parameters used in the subsequent sections. We then explain the MCMC technique implemented in the modified version of CIGALE, which we call CIGALEMC. We test our code using a mock sample of 62 galaxies already used in Giovannoli et al. (2011) and apply it to a real galaxy sample with data from the Spitzer Infrared Nearby Galaxy Survey (SINGS; see Kennicutt et al. 2003). We always consider a flat cosmological model with \( \Omega_m = 0.3 \), \( \Omega_k = 0.7 \), and \( H_0 = 70 \, \text{km s}^{-1} \, \text{Mpc}^{-1} \). Finally, we give our results and conclusions.

---

1 http://www.oamp.fr/cigale/
5 http://www.oamp.fr/cigale/
6 During the completion of this work, we noticed that Acquaviva et al. (2011) have performed a similar work in the context of the code GALAXEV developed by Bruzual & Charlot (2003).
2. THE CODE CIGALE

CIGALE calculates a grid of theoretical SEDs and fits to observational input data constituted by photometric filter fluxes ranging from UV to IR. For a detailed description of the code and its application to real data, we refer the interested reader to a number of other papers (Burgarella et al. 2005; Noll et al. 2009; Giovannoli et al. 2011; Buat et al. 2010, 2011b). In the following, we briefly summarize its main characteristics and the basic parameters used in the next sections. Our notation follows the one introduced in Giovannoli et al. (2011).

2.1. Stellar Populations and Star Formation Rate

CIGALE combines both old and young stellar populations using single stellar populations of Maraston (2005) or PEGASE (Fioc & Rocca-Volmerange 1997). In this paper we will only use Maraston models; we assume star formation histories (SFHs) with either exponentially decreasing SFR in function of time ("τ models") as

\[
\text{SFR}_\text{old}(t) = \text{SFR}_0,\text{old} \cdot e^{-\frac{(t-t_1)}{\tau_1}}
\]

\[
\text{SFR}_\text{young}(t) = \text{SFR}_0,\text{young} \cdot e^{-\frac{(t-t_2)}{\tau_2}}
\]

(1)

or “box models” characterized by constant SFR over a limited period of time; in this case the instantaneous SFR at look-back time \(t' = 0\) is given by the galaxy mass divided by the age \(t\) of the population, i.e., \(M_{\text{gal}}/t\). Labels 1 and 2 refer to the old and young stellar populations while \(t_{1,2}\) and \(t_{1,2}\) (both is units of Gyr) are their e-folding time and age, respectively. The two stellar populations are linked and weighted through their mass fraction; the parameter \(f_{\text{SP}}\) represents the fraction of the young stellar mass over the total mass, so that the total instantaneous SFR (output parameter of CIGALE) is expressed as

\[
\text{SFR} = (1 - f_{\text{SP}})\text{SFR}_\text{old}(t) + f_{\text{SP}}\text{SFR}_\text{young}(t).
\]

(2)

2.2. Absorption and Emission by Dust and Gas

In CIGALE, the absorption of star light by dust is described by a Calzetti attenuation curve (Calzetti et al. 1994, 2000); possible modifications of the curve include both the addition of a UV bump at about 2175 Å (Fitzpatrick & Massa 1990, 2007; Noll et al. 2009) and the change of the slope of the attenuation by a power law \((\lambda/\lambda_V)^\delta\), where \(\lambda_V = 5500 \text{ Å}\) is the reference wavelength for the V filter; both the amplitude of the bump and the slope \(\delta\) are free parameters of the code. The attenuation correction is applied to both stellar populations individually using the visual attenuation parameter of the young stellar populations \(A_{\text{SP}}\), in units of magnitudes) and a reduction factor of the attenuation for the old model \(f_{\text{SP}}\) as free parameters.7 The code also calculates three additional parameters, derived from the final model SEDs; \(A_{\text{FUV}}\) and \(A_V\) are defined as the effective attenuation factors in magnitudes at 1500 ± 100 Å and 5500 ± 100 Å while the age \(t_{\text{D4000}}\) is derived from the D4000 break (see Balogh et al. 1999) of the unreddened SED for a single stellar population.8

Dust emission in the IR is taken into account using 64 templates of Dale & Helou (2002). These models are parameterized

by \(\alpha\), the power-law slope of the dust mass over heating intensity, defined as follows:

\[
dM_d(U) = U^{-\alpha} dU,
\]

where \(M_d(U)\) is the dust mass heated by a radiation field of intensity \(U\).

Bolometric and dust luminosities \((L_{\text{bol}}\text{ and } L_{\text{dust}}\text{, respectively})\) are derived from all the basic parameters (see Noll et al. 2009) and dust emission due to non-thermal sources such as AGNs can also be added; the fraction \(f_{\text{AGN}}\) of dust luminosity \(L_{\text{dust}}\) (in \(L_{\odot}\)) due to an AGN is estimated using AGN templates from Siebenmorgen et al. (2004a, 2004b).

The spectral line correction due to interstellar gas is performed as in Noll et al. (2009); for the optical band, empirical line templates are taken from the Kinney et al. (1996) starburst spectra while for the UV we use templates derived from SEDs presented in Noll et al. (2004). A correction for the redshift-dependent absorption of the intergalactic medium shortward of the Lyα line is also included using the algorithm of Meiksin (2006).

2.3. Comparison with Data

A grid of theoretical photometric fluxes is calculated at the redshift of the objects considered and a Bayesian analysis is performed through the calculation of the \(\chi^2\) of each model:

\[
\chi^2(M_{\text{gal}}) = \sum_{i=1}^{k} \frac{(M_{\text{gal}} f_{\text{mod},i} - f_{\text{obs},i})^2}{\sigma_{\text{obs},i}^2} \equiv -2 \ln(L),
\]

(4)

here the galaxy mass \(M_{\text{gal}}\) (in \(M_{\odot}\)) is treated as a free parameter, \(f_{\text{mod},i}\) and \(f_{\text{obs},i}\) are the theoretical and experimental fluxes, respectively, the statistical photometry errors are considered in the term \(\sigma_{\text{obs},i}\), and \(L\) is the normalized likelihood function.

3. MCMC TECHNIQUE AND COSMOMC

In Bayesian inference, the posterior probability of the parameters \(\{\theta\}\) of a model in the light of the observed data \(\{d\}\) is given by

\[
P(\theta|d) = \frac{P(d|\theta) P(\theta)}{P(d)};
\]

(5)

here \(P(d|\theta) \equiv L(\theta)\) is the likelihood of the data given the model, \(P(\theta)\) is the prior on the parameters, which quantifies our a priori knowledge of the parameters, and \(P(d)\) (called evidence) is a normalization factor. In our case, \(d\) represents the SED of each galaxy while \(\theta\) represents the astrophysical parameters of CIGALE, as \(\{\theta\} \equiv \{t_1, t_2, f_{\text{SP}}, \ldots\}\). An MCMC sampler provides an efficient way to explore the posterior distribution and ensures that the number density of samples is asymptotically proportional to the probability density.

3.1. Metropolis–Hastings Algorithm

The code COSMOMC uses the Metropolis–Hastings algorithm to generate samples; each chain moves according to a transition probability \(T(\theta_i, \theta_{i+1})\) which is determined so that the Markov Chain has a stationary asymptotic distribution equal to the posterior distribution \(P(\theta)\) that we want to sample from. Given an arbitrary proposal density distribution \(q(\theta_i, \theta_{i+1})\) to propose

---

7 The parameter \(f_{\text{V}}\) has been labeled \(f_{\text{AGN}}\) in some previous papers.

8 In its current version, CIGALE directly outputs the dust-free D4000 break, see Buat et al. (2011b).
This ensures that each point of the chain depends only on its predecessor; in this sense the chain is a Monte Carlo Markov chain. In the grid-based approach, the parameter space is sampled in the same “blind” way for both reliability of results and computation time, as also pointed out in Noll et al. (2009) and Acquaviva et al. (2011). In fact, local minima and degeneracies between parameters can be easily missed or undersampled without a good a priori knowledge of the parameter space; the oversampling of an ill-constrained parameter can also lead to a slight degradation of the estimates of well-constrained parameters and many points can be generated in a region where the posterior is low, resulting in a waste of CPU time. This is not the case when MCMC chains are used because each chain “learns” where to move in the parameter space through the Metropolis–Hastings algorithm so that the density of samples is proportional to the posterior distribution. Degeneracies between parameters are more easily found, especially if many chains, starting from different regions in the parameter space, are used. In other words, the CIGALEMC user needs to specify the prior parameter space (number of parameters and their limits) but not the density of points for each parameter. In the following section, we will provide a comparison of CPU time between the original CIGALE and CIGALEMC when evaluating physical properties of a mock sample. The code also calculates the covariance between various parameters so that an initial run can be made and the covariance matrix obtained can be used to improve the efficiency of sampling for subsequent runs.

Since each MCMC chain starts at a random position in the parameter space, it will take a little time before the chains equilibrate and start sampling the posterior distribution. This period of initial convergence is called burn in period and the first burn in points of each chains will be discarded when doing any statistical analysis. In order to obtain uncorrelated samples of the posterior each chain is also “thinned” by using only occasional points of it; the thinning factor varies according to the number of parameters involved and it is typically in the range 25–50. The code allows to choose the burn in fraction of the chain we want to discard and automatically thins out the chains.

In our analyses, we would not use any initial covariance matrix for the parameters and, in order to be conservative, we compute statistical quantities using only the second-half of each chain. In Figure 2, we plot the points of an MCMC chain for a galaxy sample in the plane $L_{\text{dust}}$ versus $f_{\text{AGN}}$; the chain reaches the sensitive region of the parameter space after only a few “burn in” points characterized by very low values of $L_{\text{dust}}$. Having a set of samples from the full posterior distribution, it is possible to calculate statistical quantities for the parameters of interest. Since the number density of sampling points is proportional to the posterior density, it is easy to calculate mean values and marginalized one-dimensional distribution for each parameter $\theta^i$ by simply counting the number $N$ of samples within binned
ranges of parameter values:

\[ \langle \theta^i \rangle \sim \frac{1}{N} \sum_{j=1}^{N} \theta^j, \]  

(9)

while this is much more difficult in the context of the numerical grid integration because the calculation time grows exponentially with the number of dimensions.

In order to be sure that MCMC chains are efficiently sampling the posterior distribution (and then obtain robust statistics for each parameter), it is important to check their convergence. The code COSMOMC provides two convergence criteria for runs with one single chain (Gelman & Rubin 1992) and with multiple chains (Gelman & Rubin 1992). In the following analysis, we will run multiple chains using the Gelman & Rubin diagnostic, which is characterized by the “variance of chain means”/“mean of chain variances” parameter \( R \); \( |R - 1| \leq 0.03 \) is usually enough to reach convergence and stop the chains.

In this work we use COSMOMC as a generic MCMC sampler and link it to CIGALE in order to allow a faster exploration of the astrophysical parameter space. Our modified CIGALEMC code will be publicly available very soon\(^9\) and, since it is based on COSMOMC for the sampling options, convergence criteria and statistical quantities provided, we refer the reader to the Web site\(^10\) and to Lewis & Bridle (2002) and references therein for a detailed explanation of the code and MCMC methods in general.

4. ANALYSIS OF A MOCK SAMPLE

We test our CIGALEMC code with a mock sample already used in Giovannoli et al. (2011). We consider 62 artificial galaxy SEDs corresponding to Luminous InfraRed Galaxies at redshift \( \bar{z} \sim 0.7 \) and obtained by varying the input parameters of CIGALE in the following ranges: \([0.1 \text{ Gyr} \leq t_1 \leq 10 \text{ Gyr}; 0.025 \text{ Gyr} \leq t_2 \leq 0.7 \text{ Gyr}; 0 \leq f_{\text{fSPI}} \leq 1; 0.6 \text{ mag} \leq A_{\text{SPI}} \leq 2.1 \text{ mag}; 0 \leq f_V \leq 1; 1.0 \leq \alpha \leq 2.5; 0 \leq f_{\text{AGN}} \leq 0.3\); such a wide multidimensional parameter space allows us to model very different SEDs of galaxies characterized by a wide range of possible SFHs, absorption, and emission by gas and dust and AGN contamination. All galaxies are based on a Salpeter initial mass function; the age of the old stellar population is fixed at 7 Gyr; we consider a constant SFR for the young population model \( (t_2 = 20 \text{ Gyr}) \) and do not add any modification to the original Calzetti attenuation curve (no UV bump and \( \delta = 0 \)).

Theoretical fluxes are calculated in the following 17 bands from UV to IR: 0.231 \( \mu m \) for the Galaxy Evolution Explorer (GALEX), \([0.35–0.36–0.46–0.54–0.65–0.87–0.90–1.2–1.6–2.1] \mu m \) (corresponding to MUSYC bands), \([3.6–4.5–5.8–8.0] \mu m \) (corresponding to IRAC photometry), and \([24–70] \mu m \) (corresponding to MIPS photometry). We add a Gaussian distributed uncertainty \( \sigma \) to each theoretical flux; its value is 10% of the corresponding flux, a reasonable choice for the measurements considered above. For each galaxy we run eight chains with initial positions randomly chosen in the parameter space determined by the following set of input parameters:

\[ \{t_1, t_2, f_{\text{fSPI}}, A_{\text{SPI}}, f_V, \alpha, f_{\text{AGN}}, M_{\text{gal}}\}, \]  

(10)

from which statistical quantities of interest can be calculated for the following set of derived parameters:

\[ \{\text{SFR}, t_{\text{D4000}}, L_{\text{bol}}, L_{\text{dust}}, A_{\text{FUV}}, A_{\text{V}}, M_{\text{star}}\}. \]  

(11)

We assume flat priors in the following parameters space: \([0.1 \text{ Gyr} \leq t_1 \leq 10 \text{ Gyr}; 0.025 \text{ Gyr} \leq t_2 \leq 2 \text{ Gyr}; 0 \leq f_{\text{fSPI}} \leq 1; 0 \text{ mag} \leq A_{\text{SPI}} \leq 3 \text{ mag}; 0 \leq f_V \leq 1; 0.5 \leq \alpha \leq 3; 0 \leq f_{\text{AGN}} \leq 1; 8 M_\odot \leq M_{\text{gal}} \leq 14 M_\odot\}; \) chains are stopped when the Gelman & Rubin R-1 parameter is \(|R - 1| \sim 0.03\).

First of all, we want to check that our results are statistically in agreement with the input values for the mock sample. As a tool derived from COSMOMC, CIGALEMC allows to calculate and plot the mean likelihood and marginalized distribution for each parameter.

The marginalized distribution in a given direction of the parameter space \( \tilde{d} = h(\bar{\theta}) \) (where \( h(\bar{\theta}) \) is the projector operator in one of the parameters considered, as \( h(\bar{\theta}) = \theta_i \)) is proportional to the number of samples at \( \tilde{d} \) and it can be expressed as

\[ P(\bar{v}) = M(\bar{P}, \bar{v}) \equiv \int d\bar{\theta} P(\bar{\theta}|\bar{\theta}) \delta(h(\bar{\theta}) - \bar{v}), \]  

(12)

where \( P(\bar{\theta}) \) is the posterior distribution. Assuming flat priors on \( \bar{\theta} \), the mean likelihood of samples with \( \tilde{d} = h(\bar{\theta}) \) can be expressed as

\[ \langle P(\bar{\theta}; h(\bar{\theta}) = \bar{v}) \rangle \equiv \frac{\int d\bar{\theta} P(\bar{\theta}|\bar{\theta}) \delta(h(\bar{\theta}) - \bar{v})}{\int d\bar{\theta} P(\bar{\theta}|\bar{\theta})} = \frac{M(\bar{P}^2, \bar{v})}{M(\bar{P}, \bar{v})}. \]  

(13)

If \( P(\bar{\theta}) \) is a multivariate Gaussian distribution, it is possible to demonstrate that both mean likelihood and marginalized distribution are Gaussian and proportional so that they look the same: differences in these distributions will be then a signal of non-Gaussianity, which can be either intrinsic or due to parameters not very well constrained. In Figure 3 we show, for an artificial galaxy of the mock sample considered, both marginalized distributions (black solid lines) and mean likelihood (dotted lines) for some parameters of interest: we can see that \( t_1, t_2, f_{\text{fSPI}}, \) and \( \alpha \) are not very well constrained by the code. Similar results have been found in the analyses by Noll et al. (2009), Giovannoli et al. (2011), and Buat et al. (2011b).

In order to study the goodness of the fit in a quantitative way for the whole sample of galaxies, we introduce the quantity

\[ X' \equiv \frac{1}{N} \sum_j O_j - I_j}, \]  

(14)

here \( i \) runs for all the parameters considered by the code (so that \( \{X' = \{t_1, t_2, f_{\text{fSPI}}, \ldots\}\}, j \) runs for the \( N = 62 \) objects of the mock, \( O' \) is the set of the best-fit values obtained as output from CIGALEMC, \( P \) is the vector of input parameters, and \( \sigma' \) is the vector of the 68% confidence level marginalized uncertainties. As we can see from Figure 4, all \( X \) values are compatible with \( X = 0 \), which means that the code is able to find the best-fit values of the parameters with high confidence. However, some distributions are slightly skewed: this can be due to chains being stuck in local minima of the likelihood function so that, in some cases, the best fit found does not correspond to the input value. This situation is typical when parameters are not well constrained (as for \( t_2 \)) or in the presence of strong degeneracies between parameters. As an example, Figure 5 shows the 68% and 95% confidence levels in the plane \( f_{\text{SPI}} \) versus \( t_2 \) for one sample galaxy of the mock: as we can see, limits on these parameters are not very strong; we also note a partial degeneracy for high values of \( f_{\text{SPI}} \), which is due to the fact that these two
parameters affect the galactic SFR in the same way since we consider constant SFR for the young model:

$$\text{SFR} \sim f_{ySP} \frac{M_{\text{gal}}}{t_2}. \quad (15)$$

In order to check the reliability of our MCMC algorithm, we tested the parameter uncertainty estimation. We created and fit the SEDs of 100 artificial galaxies in 21 bands from IR to UV, built with the same input parameter set and with a 10% scatter in flux, corresponding to the photometric error considered. We checked that in general we are able to find the true values within the region allowed at 68% (95%) confidence 68% (95%) of the time within the Poisson fluctuation error for different choices for both the input parameters and the prior ranges allowed in the fit.

Finally, we checked the consistency of our results by calculating the Pearson correlation coefficient $r$ between the input values of the parameters used to generate the mock and the best-fit values found with CIGALEMC. The Pearson correlation coefficient quantifies the amount of correlation between two variables $X$ and $Y$ and it assumes values in the range $r \in [-1 : 1]$ (see Cohen 1988). For samples $X_i$ and $Y_i$, it can be written as

$$r = \frac{\sum_{i=1}^{N}(X_i - \bar{X})(Y_i - \bar{Y})}{(\sum_{i=1}^{N}(X_i - \bar{X})^2)^{1/2}(\sum_{i=1}^{N}(Y_i - \bar{Y})^2)^{1/2}}. \quad (16)$$
here $\bar{X}$ and $\bar{Y}$ denote the mean values of $X_i$ and $Y_i$. The interpretation of the strength of the correlation can depend on the context; however, a widely used standard introduced by Cohen (1988) considers Pearson values $|r| > 0.5$ as a signal of large correlation between variables, while $|r| < 0.3$ denotes poor correlation. Again, as we can see from Table 1, some parameters ($\tau_1$, $t_2$, $f_{\nu SP}$) have small values of $r$; this is expected since these parameters are mostly unconstrained; however, even if the exact values of the input parameters are not found, they always fall inside the statistically significant region of the parameter space. We also explicitly checked that results do not significantly change when increasing the prior space for some parameters not very well constrained as $\tau_1$, $t_2$, and $\alpha$. Finally, we tested, for a galaxy of the mock, how results change when considering a Gaussian prior, with different central values and width $\sigma = 0.2$, on one of the most unconstrained parameters, $t_2$; we found that results on the other parameters are always statistically consistent with respect to the choice of a flat prior on $t_2$.

In general, it is a good idea to choose the largest possible prior space for unconstrained parameters in order to avoid possible biases due to the prior choice.

It is useful to compare the performance of CIGALE and CIGALEMC in terms of computing time, especially since computation can become prohibitive for any grid-based method if the number of parameters involved is sufficiently high. The CPU time required to obtain convergence of the chains for each galaxy

Figure 4. Distribution of the variable $X$ for some parameters considered; $X$ values are compatible with 0 at 68% confidence level.

(A color version of this figure is available in the online journal.)
mainly depends on both the quality of the data and the number of parameters considered. Running eight chains in parallel (each one on a 2.40 GHz Intel Xeon E5530), for each galaxy of the mock we typically reach a good convergence with ∼20,000 points for each chain, which means a total of ∼160,000 points. The grid built in Giovannoli et al. (2011) to analyze the same sample with the same number of free parameters and bounds contained ∼3.5 × 10^6 points; this means a gain of order ∼20 in efficiency in the estimation of the parameters but a more dramatic efficiency can be easily reached when we need to use either more parameters or a fine sampling in a given direction of the parameter space or both. The average CPU time to reach a good convergence for a galaxy of this mock is about 35 s, which translates in 280 s of total CPU time.

5. ANALYSIS OF REAL DATA: THE SINGS SAMPLE

We now use CIGALEMC to infer physical properties of the well-known SINGS (Kennicutt et al. 2003) sample. In order to make a comparison with results obtained using the grid-based CIGALE, we use the same 39 galaxies already considered in Noll et al. (2009) with the same spectral coverage: GALEX far-UV (≈1500 Å) and near-UV (≈2300 Å) filters (Gil de Paz et al. 2007), Two Micron All Sky Survey (2MASS) data for J, H, and Ks, and SDSS u, g, r, i, z filters. Their sample contains 740:22 (11pp), 2011 October 10

In our analysis, we assume the following range of variation for a set of nine astrophysical parameters: [0.1 Gyr ≤ t < 10 Gyr, 0.025 Gyr ≤ t_2 < 2 Gyr, 0 ≤ f_{AGN} ≤ 1, −0.5 ≤ δ ≤ 0.5, 0 mag ≤ A_{SP} ≤ 5 mag, 0 ≤ f_V ≤ 1, 0.5 ≤ α ≤ 3.0, 0 ≤ f_{AGN} ≤ 1, 8 M_☉, M_{gal} < 13 M_☉]. We keep fixed both the age of the old stellar population (t_1 = 10 Gyr) and the e-folding time for the young stellar population (t_2 = 20 Gyr); these parameters are not well constrained by the data so that fixing them does not alter the fit. Finally, we only consider models with a Salpeter initial mass function and solar metallicity; metallicity measurements are quite uncertain due to many limiting factors (Noll et al. 2009; Moustakas & Kennicutt 2006, and references therein); Noll et al. (2009) checked the influence of different assumptions for the metallicity, concluding that deviations in the properties of the galaxies are within the uncertainties, which tend to increase by 0%–20% when half of or double the solar metallicity are considered. The only exception is the absolute value of t_{4000} because of the well-known age–metallicity degeneracy (e.g., Kodama & Arimoto 1997).

Our reference AGN model has L_3900 = 10^{12} L_☉, R = 125 pc, and A_V = 32 for the luminosity of the non-thermal source, the outer radius of a spherical dust cloud covering the AGN, and the amount of attenuation in the optical caused by the cloud, respectively.

Our findings can be summarized as follows.

1. Very good constraints are derived for the AGN fraction of all sources (f_{AGN} ≤ 0.10) except for NGC0584 and NGC 1404 for which f_{AGN} ≤ 0.6 at the 68% confidence level. We note that the flux at 160 μm for NGC 0584 can be contaminated by some foreground/background emission and the most recent Herschel data conclusively confirm that a background source contaminates both fluxes at 70 and 160 μm for NGC 1404 (D. A. Dale 2011, 

Table 1

| Parameters | Mock Sample |
|-----------|-------------|
| τ_1       | 0.33        |
| log_{10} M_{dar} | 0.93 |
| log_{10} L_{bol} | 0.92 |
| log_{10} L_{dust} | 0.88 |
| log_{10} SFR | 0.81 |
| t_2       | 0.26        |
| f_{SP}    | 0.30        |
| A_{SP}    | 0.81        |
| f_V       | 0.70        |
| α         | 0.65        |
| f_{AGN}   | 0.71        |

Table 2

| Filters | Rel. Errors |
|---------|-------------|
| GALEX FUV, NUV | 15% |
| Dale et al. B, V, R, I | 16% |
| SDSS u, g, r, i, z | 3% |
| 2MASS J, H, Ks | 1% |
| IRAC 3.6, 4.5, 5.8, 8.0 μm | 11% |
| MIPS 24 μm | 5% |
| MIPS 70 μm | 7% |
| MIPS 160 μm | 13% |

Figure 5. Two-dimensional marginalized distribution showing the 68% and 95% confidence level contours for f_{SP} and t_2 for a galaxy sample; the parameter t_2 is unconstrained by the data and it is partially degenerate with f_{SP}.

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

Table 1

Estimation of the Linear Correlation Coefficient of Pearson between the Exact Value and the Value Estimated by the Code CIGALEMC for Some Parameters of the Mock Catalog

Table 2

Mean Photometric Uncertainties for Our SINGS Sample
private communication); the low quality of these data is responsible for the big uncertainties obtained for other parameters such as SFR, \(L_{\text{bol}}\), \(L_{\text{dust}}\), and \(A_{\text{FUV}}\); in Figure 6 we plot the two-dimensional marginalized distribution for SFR versus \(f_{\text{AGN}}\) for NGC 1404; the double-peaked likelihood function is most probably an artifact due to the low quality of data for this source. In general, we note that NGC 0584 and NGC 1404 are elliptical galaxies which tend to have very weak but warm dust emission; it is not surprising that they show an apparently high AGN fraction. We decided not to use these sources in the rest of our analysis.

2. We are not able to put strong constraints on “phenomenological” parameters as \(t_1, t_2\), and \(\delta\); limits on these parameters depend on the assumed prior range. The poor determination of \(\delta\) is essentially due to the low number of data in UV. In general, both \(\delta\) and the possible UV bump allowed by the code are very difficult to constrain with few only broadband data (Buat et al. 2011a, 2011b). The fraction of the young stellar mass over the total mass is well constrained, with values \(f_{\text{SP}} < 0.26\) for all the galaxies, except for NGC 1705, NGC 2798, and NGC 4631, for which \(f_{\text{SP}}\) is unconstrained. The mean value for the parameter \(\alpha\) of the Dale & Helou (2002) templates is \(\alpha = 2.44\) (in agreement with Noll et al. 2009); we note that, while low values of \(\alpha\) are well constrained, no constraints are found on high values, with \(\alpha = 3\) compatible at the 68% confidence level for all galaxies expect NGC 2798, for which \(\alpha < 1.77\) at the 68% confidence level; weak constraints on high values for \(\alpha\) are due to the degeneracy of the dust emission models for wavelengths short of the emission peak, so that, for \(\alpha > 2.5\), the flux ratio \(f_{\lambda}(60)\mu m/f_{\lambda}(100)\mu m\) is almost constant (see Noll et al. 2009).

3. A frequency-dependent \(\chi^2\) analysis shows which bands mostly contribute to the total \(\chi^2\) for each galaxy. We introduce the averaged \(\chi^2\):

\[
\langle \chi^2(\lambda^j) \rangle = \frac{1}{N_j} \sum_{i=1}^{N_j} \left( \frac{\text{Th}_i^j - d_i^j}{\sigma_i^j} \right)^2;
\]

where \(\text{Th}_i^j\) and \(d_i^j\) are, respectively, the theoretical best-fit SEDs and the data points for the \(j\)th galaxies at the \(i\)th frequency and the sum runs over our sample of galaxies. From Figure 7, we see that the code is able to find a good agreement with the data for all the bands considered. The frequency-dependent \(\chi^2\) is particularly low for bands with large relative errors as UV GALEX bands and optical B, V, R, and I bands. The large uncertainty found for the MIPS 70 \(\mu m\) filter is mainly due to the galaxies NGC 1715 and NGC 5866; the code is not able to find a good fit for the 70 \(\mu m\) filter and the reduced \(\chi^2\) is high (\(\chi^2/\text{degrees of freedom (dof)} = 7.8\) for both galaxies, see Table 3). A previous analysis by Cannon et al. (2006), based on Spitzer observations, found similar results for NGC 1705, showing in particular that the models of Li & Draine (2001, 2002) give a better fit to IR data than the Dale & Helou (2002) models used in this paper (see Figure 3 of Cannon et al. 2006). Interestingly enough, NGC 1715 and NGC 5866 are, respectively, the only irregular and S0 galaxies in our sample.

4. Good constraints can be derived for the mass-dependent parameters \(M_{\text{star}}, \text{SFR}, L_{\text{bol}}, \text{and } L_{\text{dust}}\) as shown in Table 3. The Pearson correlation coefficient with results from Noll et al. (2009) is \(r = 0.99, 0.98, 0.94, 1.0\), 1.0,
and 0.94 for \(M_{\text{star}}\), SFR, \(l_{\text{D4000}}\), \(L_{\text{bol}}\), \(L_{\text{dust}}\), and \(A_{\text{FUV}}\), respectively. In Figure 8, we show a quantitative comparison with the analysis performed by Noll et al. (2009) with the original code CIGALE by plotting the ratio, \(Q \equiv (\text{CIGALEMC} - \text{CIGALE})/\sigma_{\text{CIGALEMC}}\), where CIGALEMC and CIGALE refer to the mean value of the parameters quoted in Table 3 for this work and Noll et al. (2009), respectively. Possible systematic differences between the results of both methods can be studied by considering mean, standard deviation, and skewness for the parameters of interest. As we can see from Table 4, no significant difference between results in this paper and in Noll et al. (2009) is found for SFR, log\(l_{\text{D4000}}\), log\(L_{\text{bol}}\), and \(A_{\text{FUV}}\), with each value compatible with 0 at less or about the 68% confidence level; however, there is some skewness associated with SFR, \(l_{\text{D4000}}\), and \(L_{\text{bol}}\). In the cases of \(M_{\text{star}}\) and \(L_{\text{dust}}\), the mean values are lower from 0 at about the 95% confidence level, indicating the existence of some offset between results in this paper and in Noll et al. (2009). A further comparison of the uncertainties between our results and Noll et al. (2009) shows that we are able to put stronger constraints on the galaxy SFR, while our constraints on \(M_{\text{star}}\), \(L_{\text{bol}}\), \(L_{\text{dust}}\), and \(A_{\text{FUV}}\) are weaker. Differences between our results and results in Noll et al. (2009) can be due to the different choice of the input parameter space in terms of both number and type of parameters used in the analysis; in particular, we consider a wider range of variation for \(\alpha\), \(\delta\), \(A_{\text{sp}}\), \(t_{2}\), and \(f_{\text{AGN}}\), while we decide to fix the parameter \(t_{2}\) (see Table 3 of Noll et al.)
2009 for their choice of the parameter space); however, it is reassuring to see how different theoretical assumptions lead to compatible results.

6. CONCLUSIONS

In this paper, we have introduced an MCMC sampling method for the astrophysical parameter estimation from SED fitting with CIGALE. We have shown the following advantages of our modified CIGALEMC code over the usual grid-based CIGALE:

1. Its efficiency, in terms of CPU time, through the Metropolis–Hastings algorithm. Most of the sampling points are drawn in the region where the posterior probability is high, while in the grid-based approach all regions are sampled in the same way. Moreover, marginalized one-dimensional probability distributions for the parameters of interest are calculated by simply counting the number of samples within a binned range of parameter values, the density of sampling points being proportional to the posterior probability. It is hard to do the same with the usual grid approach, since the integration calculation scales exponentially with the number of dimensions. The analysis of a mock sample shows that CIGALEMC needs 20 times less points than CIGALE to reach convergence for a given galaxy but, in general, results depend on both the number of parameters and the prior used.

2. Its accuracy; degeneracies between parameters are easily found, convergence criteria (already implemented in the code) ensure that statistical quantities for the parameters of interest are robustly determined and cross-checks through statistical analysis of mock catalogs are not necessary.

3. Its “user friendly” characteristics. The user does not need to decide a priori the number density of samples for each region, trying to find a compromise between the accuracy of the results and the speed of the code: only the prior range must be chosen in advance for CIGALEMC; in fact the Metropolis–Hastings algorithm automatically samples adequately the posterior probability according to its values.

Our code will be available very soon at this Web site: http://www.oamp.fr/cigale/.

P.S. thanks Denny Dale for very useful discussions. We thank Antony Lewis for providing a publicly available version of COSMOMC.

REFERENCES

Acquaviva, V., Gawiser, E., Guita, L., et al. 2011, ApJ, in press
Balogh, M. L., Morris, S. L., Yee, H. K. C., et al. 1999, ApJ, 527, 54
Bruzual, G., & Charlot, S. 2003, MNRAS, 344, 1000
Buat, V., Giovannoli, E., Burgarella, D., et al. (Herschel Collaboration) 2010, MNRAS, 409, 11
Buat, V., Giovannoli, E., Heinis, S., et al. 2011a, A&A, in press (arXiv:1107.1049)
Buat, V., Giovannoli, E., Takeuchi, T. T., et al. 2011b, A&A, 529, A22
Burgarella, D., Buat, V., & Iglesias-Páramo, J. 2005, MNRAS, 360, 1413
Calzetti, D., Armus, L., Bohlin, R. C., et al. 2000, ApJ, 533, 682
Calzetti, D., Kinney, A., & Storchi-Bergmann, T. 1994, ApJ, 429, 582
Cannon, J. M., Smith, J.-D. T., Walter, F., et al. 2006, ApJ, 647, 293
Chary, R., & Elbaz, D. 2001, ApJ, 556, 562
Christensen, N., Meyer, R., Knox, L., & Luzy, B. 2001, Class. Quantum Grav., 18, 2677
Cohen, J. 1988, Statistical Power Analysis for the Behavioral Sciences (Hillsdale, NJ: LEA)
Conroy, C., Gunn, J. E., & White, M. 2009, ApJ, 699, 486
da Cunha, E., Charlot, S., & Elbaz, D. 2008, MNRAS, 388, 1595
Dale, D. A., Bendo, G. J., Engelbracht, C. W., et al. 2005, ApJ, 633, 857
Dale, D. A., Gil de Paz, A., Gordon, K. D., et al. 2007, ApJ, 655, 863
Dale, D. A., Gil de Paz, A., Gordon, K. D., et al. 2008, ApJ, 672, 735
Dale, D. A., & Helou, G. 2002, ApJ, 576, 159
Dopita, M. A., Groves, B. A., Fischera, J., et al. 2005, ApJ, 619, 755
Fioc, M., & Rocca-Volmerange, B. 1997, A&A, 326, 950
Fitzpatrick, E. L., & Massa, D. 2007, ApJ, 663, 320
Gelman, A., & Rubin, D. B. 1992, Stat. Sci., 7, 457
Gil de Paz, A., Boissier, S., Madore, B. F., et al. 2007, ApJS, 173, 185
Giovannoli, E., Buat, V., Noll, S., Burgarella, D., & Magnelli, B. 2011, \textit{A&A}, 525, A150
Jarrett, T. H., Chester, T., Cutri, R., Schneider, S. E., & Huchra, J. P. 2003, \textit{AJ}, 125, 525
Kennicutt, R. C., Jr., Armus, L., Bendo, G., et al. 2003, \textit{PASP}, 115, 928
Kinney, A. L., Calzetti, D., Bohlin, R. C., et al. 1996, \textit{ApJ}, 467, 38
Kodama, T., & Arimoto, N. 1997, \textit{A&A}, 320, 41
Lagache, G., Dole, H., Puget, J. L., et al. 2003, \textit{MNRAS}, 338, 555
Lagache, G., Dole, H., Puget, J. L., et al. 2004, \textit{ApJS}, 154, 112
Lewis, A., & Bridle, S. 2002, \textit{Phys. Rev. D}, 66, 103511
Li, A., & Draine, B. T. 2001, \textit{ApJ}, 554, 778
Li, A., & Draine, B. T. 2002, \textit{ApJ}, 576, 762
Maraston, C. 2005, \textit{MNRAS}, 362, 799
Meiksin, A. 2006, \textit{MNRAS}, 365, 807
Moustakas, J., & Kennicutt, R. C., Jr. 2006, \textit{ApJ}, 651, 155
Muñoz-Mateos, J. C., Gil de Paz, A., Zamorano, J., et al. 2009, \textit{ApJ}, 703, 1569
Noll, S., Burgarella, D., Giovannoli, E., et al. 2009, \textit{A&A}, 507, 1793
Noll, S., Mehlert, D., Appenzeller, I., et al. 2004, \textit{A&A}, 418, 885
Raftery, A. E., & Lewis, S. M. 1992, \textit{Stat. Sci.}, 7, 493
Siebenmorgen, R., Freundling, W., Krügel, E., & Haas, M. 2004b, \textit{A&A}, 421, 129
Siebenmorgen, R., & Krügel, E. 2007, \textit{A&A}, 461, 445
Siebenmorgen, R., Krügel, E., & Spoon, H. W. W. 2004a, \textit{A&A}, 414, 123
Silva, L., Granato, G. L., Bressan, A., & Danese, L. 1998, \textit{ApJ}, 509, 103
Stoughton, S., Lupton, R. H., Bernardi, M., et al. 2002, \textit{AJ}, 123, 485
Witt, A. N., & Gordon, K. D. 2000, \textit{ApJ}, 528, 799